



ISSUES RELATED TO MISSING ATTRIBUTES IN A-POSTERIORI NOVELTY ASSESSMENTS

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

Abstract

Novelty assessment is a fundamental activity for creativity evaluation of ideas or concepts. Accordingly, literature acknowledges a variety of suitable metrics, among which some “a-posteriori” versions are often used for assessing idea generated in experimental sessions. We observed that when in presence of tasks requiring multiple functions, the application of the metric becomes problematic. In this paper, we argue about the recalled issue, providing detailed suggestions for a correct “a-posteriori” novelty assessment of ideas generated from tasks with multiple required functions.

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

1. Introduction

Nowadays, innovation is one of the most yearned targets for both industry and academia. Indeed, while the industrial stakeholders constantly search for newer and more efficient products, processes and services, scholars are involved in several studies about the recalled matters.

Accordingly, several research activities exist around the identification of more comprehensive definitions of the interrelated concepts of creativity and innovation, which are also similarly related to the concept of newness and usefulness (Oman et al., 2013). Consequently, many definitions of creativity can be found in literature (Sarkar and Chakrabarti, 2008), and for some of them the concept of novelty plays a key role (Sarkar and Chakrabarti, 2011; Corazza and Agnoli, 2016).

Moreover, in order to assist and enhance creativity and innovation, more and more contributions can be found in literature about cognitive aspects, development of comprehensive design methods and tools or, in general, idea generation approaches and their mutual integration, e.g. (Shah et al., 2001; Howard et al., 2010; Frillici et al., 2015; Fiorineschi et al., 2018).

Therefore, in order to evaluate and validate the effectiveness of the proposed methodological approaches, researchers need to “assess creativity” of generated ideas. Accordingly, several studies exist, comparing the effectiveness of different approaches for idea generation, e.g. (Chulvi et al., 2012; Vargas-Hernandez et al., 2013).

A comprehensive review of the available metrics for assessing creativity is not in the scope of this paper, but among them, the metric proposed by Shah et al. (2003) is one of the most acknowledged and used by researchers, where “ideation effectiveness” is assessed by means of four distinct metrics, i.e. “quantity”, “quality”, “variety” and “novelty”.

However, the experiences we had during practical research activities highlighted non-negligible problems in applying the “novelty” metric. In particular, we refer to the “a-posteriori” version of novelty, which is claimed to measure the unusualness or “uncommonness” of a specific idea in relation to a given group of other ideas. More specifically, the recalled problems arose when assessing ideas coming from design tasks where multiple functions were required for the design outcomes, but the latter implemented

only a part of the requested functionalities. In the recalled cases, we observed that the novelty metric of Shah et al. (2003) led to incoherent results and consequently we also observed that managing the problem of novelty assessment becomes not trivial.

Therefore, with this paper we aim at comprehensively introducing the above mentioned applicability problem of the novelty metric proposed by Shah et al., arguing about it, and proposing suggestions for its resolution. In order to accomplish 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 argues about the application of some trivial solutions. In Section 4, suggestions are reported about comprehensive solutions to overcome the observed applicability problem. Eventually, final considerations and possible research hints are reported in the last section.

2. Related work

Literature acknowledges different ways to assess creativity, among which, some of them adopt specific metrics to evaluate ideas, new products, or even people (Johnson et al., 2016). Especially concerning product and process creativity evaluations, it is possible to identify different metrics based on a multitude of parameters like originality, usefulness, fluency, flexibility, etc. (Vargas-Hernandez et al., 2012). Among the recalled metrics, one of the most diffused is that of Shah et al. (2003), 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 the feasibility of ideas and their relevance to 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. Shah et al. (2003) also proposed four well-known metrics, i.e. one for each of the above mentioned parameters. However, a first distinction can be made, since 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). For Variety assessment, the recalled 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. Successively, 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 (see Figure 1). 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.

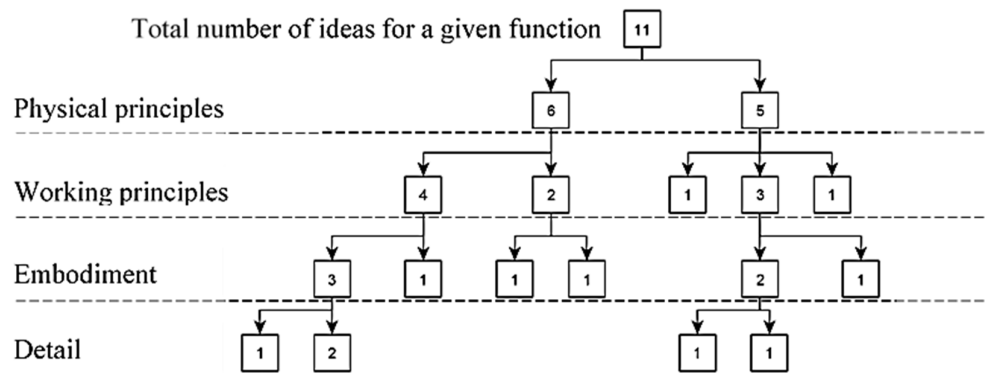


Figure 1. Example of genealogy tree from Shah et al. (2003)

For assessing quality, Shah et al. 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 quality.

Concerning Novelty, Shah et al. (2003) 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 of the designed system, 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 existing systems.

In the approach proposed by Shah et al. (2003), i.e. the "a-posteriori" one (SNM), they suggest to identify some recurring "key-attributes" among the set of ideas to be examined, and to find the different ways by which each of them has been satisfied. Moreover, each of the identified key attribute is associated to a normalized weight, indicating its relative importance. Once the key attributes have been identified and the related weight assigned, novelty is assessed by scoring the ideas generated for each key attribute. Then, scores are summed together by multiplying each of them for the related normalized weight. More precisely the novelty score for each attribute S is calculated by (1):

$$S_{ij} = \frac{T_{ij}-C_{ij}}{T_{ij}} \times 10 \quad (1)$$

Where T_{ij} is the total number of solutions (or ideas) conceived for the key attribute i , and design stage j ; C_{ij} is the count of the current solution for the attribute i , and design stage j . Then, the overall novelty of each idea M is calculated through (2):

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

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 .

The example of the "a-posteriori" assessment reported in Shah et al. (2003), concerns a single stage design task where designers (students) identified 46 ideas characterized by three key attributes.

In the recalled example, 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.

One of the most acknowledged limits of 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 with those of past idea generation sessions or with marketed products (Srivathsavai et al., 2010). Moreover, it has been observed that the SNM novelty score is dependent on the number of similar ideas conceived in a given set, therefore the greater the number of similar ideas, the lower the novelty score (Srivathsavai et al., 2010).

Brown (2014) highlighted several issues concerning SNM, among which, the subjectivity affecting 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 of solutions. 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 sensitive only to differences between concepts at the embodiment level, excluding more abstract differences.

The above mentioned issues have been partially solved by some variant of SNM (described in the following paragraph), but we noticed that the particular problem that we faced has not been mentioned. More precisely, Sluis-Thiescheffer et al. (2016) recently observed that the example in the paper of Shah et al. (2003) 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. Therefore, according to Sluis-Thiescheffer et al. (2016), we found that depending on the considered sample of designers, the design task, and the particular examined design stage, it is possible that in some cases, some of the conceived solutions only partially satisfy the requirement set, i.e. sometimes important attributes (or functions) are neglected. As shown in the following paragraph, to date the recalled issue has never been comprehensively faced in literature.

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.

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 (see Figure 1). 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 , as shown in Equation 3:

$$S_{ij} = \frac{1}{C_{wp}m} \quad (3)$$

Where C_{wp} is the count of occurrences in specific WP branch, and m is the number of branches at the WP level (see Figure 1). 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 observed for SNM (among which, the lack of consideration for uncommonness of PPs and WPs), they proposed a new metric variant based on GT, where an additional level has been added upon the PP one, i.e. the "Strategy" level. Moreover, they proposed a new formula for calculating M (for a single function), as reported in Equation 4:

$$M = \sum_{j=1}^m \frac{(1-P_j)S_j}{2} \quad (4)$$

Where P_j is the count of responses at the j level, divided by the count of responses of its parents.

Eventually, Sluis-Thiescheffer et al. (2016) proposed to assess novelty with a binary metric (novel or not novel), where an arbitrary threshold is considered for the identification of less frequent (novel) solutions. However, although the same authors highlight the possibility to deal with missing attributes (i.e. with concepts or ideas satisfying only part of the key attributes), they fail in comprehensively explain how their proposal solves the problem, especially for single idea assessments.

Therefore, on the light of the current improvements of SMN reviewed here, it is possible to assert that the problem of missing attributes in novelty assessments has not been examined in deep, and then still deserves to be investigated appropriately.

3. Missing attributes in a-posteriori novelty assessments

This section describes the issues that we identified for a-posteriori novelty assessments applied to ideas that miss to satisfy some of the required functions and/or some of the key attributes.

In particular, the related problems are introduced and analysed, and some examples for SNM and a particular GT-based variant of the metric are reported.

3.1. Introducing the problem

Except for that of Sluis-Thiescheffer et al. (2016), the examples reported in the contributions cited in Section 2 concern particular cases where the ideas to be assessed satisfy all attributes and/or functions. However, we have been recently involved in some experimental activities devoted to compare the ideation effectiveness of two conceptual design methods, i.e. the well-known Functional Decomposition and Morphology (FDM) (Pahl et al., 2007) vs. the Problem Solution Network approach (Fiorineschi et al., 2016), which is claimed to overcome some flaws ascribed to FDM. In this particular case, we assigned a design task (toothbrush device) to a sample of engineering students, and we required the implementation of two main functions for the system, that were "removing dirt from teeth" and "removing bacteria from the device after its use". For novelty assessment of ideas we considered a variant of the metric proposed by Peeters et al. (2010), where we simply substituted "EMB" item with

a more general "structure" (STR) one. Indeed, in our case the structural information about proposed solutions was represented only by rough conceptual design sketches, without accessing to the Embodiment design stage. Additionally, since we considered more than one function, we also needed to take into consideration their related importance by assigning weights. Therefore, Equation 2, for only one design stage (conceptual), was modified in Equation 5:

$$M_{GT} = \sum_{i=1}^m f_i \sum_{j=1}^o S_{ij} w_{ij} \quad (5)$$

Where f_i is the normalized weight of function i , m is the number of functions, o is the number of items (PP, WP and STR), and w_{ij} is the normalized weight assigned to the item j , related to the function i .

When analysing the ideas generated by students, we noticed that in some cases, the second functional requirement was not satisfied, implying the impossibility to score the novelty of the related items (see Idea 1 in Table 1). More precisely, applying Equations 1 and 2 (with items in place of attributes), the novelty score of the "incomplete" ideas resulted to be not comparable with others "complete" ones, since one half of the summation (multiplied by the function's weight) is missing.

For example, considering a normalized set of weights for the three items (0,59 for PP, 0,29 for WP and 0,12 for STR), and considering the same weight for each function ($f_i=0,5$), Equation 5 applied to generic S values, leads to the novelty scores listed in Table 1.

Table 1. Generic examples with a missing function and/or items in the GT. Overall novelty score calculated by Equation 5

	S scores for Function 1 ($f_i=0,5$)			S scores for Function 2 ($f_i=0,5$)			Overall Scores M_{GT}
	PP ($w_{ij}=0,59$)	WP ($w_{ij}=0,29$)	STR ($w_{ij}=0,12$)	PP ($w_{ij}=0,59$)	WP ($w_{ij}=0,29$)	STR ($w_{ij}=0,12$)	
Idea 1	10	10	10	-	-	-	5
Idea 2	5	5	5	5	5	5	5
Idea 3	10	10	10	10	10	10	10
Idea 4	10	10	10	5	-	-	6,48

As shown in Table 1, although Idea 1 reached the highest S-score for items related to Function 1, its overall novelty score is the same of Idea 2, which reached a mediocre score for each item, but for all the requested functions.

The problem can be extended to SNM, in a similar case where the key attributes are implicitly related to one or more functions. In such a case, the novelty scores calculated with Equation 2 (with generic f_i assignments) are presented in Table 2.

As shown in Table 2, due to the weights of the missing attributes, the overall novelty score of Idea 1 is higher than that of Idea 2, but lower than that of Idea 3.

The mere application of Equations 5 and 2 respectively to the examples of Table 1 and 2, is somehow equivalent to assign S=0 for missing items or attributes. However, assigning a "0" score to the novelty of the missing items is a non-sense, because it is not possible to assess the novelty of something that does not exist.

Table 2. Generic examples with missing SNM attributes. Overall novelty score calculated by Equation 2 for a single design stage

	S scores for key attributes						Overall Scores M_{SNM}
	Att. 1 ($f_i=0,3$)	Att. 2 ($f_i=0,3$)	Att. 3 ($f_i=0,15$)	Att. 4 ($f_i=0,15$)	Att. 5 ($f_i=0,05$)	Att. 6 ($f_i=0,05$)	
Idea 1	10	10	10	-	-	-	7,50
Idea 2	5	5	5	5	5	5	5
Idea 3	10	10	10	10	10	10	10
Idea 4	10	10	10	5	-	-	8,25

The problem is not trivial to solve, also because possible solutions have different consequences if applied on classical SNM rather than on the above mentioned GT-based novelty metric for multiple functions.

3.2. Possible solutions and related issues for SNM

To overcome the issues related to missing attributes, some quite trivial solutions may be inferred for SNM. For example, it would be hypothetically possible to select only those key attributes that appear on the whole set of ideas to be assessed, i.e. considering an intersection of the key attributes characterizing each single idea. Unfortunately, in this way it is possible to lose a lot of important information about the actual novelty score: if even only one solution does not consider a key attribute (e.g. related to an important function), the assessment of the whole sample of ideas cannot consider such a parameter. Therefore, the intersection option can be too drastic for assessing ideas with SNM.

Another drastic solution could be that of "excluding" those ideas that fail in satisfying all the considered key attributes. However, it seems to be not correct, because "missing attributes" should affect the "quality" assessment (i.e. how much the functional requirements have been fulfilled), not the novelty one. Therefore, both the intersection of attributes and the exclusion of "imperfect" ideas are unsatisfactory.

3.3. Possible solutions and related issues for GT-based versions

The considerations reported here are based on the GT metric introduced in this section, which uses the three specific items (PP, WP and STR) for each function (see the example reported in Table 1). Nevertheless, it is possible to assert that the same issues characterize all of the GT-based novelty metrics introduced in Section 2, if opportunely adjusted for considering multiple functions.

Referring to what observed for SNM, the same considerations can be made here, simply by considering functions in place of attributes. However, in the GT case, an additional set of parameters exists, i.e. the items. Therefore, even if all functions have been implemented in a given idea, it is possible that the designer has represented only part of the related items. Accordingly, considering for example the Idea 4 in Table 1, the second function has been implemented, but the related solution has been represented only in terms of PP (e.g. "use electromagnetic waves"). Even in this case, although most of the present items reached the highest novelty score, Equation 5 leads to a mediocre overall quality score, slightly superior to that of Idea 1.

Therefore, similar to what observed for key attributes in SNM, some issues arise, concerning an interrelationship between quality and novelty assessments. However, here it is necessary to distinguish two different types of quality:

- Missing functions should negatively affect quality in terms of satisfaction of functional requirements.
- Missing items should negatively affect a metric based on the representation quality.

Indeed, quality metrics exists, which also consider "how well" or "how much" the idea has been represented (Choo et al., 2014; Berthelsdorf et al., 2016). In case of a missing items belonging to the GT, it would be possible to assert that the designer failed in reporting all the requested information (e.g. indicating only the PP, but missing the WP and the related structural details).

In any case, according to the measurement of ideation effectiveness originally proposed by Shah et al. (2003), novelty and quality assessments should be performed independently each other.

3.4. Possible consequences from missing attributes/functions/items

The recalled issues (for both SNM and GT-based metrics) could lead to wrong evaluations and/or selections in experimental and/or design contexts. More precisely, if missing attributes are not managed properly during ideas or methods assessments, scores could be unreliable, and results from different experimental sessions could be incomparable.

3.4.1. Consequences in assessing ideas

As shown in the examples reported in Table 2, missing attributes lead to overall novelty scores that sometimes do not reflect the actual "uncommonness" of the assessed idea. However, the two drastic

strategies inferred in this section potentially lead to unprecise results, where also quality considerations are implicitly involved. Consequently, the presence of non-standardized ways to face the problem of missing attributes, could somehow lead to confusion when comparing results from different activities.

More precisely, according to the generic example of Table 2, the exclusion of imperfect ideas leads to a false novelty ranking. Indeed, according to the recalled strategy, Idea 1 should be excluded, but it is not possible to assert that it is less novel than Idea 2, or even than Idea 3. Maybe, it would be possible to assert that Idea 3 satisfied more requirements than Idea 1, but this consideration should be related to quality aspects. Similar criticalities can be observed by performing the assessment also only on the intersection of the attributes.

As observed for SNM, missing functions and/or items in the GT case lead to overall novelty scores that sometimes do not reflect the actual uncommonness of the assessed idea (see Table 1). Similarly, the two drastic strategies (exclusion and intersection) potentially lead to unprecise and different results, which in this case are implicitly affected by two different quality types. Moreover, while the identification of the key attributes is a very subjective activity, the identification of functions and items are directly related to the requirement list characterizing the assigned design task. Therefore, in the GT case, it is even more evident that required attributes (here functions and items) should be excluded from the assessment.

3.4.2. Consequences in assessing design methods or idea generation approaches

The issues highlighted for ideas assessment also affect methods assessment. Indeed, in the SNM case the mean novelty score of a set of ideas generated by exploiting a given method/approach can be calculated simply by considering the overall scores of each generated idea, as shown in Figure 2 (say, horizontal-vertical method, or HV). However, the mean novelty score for a set of ideas can be calculated also by considering the sum of the mean S-scores of each attribute, each of them multiplied for the related weight as shown in Figure 3 (say, vertical-horizontal method, or VH). Similarly, also in the GT case the mean novelty score can be calculated by HV, but in this case the VH should also take into consideration the weights of the functions, which are implemented by each specific set of items.

However, when all attributes are present, both in the SNM and the GT cases, the recalled methods for calculating the mean novelty score lead to the same value. Nevertheless, in presence of incomplete solutions the two methods lead to different scores, where the difference is dependent on the percentage of missing attributes/function/items and the entity of their weights.

Therefore, non-negligible problems may arise when comparing results from different experiments, especially if conducted by different research staffs. Indeed, in case of a different management of missing attributes and/or also in case of different ways for the calculation of the mean novelty score, the assessments may be incomparable. Although research works on this argument are usually presented by publications with detailed descriptions of the adopted methodology, it is possible that recalled details elude the attention of the authors and/or the readers, invalidating successive considerations or comparisons between similar experiments. It is easy to infer that such an eventuality should be avoided.

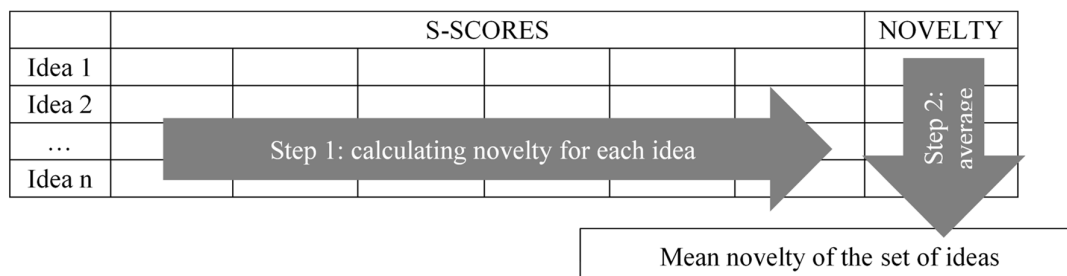


Figure 2. Horizontal-Vertical approach (HV) for calculating the mean novelty score of a set of ideas

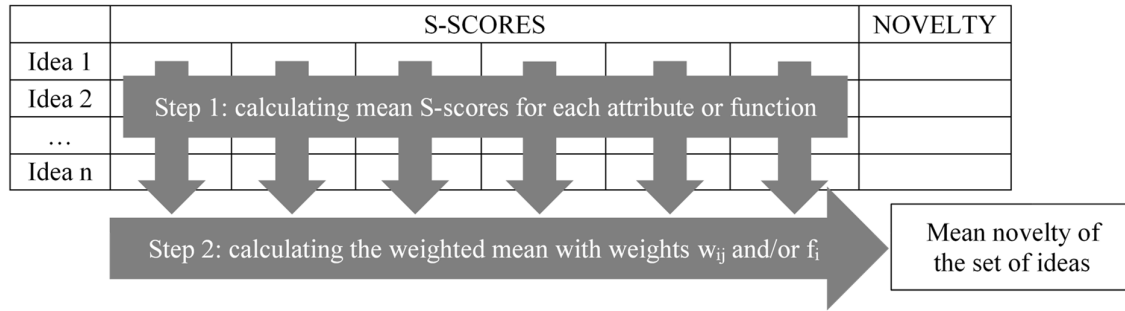


Figure 3. Vertical-Horizontal approach (VH) for calculating the mean novelty score of a set of ideas

4. Suggestions for a comprehensive a-posteriori novelty assessment

In this section, we provide specific suggestions for performing comprehensive a-posteriori novelty assessments and we propose a procedure for assessing novelty in idea generation or design methods.

4.1. Suggestions for idea assessment

As shown in Section 3, in case of missing attributes it can be difficult to rank idea or concept variants only in terms of novelty. Indeed, besides the novelty score, it is necessary to consider quality aspects, e.g. by opportunely selected quality metrics (e.g. fulfilment of functional requirements). However, the formulation and the validation of the recalled metrics are out of the scope of this paper, which is focused on pure novelty aspects.

Here in the following paragraph, specific suggestions are reported for both SNM and GT cases.

4.1.1. SNM case

In order to obtain overall novelty scores that are comparable among complete and incomplete ideas, we suggest to adopt Equation 6:

$$M'_{SNM} = M_{SNM} \frac{1}{\sum_{i=1}^p f_i} \quad (6)$$

Where p is the number of attributes actually involved, while the other terms are the same used in Equation 2. Note that the numerator is the sum of the normalized weights of all attributes, i.e. equal to 1. Therefore, in presence of a set of ideas satisfying all attributes, M'_{SNM} and M_{SNM} lead to the same novelty scores.

As can be observed in Table 3, Idea 1 (partial scores from Table 2) now reached the same overall novelty score of Idea 3. It is correct (in our opinion), because even if the first idea satisfies only part of the functional requirements related to the key attributes, the novelty scores assigned to the related solutions are the same of all of those involved in Idea 3. Therefore, from a pure novelty point of view, Idea 1 and Idea 3 are equivalent.

Table 3. Comparison between the results calculated by Equation 2 and Equation 6, referring to the S-scores listed in Table 2

	M_{SNM}	M'_{SNM}
Idea 1	7,50	10
Idea 2	5	5
Idea 3	10	10
Idea 4	8,25	9,17

However, besides the novelty score it is important to keep track of the different quality aspects deriving from the presence of missing attributes, and then, it is necessary to associate novelty scores with quality

scores from quality metrics capable to take the recalled aspects into consideration (e.g. the number of satisfied attributes).

Note that in ideation effectiveness measurements, to date, novelty and quality metrics have been always considered as independent each other, while we are now suggesting that in order to properly manage missing attributes, the two metrics are somehow interrelated.

4.1.2. GT case

Similarly to what suggested for the SNM case, here we suggest to divide the novelty score for the sum of the normalized weights of the items and the functions actually involved. More precisely, we suggest to calculate the overall novelty of ideas, by means of Equation 7:

$$M'_{GT} = M_{GT} \frac{1}{\sum_{i=1}^q f_i \sum_{j=1}^r w_{ij}} \quad (7)$$

Where q is the number of functions actually involved, r is the number of items actually involved for a given function i , and the other terms are the same used in Equation 5. Note that the numerator is the sum of the normalized weights of all functions and items, i.e. equal to 1. Therefore, in presence of a set of ideas implementing all functions and representing all items, M'_{GT} and M_{GT} lead to the same novelty scores.

As can be observed in Table 4, Idea 1 (S-scores from Table 1) now reached the same novelty score of Idea 3. As previously explained for the SNM case, this is (in our opinion) correct, because strictly in terms of novelty, the two ideas are equivalent.

Table 4. Comparison between the results calculated by Equation 5 and Equation 7, referring to the S score listed in Table 1

	M_{GT}	M'_{GT}
Idea 1	5	10
Idea 2	5	5
Idea 3	10	10
Idea 4	6,48	8,14

Also in this case it is important to keep track of quality aspects by associating novelty scores to specific quality ones. However, here the quality metric to be considered for a comprehensive ideation effectiveness measurement, should be capable to take into consideration aspects related to both missing functions and missing items. However, the definition of a comprehensive quality metric is out of the scope of this paper.

4.2. Suggestions for methods assessment

According to what stated in Section 3.4.2, both HV and VH methods can be used for calculating the mean novelty score of a set of ideas. More precisely, according to SNM, the GT version considered here, and the proposed Equations 6 and 7, it would be possible to calculate three different values for both SNM and GT cases, i.e. 2 HV + 1 VH (see Table 5).

Table 5. Different mean novelty scores calculated from the data listed in Table 1 (for the GT case) and Table 2 (for the SNM case)

	Horizontal-Vertical (on M_{SNM} and M_{GT})	Horizontal-Vertical (on M'_{SNM} and M'_{GT})	Vertical-Horizontal
SNM (Table 2)	7,69	8,54	8,31
GT (Table 1)	6,62	8,29	7,88

As shown in Table 5, the mean novelty scores calculated with the three different approaches are characterized by some differences, which depend on the considered case (SNM vs GT), the percentage of missing attributes, and the considered method for the calculation of the mean values (VH vs HV).

By only observing the recalled values, it is not possible to identify the best choice. Nevertheless, some considerations can be made, to argue about the calculation methods.

Indeed, the HV approach based on M_{SNM} and M_{GT} leads to values that are strongly affected by the problems highlighted for single idea assessment, while the HV version based on M'_{SNM} and M'_{GT} actually overcomes the recalled problems. The differences between the latter approach and the VH method resides on the number of missing attributes. For example, by adding S-scores to the missing attributes/functions/items of Idea 1 on both Table 1 and 2, it is possible to observe that if the added values are equal to 10 (i.e. equal to the other S-scores of Idea 1), M'_{SNM} and M'_{GT} of the same idea remain unchanged. Therefore, also the related HV method applied on the updated versions of Table 1 and 2 leads to unchanged novelty scores. Differently, the addition of the previously mentioned values on the empty cells of Idea 1 (both in the Table 1 and Table 2), leads to a modification of the mean novelty scores calculated by the VH approach. The recalled differences can be observed by comparing the results shown in Table 5 with those shown in Table 6.

Table 6. Different mean novelty scores calculated on the modified versions of Table 1 and Table 2 (score=10 added to the empty cells of Idea 1)

	Horizontal-Vertical (on M_{SNM} and M_{GT})	Horizontal-Vertical (on M'_{SNM} and M'_{GT})	Vertical-Horizontal
SNM	8,31	8,54	8,52
GT	7,87	8,29	8,30

Nevertheless, that reported above is only a particular example, where the presence or absence of scores related to some attributes actually does not modify M'_{SNM} and M'_{GT} and the related mean values calculated by the HV method. However, in a general case, where the added values actually modify the M'_{SNM} and M'_{GT} of an idea (e.g. by adding a score of 5 to empty cells of Idea 1 on Table 1 and Table 2), the same HV method is sensitive to score changes (see Table 7 vs Tables 6 and 5).

Table 7. Different mean novelty scores calculated on the modified versions of Table 1 and Table 2 (score=5 added to the empty cells of Idea 1)

	Horizontal-Vertical (on M_{SNM} and M_{GT})	Horizontal-Vertical (on M'_{SNM} and M'_{GT})	Vertical-Horizontal
SNM	8	8,23	8,17
GT	7,24	7,66	7,59

Therefore, strictly from the a-posteriori novelty assessment point of view, we suggest to adopt the HV method based on M'_{SNM} and M'_{GT} (depending on the adopted metric), since the calculated mean scores are sensitive to the number of "scored attributes" only when the recalled scores actually impact on novelty of ideas, without implicitly considering quality aspects.

Nevertheless, also for the assessment of design methods and ideation approaches, it is necessary to keep track of the quality issues related to missing attributes. Therefore, depending on the selected novelty assessment approach, also in this case it is necessary to identify an appropriate quality metric.

5. Conclusions

In our experimental activities we applied a variant of the a-posteriori novelty metric originally proposed by Shah et al. (2003), and we faced some issues when assessing set of ideas with "incomplete" entries, i.e. ideas implementing only part of the required functions. More precisely, we initially observed that in presence of design tasks characterized by multiple functions, if some assessed ideas neglect one or more of the required functionalities, the novelty assessment might lead to incorrect results. By generalizing the problem, we noticed that the same issues exist also for the a-posteriori novelty metric originally proposed by Shah et al. (2003), which seems unreliable if applied on a set of ideas with missing attributes.

The objective of this paper was to argue about the recalled problems and to propose plausible suggestions to overcome them for the assessment of both single ideas and group of ideas. Accordingly, we presented some generic a-posteriori novelty assessment examples for the original metric of Shah et al. (2003) and for an appositely conceived variant, capable to exploit genealogy trees for the assessment of conceptual design outcomes with multiple functions. The same examples have been used to show the differences between different possible ways to overcome the problem, highlighting that the absence of a standardized assessment approach could potentially lead to unreliable evaluations, and consequently, to unreliable experimental activities. Moreover, we also observed that for some trivial solutions to the identified applicability problems, the assessment results are somehow implicitly affected by quality-related considerations. Therefore, we have suggested two calculation methods for both single ideas and group of ideas (the latter generally used for evaluating idea generation approaches), which allows to separate pure novelty assessments from quality-related ones. However, while the assessments of quality and novelty should be performed independently, we also observed that depending on the selected novelty metric, it is important to adopt other metrics capable to keep track of the quality aspects related to missing attributes or functions. Consequently, the work presented in this paper has multiple implications. Indeed, the suggestions presented here can be used in future as "standard" ways to assess novelty, independently on the number of attributes, functions and independently on how many of them have been implemented. Moreover, the observation made about the need to associate the proposed novelty metrics to specific quality ones, lead to important research hints about the definition of comprehensive quality metrics also in relation to novelty aspects, fostering the definition of more comprehensive ideation effectiveness measurement approaches.

References

- Berthelsdorf, F.A., Arlitt, R., Armstrong, J., Wilson, A. and Stone, R. (2016), "Creative Personalities: How Personality Influences Creativity of Concept Ideation Methods", *Proceedings of The 4th International Conference on Design Creativity, Atlanta, USA*, The Design Society, pp. 1–8.
- 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>
- Choo, P.K., Lou, Z.N., Camburn, B.A., Wood, K.L., Koo, B. and Grey, F. (2014), "Ideation Methods: A First Study on Measured Outcomes with Personality Type", *Proceedings of the ASME 2014 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference IDETC/CIE 2014, Buffalo, New York, USA, August 17-20, 2014*, ASME, pp. V007T07A019. <https://doi.org/10.1115/DETC2014-34954>
- 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>
- 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>

- 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>
- Oman, S.K., Tumer, I.Y., Wood, K. and Seepersad, C. (2013), "A comparison of creativity and innovation metrics and sample validation through in-class design projects", *Research in Engineering Design*, Vol. 24 No. 1, pp. 65–92. <https://doi.org/10.1007/s00163-012-0138-9>
- 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. (2008), "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. (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., Smith, S.M. and Vargas-Hernandez, N. (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](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>
- 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>
- 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.-Ing. Lorenzo Fiorineschi, Research Assistant
 University of Florence, Industrial Engineering
 via santa marta 3, 50139 Firenze, Italy
 Email: lorenzo.fiorineschi@unifi.it