AUTHOR QUERY FORM

SOCIETY FOR DESIGN AND PROCESS SCIENCE S D P S	Trans. SDPS: J. Integr. Des. Process Sci. Article Number: JIDPS-[No. 2021-10-01-1999-00000745_R]	IOS Press

Dear Author(s),

The queries associated with your article for proof are listed below. Please answer all of these queries and provide any corrections by annotating this PDF before returning it to the editorial office. The latest version of Adobe Reader is strongly recommended for annotating to avoid any potential conflict. Production and publication of your paper will continue after you return the proofed PDF.

Location in article	Query (Please click the linked query number or text to navigate to the right location in the proof. Add your PDF annotations there.)
Line 21	“is” has been changed to “are”.
Line 40	“searching suitable solutions” has been changed to “searching for suitable solutions”.
Line 60	“their effectiveness captured” has been changed to “their effectiveness has captured”.
Line 62	“bot” has been changed to “both”.
Line 111	“is” has been changed to “are”.
Line 126, 639, 680	“A first” has been changed to “The first”.
Line 171	“define” has been changed to “defines”.
Line 187	“,” has been changed to “;”.
Line 255	“an” has been deleted.
Line 255, 256	“3” has been changed to “three”.
Line 317	“repeated” has been changed to “was repeated”.
Line 344	“provide” has been changed to “provides”.
Line 409	“already commented” has been changed to “has already commented”.
Line 414, 527	“requirement” has been changed to “requirements”.
Line 424	“Table” has been changed to “Tables”.
Line 567	“payed” has been changed to “paid”.
Line 617	“A further development” has been changed to “Further development”.
Line 642	“outcome” has been changed to “outcomes”.

Heading	All of the *heading 2 and 3* have not been capitalized the first letter for the word.
Figure, Table	All of the *Figure and Table* forms have been capitalized the first letter for every word.
Key words	Key words have been changed to lowercase.
Page header	Every page header has been changed to the specific title.
Reference	All of the *Reference* forms have been added hyperlink and colored with Bule. All of the *Reference* forms have been capitalized the first letter for the word.

Please return your proof to: JIDPS editorial system within 48 hours. Thank you for your cooperation.

Transactions of the SDPS:
Journal of Integrated Design and Process Science
XX (XXXX) XX-XX
DOI 10.3233/jid-210015
http://www.sdpsnet.org

Effectiveness of different requirements checklists for novice designers

Stefania Altavilla[a], Niccolò Becattini[b*], Lorenzo Fiorineschi[c] and Federico Rotini[c]

[a] *Faculty of Science and Technology, Free University of Bozen-Bolzano, Bolzano, Italy*
[b] *Department of Mechanical Engineering, Politecnico di Milano, Milano, Italy*
[c] *Department of Industrial Engineering, Università di Firenze, Firenze, Italy*

Abstract Working under constrained conditions can boost or kill creativity, depending on the nature of the constraints (organizational, personal or task-related). However, a design process without clearly identified constraints, which set the project objectives, could lead to inefficiencies and unfruitful iterations. Some of the most acknowledged procedures to support requirement definition are focused on the use of specific checklists. However, notwithstanding the importance of the task, little attention was dedicated to the verification of the effectiveness of these tools. In such a context, the paper presents an investigation aimed at assessing the performance of three checklists that exploit different strategies to elicit requirements. To that purpose, a sample of fifty engineering students was asked to use the checklists to define the requirements for a specific design case. The outcomes of the experiment were assessed according to well-acknowledged effectiveness metrics, i.e. quantity, operationality, validity, non-redundancy, and completeness. The result of the assessment highlights that checklists based on more general questions or abstract stimuli can better support novice designers in making explicit internally felt design constraints that can potentially lead to more innovative design.

Keywords: Conceptual design; design tools; product development; requirements elicitation

1. Introduction

~~Translating customer requirements into technical requirements are~~ typically addressed by the Quality Function Deployment (QFD) method (Akao, 1990), which is well known to support designers to transform customer requirements into technical requirements. It maps Customer Attributes (or requirements) and Engineering Characteristics (technical requirements) for the product. Moreover, QFD allows to rank customer requirements according to the perceived level of importance, and to consider more sophisticated requirements classifications (Kano et al., 1984; Matzler & Hintertuber, 1998). QFD, however, leaves the definition of technical requirements to the designers' talent as it just maps the relationships between customer attributes and technical requirements. In other words, QFD supports the designers in translating external constraints (the brief and the set of customer attributes) into technical requirements, but it does not provide any help in defining internal constraints, which are claimed to push for novelty.

An effective and efficient design process needs a set of properly identified and formalized requirements, as this influences the creativity of the related outcomes (Arrighi, Le Masson, & Weil, 2015; Johnson-Laird, 1988; Finke, 1990; Stokes, 2001), both in terms of novelty and variety of the generated ideas (Worinkeng,

* Corresponding author. Email: niccolo.becattini@polimi.it. Tel: (+39)02-2399 8680.

 Published by IOS Press

Joshi, & Summers, 2015; Herrmann, Goldschmidt and Miron-Spektor, 2018). Accordingly, the most acknowledged engineering design handbooks (e.g. Eder & Hosnedl, 2008; Cross 2000; Pahl et al., 2007; Ullman, 2010; Pugh, 1991; Ulrich & Eppinger, 2012) consider requirements as the technical description of objectives that characterize the design process since the very beginning. They help manage the problem(s) and set the goals for the design, potentially reducing the complexity of choices (due to a limited number of available alternatives to consider for the achievement of a goal). Moreover, requirements constrain the boundaries of the design space that the designer explores for searching for suitable solutions. As said above, an excessive use of constraints can also hinder creativity (Roskes, 2015; Caniëls & Rietzschel, 2015; Peterson et al., 2013) as they might trigger a cognitive overload. In fact, constraints play an active role across the cognition-demanding design activities of analysis, synthesis and evaluation (Cascini, Fantoni, & Montagna, 2013; Evbuomwan, Sivaloganathan, & Jebb, 1996; Fiorineschi, Rotini, & Rissone, 2016; Nikulin, Cascini, Viveros, R., & Barbera, 2014). For what concerns the analysis step, their formalization represents a structured list of objectives to achieve. Besides, requirements also support the synthesis of solutions, since turning them into a design proposal deals with the generative process of ideation (Boden, 2009). Additionally, the generation of new solutions can be fostered when in presence of conflicting requirements (Baldussu, Becattini, & Cascini, 2011; C. M. Eckert, Stacey, Wyatt, & Garthwaite, 2011). Concerning the Evaluation step, requirements provide the fundamentals of the evaluation parameters needed to perform the comparisons among candidate solutions, for evaluating them and selecting the most promising.

However, the design specification is constantly updated and drives the design process across all the stages. On the one hand, this progressive refinement is due to the increasing detail level of the solution as the design process proceeds across its stages (e.g. conceptual, embodiment and detail design). On the other hand, this might also depend on a poor planning of the Fuzzy front-end (Bacciotti, Borgianni, Cascini, & Rotini, 2016). While the former case is predictable and expected, the need to adjust the specification "on the fly" due to poor planning can trigger unexpected consequences (e.g. significant additional expenses).

Checklists for the design specification support designers in the hard task of defining requirements with a list of items to be considered potentially relevant as design objectives, but their effectiveness has captured little attention so far. Unfortunately, a comparison between checklists to provide evidences about their pros and cons is currently lacking. The literature claims that design constraints affect (both positively and negatively) the creative process (Bonnardel & Bouchard, 2017) and that their formalization can help the designers in problem decomposition and management (Caniëls & Rietzschel, 2015). Thus, the number of requirements, their distribution across the different phases of the product life-cycle as well as their completeness are therefore crucial for an effective design process. This leads to the formulation of the following research question:

"Does any difference emerge in the outcomes of a requirements definition task with the use of alternative and/or competing requirements checklists?"

as a properly defined requirements checklist allows the designer to externalize its internally perceived design constraints, have a clear list of goals and objectives to attain and directions/licit moves for the generation of ideas.

To answer this question, the authors investigated three checklists, which appear to be suitable for a benchmarking study, since they have by different structures and principles to elicit requirements from designers and stakeholders.

According to this purpose, section 2 introduces the Requirement Checklists considered for the study and specifies the motivations behind this work. Section 3 clarifies the research method, together with the protocol for the execution of the experiment and the related acquisition of data. Section 4 presents the results of the experimental investigation with reference to acknowledged characteristics for requirements and specifications (Roozenburg & Eekels, 1991). Before the conclusion, the results get discussed with reference to the impact the new findings might have.

2. Background

The Cambridge Dictionary defines the noun "checklist" as "a list of things that you must think about, or that you must remember to do". From this perspective, "requirements checklists" are not exceptions. They aim at supporting designers to leverage their own knowledge about the design task or project they have to address, and defining the conditions that the solution should meet to satisfy needs and goals. However, a checklist is not sufficient to formulate requirements since the designer, typically, must interpret the contents of the checklists and adapt them to the situation at hand. Indeed, not all the items in the checklist should be considered relevant, while some others will probably have to be adjusted to suit the specificity of the needs and the goals behind the design. To this purpose, several requirement checklists are available in the literature.

The following subsections review the current state of the art and describe the checklists considered for this work.

2.1. Current lacks and motivations behind this work

The management of requirements is a topic per se, as there are handbooks specifically tailored to support scholars and practitioners in carrying out activities as requirements elicitation, refinement, analysis, etc. (Dick at al., 2017). So far, the literature presents most of the contributions about requirements, their elicitation and management from the perspective of the Information Technology domain. Despite the importance of this topic in engineering design, its related literature focused most on the efforts for providing new approaches (Jones & Kyoung-Yun, 2015; Mokhov et al., 2016) or improving the existing ones (Shu et al., 2017; Sumesh et al., 2020) for the management or the elicitation of requirements (Brace & Cheutet, 2012). More in general, the requirements engineering literature often fails to describe how the requirements have been identified and formalized or it simply reports the list of requirements used in the study, without providing too many details about the technique used to generate them (e.g. through experts' opinion, as witnessed in Tompkins at al. 2018). Nevertheless, the selection of the elicitation technique/approach is also gathering more and more attention in recent years (e.g. Wellsandt et al., 2014). A recent paper by Horkoff et al. (2018), still from the perspective of software development, considers the issue of selecting the "right" elicitation technique according to four dimensions: the acceptability for the user and the subject (i.e. effort required for elicitation), the perceived satisfaction and its usefulness. However, Carrizo, Ortiz, and Aguirre (2016), in their survey about requirements elicitation techniques claimed that the metrics to choose the elicitation technique are not unique and there is no common way to compare their performance and run a meaningful comparison among them.

Beyond the review of well acknowledged checklists presented by Brace & Cheutet (2012), recently other checklists appeared in literature with the purpose of targeting specific domains, such as the environment-related requirements checklist by Michelin et al. (2015) as well as the more "need-oriented" checklists presented in Becattini & Cascini (2014) or in Brglez and Dolšak (2016). Nevertheless, few studies have previously checked the effectiveness of checklists (e.g. Becattini, Cascini, & Rotini, 2015), but unfortunately, it is not possible to find a comprehensive benchmark for checklists in literature.

Moreover, the checklist can present different items and/or use different modalities to stimulate the designer towards the identification of requirements. Accordingly, the checklists considered in this paper (described in the following paragraphs) present such intrinsic differences, thus allowing for purposeful comparisons and providing evidence to answer the research question.

2.2. Checklists selected for this work

2.2.1. The Pahl and Beitz's checklist (PBCL)

Pahl et al. (2007) suggest two different requirements checklists, which can be used in different phases of the design process. The first checklist concerns the elicitation of the information to support the activities involved in the conceptual design phase, while the other checklist mainly focuses on the elicitation of specific criteria to support assessment and selection of solutions during the embodiment and detailed design phases. For the purposes of this work, the authors considered the checklist that Pahl and Beitz proposed for conceptual design (hereinafter called PBCL). The reason behind such a choice is that the role of the requirements is relevant especially in the early phases of the design process, as a poor definition of the specification might trigger several costly design iterations. This led the authors to focus on the conceptual design stage of the process. PBCL guides the exploration of requirements through the administration of a set of stimuli to the user. The stimuli cover different categories of product features, which can be briefly summarised as follows:

- functional performance: flows of force/energy, material and signal/information
- life cycle issues: assembly, transportation, operation maintenance and end-life
- human factors: safety and ergonomics
- specific features of the system: geometry and kinematics
- quality: regulations, standards and testing
- costs and schedules: investments, costs, planning and controls of the development process, time for the development.

With the aim of providing an idea about the formulation of stimuli belonging to PBCL, two examples are presented in the following for the category "human factors" (Pahl et al., 2007):

"Safety: Direct safety systems, operational and environmental safety."

"Ergonomics: Man-machine relationship, type of operation, operating height, clarity of layout, sitting comfort, lighting, shape compatibility.

As shown, the stimuli belonging to each category are lists of examples related to aspects, performance, features, and parameters of the system, presented in a general form, and textually described, which might result relevant under particular conditions or situations.

2.2.2. The Pugh's checklist (PCL)

The checklist proposed by Pugh (Pugh, 1991), hereinafter called PCL, still considers the main categories of requirements introduced in PBCL, although it is more detailed, especially concerning life cycle issues. PCL relies on questions as triggers to elicit requirements. To provide an example, the stimuli for safety and ergonomics suggested by PCL are presented as follows:

"Safety: Should any special facilities be provided for the safety of users and non-users?"

"Ergonomics: Which requirements, with regard to perceiving, understanding, using, handling, etc. does the product have to meet?"

Therefore, PCL works with a different logic if compared to PBCL since it proposes a set of specialised questions to guide the user towards the definition of the relevant requirements the system should satisfy.

2.2.3. A third checklist for the comparison (BCL)

Eventually, the third instrument considered for the comparison is the checklist suggested by Becattini & Cascini (2013), hereinafter called BCL. It is based on textual stimuli that are organised according to the terms of Ideality in TRIZ (Altshuller, 1984; Gadd, 2011; Salamatov, 1999). More in particular, according to the "law of Ideality increase" suggested by TRIZ, which states that systems evolve by increasing the delivered useful functions and by reducing generated harmful effects and consumption of resources. Therefore, the exploration of requirements according to the perspective suggested by Ideality allows the user to take into consideration future desired features, potentially relevant for system and stakeholders. The above-introduced three categories contain sub-classes that refer to specific aspects of the system at different levels, in different phases of the life cycle and for different stakeholders. The stimulus provided to the user is the textual description that defines the sub-class it belongs to. For instance, still considering issues related to safety and ergonomics, these stimuli belong to an abstract class of side effects directly due to the technical system itself:

- Side effects due to the action of the technical system:
  - Ex1: Production scraps (e.g. process waste, amount of materials to be reprocessed, etc.)

  - Ex2: Environment pollution (e.g. heat dissipation, noise level, effects of chemicals on the eco-system, etc.)
  - Ex3: Comfort, ergonomics, safety (e.g. standing/seated operator, grabbing force, # of required movements to carry out an operation, etc.)
  - Ex4: Reliability (e.g. expected mean time between failures, failure rate, etc.)
  - Ex5: … (any other issue related to side effects generated by the system)

It is quite evident that the BCL's classification of requirements strongly differs from those adopted by the other checklists, since it classifies the requirements according to a perspective explicitly based on the evolution of needs, related system features, and functional role played by the latter in satisfying the stakeholders.

2.2.4. Characterization of the main differences among the selected checklists

The three checklists presented in this section differ by their stimulation strategies; their opposed features are depicted in representative quadrants (Figure 1). In particular, the considered metrics can be discerned in terms of the specific way to provide the stimuli (Examples vs Questions in Figure 1), and in terms of abstraction of the provided stimuli (Abstract vs Context specific in Figure 1). The comparison of the considered metrics is expected to clarify what kind of strategy better stimulates the identification and formalization of requirements.

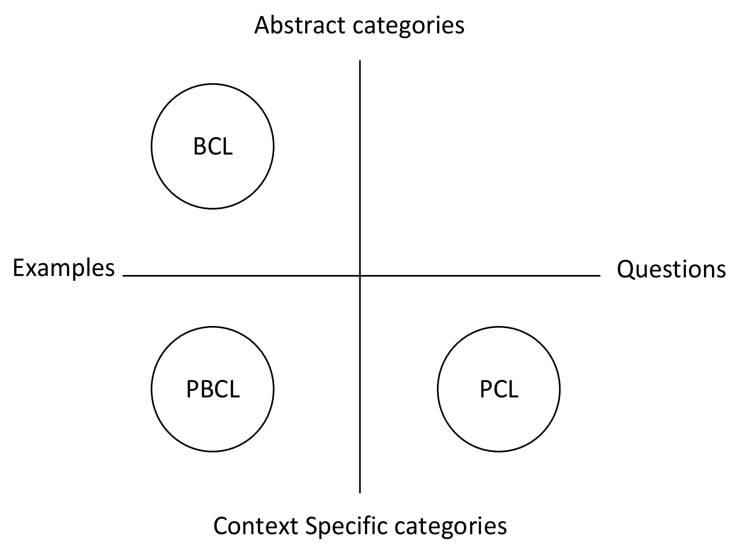


Figure 1. Checklists Classified by Characteristic of the Provided Stimuli to Generate Requirements.

3. Research Method

The experimental set-up can be schematically represented as shown in Figure 2, where the task requires novice designers to translate a set of ten product attributes into a more comprehensive set of engineering requirements.

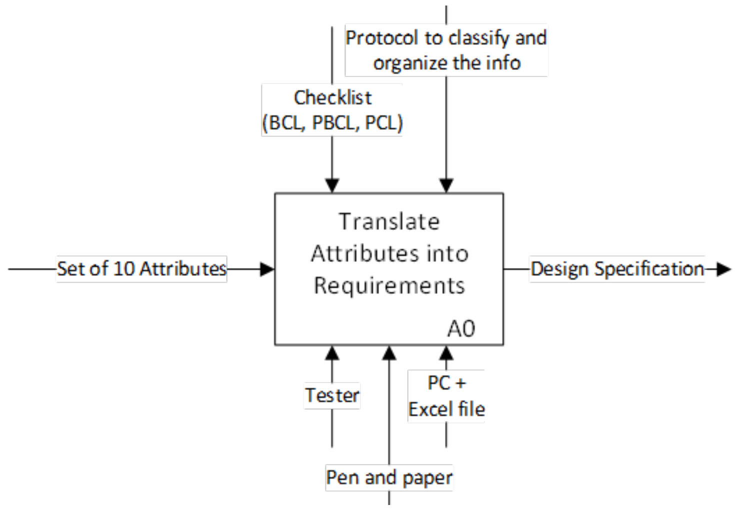


Figure 2. IDEF0 Diagram of the Experimental Set-Up (left arrow: input; right arrow: output; downwards arrows: process controlling elements; upwards arrows: mechanisms for process execution)

The detailed description of the experimental set up and the analysis process is reported in the following subsections.

3.1. Sample of participants involved in the experiment

Fifty students of the Master of Science (MS) degree in Mechanical Engineering (University of Florence, Italy) composed the sample considered for the experiment (Testers in Figure 2). They were all from Italy and only three of them were female. They attended the first part of a course that provides a framework and a set of creative tools for the analysis of stakeholders' needs, and the definition of new benefits/advantages the product to innovate should provide.

A specific definition of function was provided to students, i.e. that based on the Energy-Material-Signal (EMS) flows (Pahl et al., 2007), since it is one of the most largely taught and diffused. It is worth noticing that the concept of function can be interpreted in many ways (Eckert at al., 2011; Eckert, 2013; Vermaas & Eckert, 2013) and therefore a common framework to define functions is beneficial, at least to support the identification of functional requirements (target shared by all the three checklists, without exceptions).

Eventually, a short briefing (thirty minutes overall) was performed with students before the experiment to explain the checklists' logic and their use. Then, the sample was subdivided into three groups (one for each checklist – according to the left downward arrow of Figure 2): 18 students worked with the PCL, 18 with the PBCL and 14 with the BCL (one female for each group). The numerical differences between groups are due to the room setup, as the administration of checklists was done to avoid cross-contamination among participants. More specifically, students were asked to work individually to avoid the mixing of individual thinking and to extract a greater amount of data for the subsequent analysis process. The test requires no control group as the aim is to compare different checklists. Previous studies have already verified the benefits of using a checklist against no support (N. Becattini & Cascini, 2014).

3.2. Design task

The experiment was structured by relying on a particular academic case study: "a device for teeth and mouth hygiene (e.g. an innovative toothbrush)" which also holds the attributes of Table 1, i.e., a set of given design constraints that are represented as input arrow in Figure 2.

Table 1. Initial Set of Desired Attributes

Attribute	Description of the objectives
Hygienic aspects	Performances about this attribute should be comparable to those of existing products of the same type. Nevertheless, the system to be designed should comply at least with standard safety requirements, in order to avoid problems in the oral cavity.
Comfort	No particular performances are expected in terms of comfort.
Aesthetic pleasantness	The ideal solution should be pleasant and perfectly integrated in the environment where it is normally hosted.
Versatility of use	It is expected the possibility to perform multiple cleaning operations within the oral cavity. Besides the teeth cleaning, tongue, palate, and gingival interstices should be considered.
Cleaning effectiveness	The teeth cleaning effectiveness must be maximised.
Ease of use	The system should be as easy as possible.
Multiple functions	Besides the cleaning functionalities, the system should provide other functionality types. In particular, the system should allow to listen to music and/or daily news.
Customisation	The system should allow to be configured according to the user preferences.
Size	The system can be bigger than existing products with similar functionalities.
Energy saving	No particular restrictions are provided in terms of energy consumption.

Students were asked to use the checklists as a guide to extract and/or find the design information needed for the engineering development of the product. Since no additional data were provided, students could extrapolate whatever they wanted in terms of additional data to formulate engineering requirements (represented in Figure 2 by the right arrow). For example, for the attribute "Size" (Table 1), the objective reports that the system can be bigger than the existing ones, but without explicit limits for the maximum allowable size. It is a choice of the student (conditioned by the specific checklist) to establish and indicate missing information (i.e., to externalize additional constraints not otherwise made explicit).

3.3. Testing protocol

The material administered for the experiment consisted in a paper sheet containing the list of attributes reported in Table 1, another paper sheet with a short description of the parameters composing the checklists, and a spreadsheet file containing a structured matrix (Figure 3). Accordingly, students were asked to use their own laptop to list their requirements on the spreadsheet (central and right upwards arrow in Figure 2).

	Main function 1			…			Main function M		
Checklist parameter 1	R 1.1.1	..	R 1.1.j	…	…	…	R M.1.1	..	R M.1.j
Checklist parameter 2	R 1.2.1	..	R 1.2.j	…	…	…	R M.2.1	..	R M.2.j
…	…								
Checklist parameter N	R 1.N.1	..	R 1.N.j	…	…	…	R M.N.1	..	R M.N.j

Figure 3. Generalized Version of the Matrix (N checklist parameters and M functions) Administered to Students for the Experiment (Fiorineschi, Becattini, Borgianni, & Rotini, 2020). The Number of Column Under Each Function Is Not Fixed and Can Vary By Subject. Empty Boxes Are Allowed When It Is Not Possible to Extract/Infer Relevant Requirements.

The structured matrix was provided to support students (novices in requirements definition) in performing the required task.

The time allotted for the test was 60 minutes and students were asked to save the spreadsheet file and send it to a specific email address for data collection.

3.4. Data analysis

3.4.1. Metrics for requirements evaluation

As the goal is to compare the results of the application of different requirements checklists, the results require appropriate evaluation metrics. Roozenburg and Eekels (1991) defined the three ideal characteristics of a design specification, together with the three main elements a requirement should comply with, to be included into the requirement list. They are presented here in the following, in the same order proposed by Roozenburg and Eekels (1991):

- Validity - as the capability of the requirement to discriminate the extent of the achievement of a certain objective.
- Completeness - as the capability of the whole specification to cover all the objectives in the different domains where stakeholders are involved.
- Operationality - as the capability of the requirement to make the objective measurable, to avoid subjective evaluation of (partial) solutions.
- Non-redundancy - as the capability of the specification to be free from duplicates.
- Conciseness - as the capability of the specification to contain just the meaningful requirements, without neglecting important facets to be taken into account (not too many, not too few).
- Practicability - as the capability of the requirements to be tested, e.g. with simulations or by exploiting available information.

To assess Validity and Operationality the authors relied on a coding scheme based on the ENV model (Cavallucci & Khomenko, 2005). The ENV model describes parameters of entities by clarifying the Element they belong to, their Name and the Value it takes. For instance, the sentence “A tomato is a round and red vegetable” describes two parameters: “Tomato” is the element (E), whose considered parameters are (N) colour and shape. They assume, respectively, the values (V): red and round. As Validity describes the requirement to satisfy/target to achieve, a requirement is valid just if it defines both the name of the (measurable) parameter to achieve (N) and the element (E) it belongs to. Operationality requires verifying the achievement of a target value (V) for the above parameter. A requirement is operational just if it clarifies what value to measure otherwise it is not.

Non-redundancy concerns with the exclusion of duplicates from the specification: redundant requirements (in each specific set) are those targeting the same E-N-V triad.

For the Completeness metric, the stakeholders involved in the life-cycle of the specific product have been identified with an “a-posteriori approach” on the entire set of requirements generated by all the subjects (see Table 2). Accordingly, a (more) complete specification involves a larger set of considered stakeholders.

The evaluation of conciseness becomes impossible to define with an a-priori logic, being it dependent on the direction of development a requirement holds in itself, as a piece of designer’s externalized knowledge. On the other hand, the evaluation of practicability depends on the reference simulation or testing system considered for such a purpose. The arbitrariness of these two characteristics also makes them potential biasing factors. For this reason, Conciseness and Practicability have been neglected from the final metrics.

Subjects were not aware of the metrics to apply on the results to avoid biased outcomes.

Table 2. The Stakeholders Identified from the Entire Set of Requirements Generated by Students

Stakeholder	Short description	Phase
Seller	The person who handle the product until it is out of the store	Sale
Buyer	The person who is interested in buying the product and that operates the selection	
User	The person handling the product from the first moment after the purchase, up to its retirement (except for maintenance intervals)	Use/Benefit
Beneficiary	The person receiving the benefits provided by the product (non-necessarily the same person of the user).	
Dentist	The person that is indirectly affected by the benefits provided by the product.	Other
Transporter	The person that transport the product	Transport
Mantainer	The person handling the product during the maintenance operations	Maintenance
Disposal guy	The person handling the product during the disposal operations	Disposal

3.4.2. Data collection and management

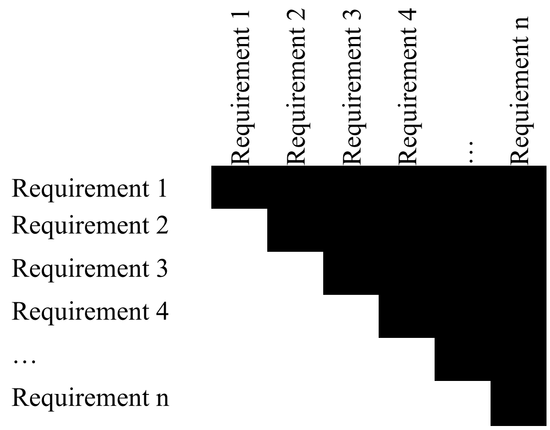
In order to manage data, spreadsheets were collected by group, so that the results can be classified by the subject participating in the data collection process (Figure 2 – right downwards arrow). In each students' spreadsheet, each requirement has been analysed to verify the presence of the three parameters of the ENV triad. Moreover, for each requirement, the affected stakeholders (Table 2) were identified by means of additional columns in the same worksheet (Table 3).

Concerning the Non-redundancy metric, each worksheet (one for each student) enabled this assessment with a specific matrix (see Table 4).

Table 3. Table Used to Assess the Requirement Sets Produced by Each Student. "1" Or "0" Are Attributed to Each of the ENV if Respectively the Parameters Are Present or Are Missing. "1" Is Assigned to Each Stakeholder Actually Affected by the Requirement.

				Affected stakeholders							
	E	**N**	**V**	**Transporter**	**Seller**	**Buyer**	**User**	**Beneficiary**	**Dentist**	**Maintainer**	**Disposal guy**
Requirement 1	1/0	1/0	1/0	1/0	1/0	1/0	1/0	1/0	1/0	1/0	1/0
Requirement 2	1/0	1/0	1/0	1/0	1/0	1/0	1/0	1/0	1/0	1/0	1/0
…						…					
Requirement n	1/0	1/0	1/0	1/0	1/0	1/0	1/0	1/0	1/0	1/0	1/0

Table 4. Non-Redundancy Assessment Matrix. The Value "1" Is Introduced in Those Boxes (In the Lower Triangle of The Matrix) Where It Has Been Identified as Redundancy Between the Requirement in the Row with the Requirement in the Column.



3.4.3. Experimental data processing

According to the considered metrics, three evaluators coded the specification produced by students. More precisely, each evaluator coded the results of two out of the three considered checklists, so that two different coders assess the results of each of the three treatments. Each evaluator worked with all the considered metrics. Krippendorff's Alpha test (Hayes & Krippendorff, 2007) returns Inter-Rater Reliability score. When Alpha scores below 0,66 for a specific metric, the coding results were shared and discussed, and the specific coding activity was repeated. This iterative process was then repeated until the Alpha scores overcome the threshold value 0,66.

The adopted metrics and the related coding activity enabled the distinction of suitable requirements from those that are not (univocally) interpretable or not measurable. This skimming process started from the whole quantity of items and progressively considered the criteria of validity and operationality (which applies on every requirement). Then, the specifications got also reduced to remove duplicates (non-redundancy). The remaining requirements, per responding subject involved in the experiment, constituted the individually generated design specification. The degree of completeness of each specification followed to the criterion described at the beginning of section 2.4.

All the individually generated tentative specifications, as well as their progressive refinements towards the final set of selected design requirements, constituted the data points of distributions by checklists. These distributions got analysed in terms of descriptive statistic estimators (average and standard deviation) to highlight the performance of each checklist and compare them against each other.

4. Results

The statistical analyses reported in the following paragraphs mainly concern the descriptive statistics and the analysis of variance test (ANOVA). The latter was performed to assess the difference in the number of requirements produced by students using the BCL, PBCL and PCL checklist. SPSS version 18.0 is used to calculate all statistical comparisons and the level of significance was set to p-value < 0.05.

4.1. Overall productivity by subject and checklist

Table 5 shows, per each of the subject involved in the investigation, the number of requirements that populate every tentative specification considered for the application of metrics. These values, then, reflect the overall productivity of each subject.

Table 5 Number of Requirements Generated by Each Subject

Checklist	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Tot.	AVG.	St. dev.
BCL	31	17	28	43	15	8	13	10	12	38	11	11	14	18					269	19.2	11.2
PBCL	13	15	7	6	17	17	20	9	28	10	21	15	17	16	11	11	5	9	247	13.7	5.9
PCL	21	13	29	10	12	15	45	23	10	17	13	11	9	4	11	10	9	17	279	15.5	9.4

Descriptive statistics, reported in Table 5, show that on average BCL provides more populated tentative checklists when requirements are formalised starting from the same set of solution attributes as input for the synthesis of solutions. PCL, then, appears to be more productive than PBCL. Both BCL and PCL also have large variability among the recorded performance, while PBCL provides less performing but more stable results. The boxplots depicted in Figure 4 graphically summarize the distribution of recorded outcomes.

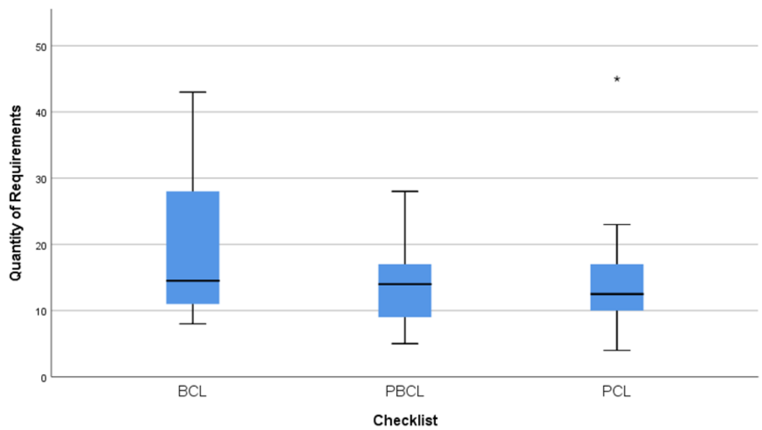


Figure 4. Boxplot of the Number of Requirements Produced by the Different Groups of Students (each of them using a different checklist).

Figure 4 also highlights that there is a generally wider distribution of results in the third and the fourth quartile for BCL, while PBCL and PCL have more narrow distributions and similar interquartile ranges, which should be statistically investigated. Moreover, the largest value for PCL application displayed in the boxplot in Figure 4, is separated from the bulk of the data, and can, therefore, be tested as an outlier. Looking at the interquartile range (IQR) computed from Tukey's hinges (Schwertman, Owens, & Adnan, 2004); the selected value exceeds three times the index, and can be labelled as an outlier, and hence dropped from the analysis. Therefore, the final distribution of the PCL requirements presents a more stable distribution, with a lower standard deviation as reported in Table 6.

On the new set of data, the analysis of variance test (ANOVA) estimates the difference among the BCL, PBCL and PCL checklist. However, no significant difference among the three groups is found.

Table 6 Number of Requirements for the PCL Checklist After the Eliminated Outlier

Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Tot.	AVG.	St. dev.
n° req.	21	13	29	10	12	15	-	23	10	17	13	11	9	4	11	10	9	17	234	13.76	6.10

4.2. Outcomes of the investigation by metrics for requirements and specifications

This subsection provides details about the validity and the operationality of requirements and the completeness and non-redundancy for the whole specification, consistently with the metrics adapted from Roozenburg and Eekels (1991).

4.2.1. *Validity*

Table 7 presents results about the validity of the requirements, based on the same set of data reported in Table 5 and Table 6. In particular, Table 7 shows, for each checklist, the total number of requirements generated by each student (rows "a" of the table) and those remaining in the specification after the check for validity, both in absolute (rows "b" of the table) and percentage terms (rows "c" of the table).

Data in Table 7 mainly confirm the results already commented in subsection 3.1. In fact, both PCL and BCL show higher values, on average, of valid requirements generated by students. Still, the BCL distribution results in the one with the largest value of standard deviation.

Figure 5 presents the normalised results according to the number of subjects exposed to each testing condition with checklists. The results are organised in clusters of requirements spanning four different contiguous design specification sizes (e.g. from 4 to 7 requirements in the specification, from 8 to 11, etc.). To make results homogeneous and fully comparable, they are normalised and scaled to 100% to make a direct estimation of the probability of finding more or less populated specifications within the samples.

There is approximately 30% of chance to get a design specification with 20 requirements or more with BCL, while this probability drops to 5% for both PCL and PBCL. To better check for differences among the three groups, an ANOVA test is performed on the percentage of valid requirements generated by each student. The Tukey post hoc test is also used to detect significant differences in the pairwise comparisons (i.e. BCL vs PBCL, BCL vs PCL and PBCL vs PCL). The level of significance is again set at p-value<0.05. The results are in Tables 8 and 9.

Table 7. The Total Number of Requirements for Each Student (rows "a" of the table) and the Number of Valid Ones in Absolute (rows "b") and Percentage Terms (rows "c").

	Number of requirements for each subject																		Tot.	AVG.	St. dev.
Checklist	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18			
BCL																					
a)	31	17	28	43	15	8	13	10	12	38	11	11	14	18					269	19.2	11.2
b)	25	13	21	34	15	6	9	8	10	29	10	9	8	14					211	15.1	8.8
c)	.8	.8	.8	.8	1	.8	.7	.8	.8	.8	.9	.8	.6	.8					-	0.8	0.1
PBCL																					
a)	13	15	7	6	17	17	20	9	28	10	21	15	17	16	11	11	5	9	247	13.7	5.9
b)	12	8	6	4	11	17	15	8	25	9	14	11	13	8	8	7	4	7	187	10.4	5.2
c)	.9	.5	.9	.7	.6	1	.8	.9	.9	.9	.7	.7	.8	.5	.7	.6	.8	.8	-	0.8	0.2
PCL																					
a)	21	13	29	10	12	15	-	23	10	17	13	11	9	4	11	10	9	17	234	13.8	6.1
b)	19	12	23	9	10	11	-	19	9	15	12	9	8	4	11	6	9	12	198	11.7	4.9
c)	.9	.9	.8	.9	.8	.7	-	.8	.9	.9	.9	.8	.9	1	1	.6	1	.7	-	0.9	0.1

Table 8 Summary of the ANOVA Test Using the Percentage of Valid Requirements (df – degrees of freedom; F – Fisher Snedecor statistics; Sig – Significance)

	Sum of Squares	**df**	**Mean Square**	**F**	**Sig.**
Between Groups	0.093	2	0.046	3.412	0.042
Within Groups	0.626	46	0.014		
Total	0.719	48			

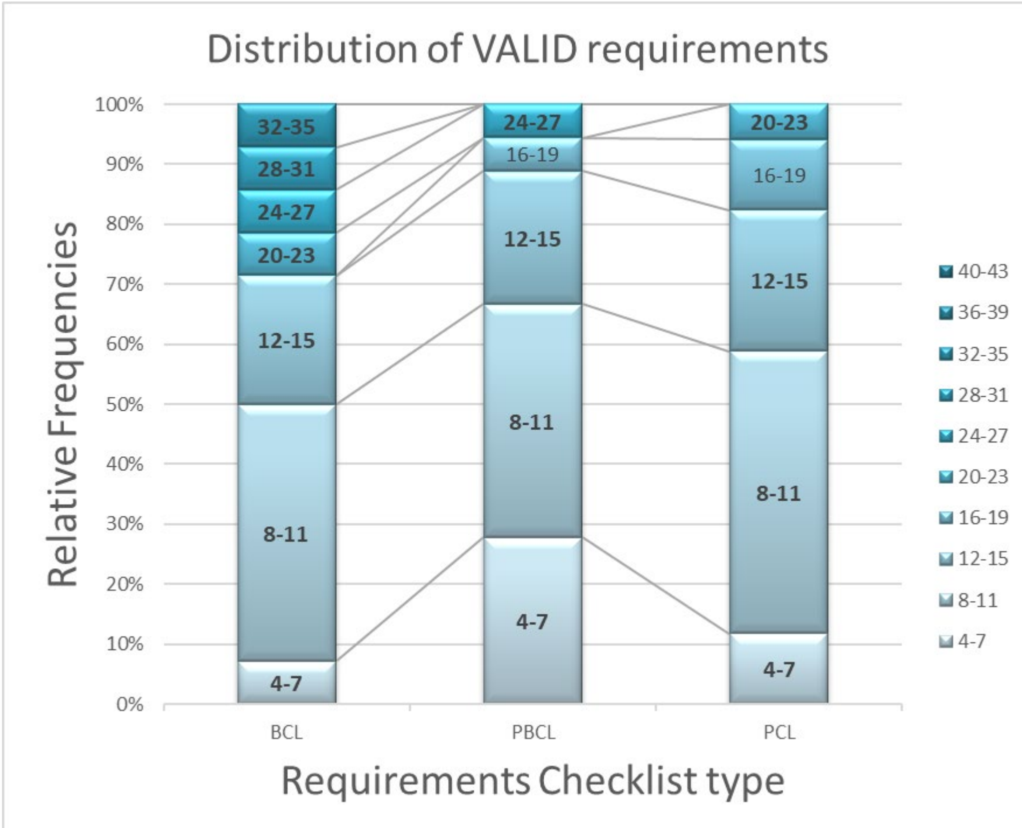


Figure 5. Normalized Results According to the Amount of Subjects Exposed to Each Testing Condition with Checklists

Table 9 Summary of the Pairwise Comparisons Between the Three Checklists Using the Percentage of Valid Requirements (I – rows; J – Columns)

		J				
		BCL	**PBCL**		**PCL**	
I	**BCL**	-	Mean diff. (I-J)	0.028	Mean diff. (I-J)	-0.073
			Std. Error	0.041	Std. Error	0.041
			Sig.	0.781	Sig.	0.069
	PBCL	-	-		Mean diff. (I-J)	-0.10
					Std. Error	0.041
					Sig.	**0.039**
	PCL	-	-		-	

As the results of Table 8 show, the difference in the production of valid requirements concerning the three checklists is significant (F=3.412*, p-value<0.05). According to the results of the Tukey post-hoc test, significant differences exist amongst the group of students using the PCL and the PBCL checklists (p-value=0.039), as reported in Table 9. In particular, it is possible to observe that PCL, on average, produces 10% more valid requirements than PBCL.

The Shapiro-Wilk test verifies the assumption that the data follows a normal distribution (null hypothesis), which is confirmed (p-value 0.197).

4.2.2. Operationality

Table 10 collects the amount of valid and operational requirements, by subject and by checklist, as for Table 7.

Table 10. Amount of Total and Valid and Operational Requirements, by Subject and by Checklist

	Number of requirements for each subject																		Tot.	AVG.	St. dev.
Checklist	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18			
BCL																					
a)	31	17	28	43	15	8	13	10	12	38	11	11	14	18					269	19.2	11.2
b)	25	13	21	34	6	6	9	8	10	29	10	9	8	14					202	14.4	9.1
c)	.8	.8	.8	.8	.4	.8	.7	.8	.8	.8	.9	.8	.6	.8					-	0.7	0.1
PBCL																					
a)	13	15	7	6	17	17	20	9	28	10	21	15	17	16	11	11	5	9	247	13.7	5.9
b)	7	8	4	1	11	10	13	8	25	7	14	8	8	8	8	7	4	7	158	8.8	5.1
c)	.5	.5	.6	.2	.6	.6	.7	.9	.9	.7	.7	.5	.5	.5	.7	.6	.8	.8	-	0.7	0.2
PCL																					
a)	21	13	29	10	12	15	-	23	10	17	13	11	9	4	11	10	9	17	234	13.8	6.1
b)	17	10	23	8	10	11	-	19	8	15	12	9	8	4	10	6	9	11	190	11.2	4.8
c)	.8	.8	.8	.8	.8	.7	-	.8	.8	.9	.9	.8	.9	1	.9	.6	1	.6	-	0.8	0.1

In general, Table 10 confirms what has already commented for the set of valid requirements. Both PCL and the BCL averagely generated a higher percentage of requirements (rows “c” of Table 10), both valid and operational. Comparing Table 10 with Table 7, it is interesting to note that BCL produces almost all valid and operational requirements, having only nine requirements dropped from the analysis (from 211 valid requirements to 202 valid and operational ones). The other two checklists, PBCL and PCL, have very different behaviours: once again PCL performs better than PBCL as the number of dropped requirements is smaller (9 against 29).

Figure 6, consistently with Figure 5, shows normalised data for clusters there represented to clarify the probability mass functions, one per each of the experimental treatments.

In such a case, Figure 6 has clusters with a shift towards lower values, so as to keep track of the whole set of experimental data. Almost 30% of specifications formalised with BCL holds more than 20 requirements, while it is approximately 5% for PCL and PBCL. In this case, as well, ANOVA and a following Tukey post-hoc test investigate the differences among the three groups. Tables 11 and 12 summarize the results.

Table 11 Summary of the ANOVA Test Using the Percentage of Valid and Operational Requirements (df – degrees of freedom; F – Fisher Snedecor statistics; Sig – Significance)

	Sum of Squares	**df**	**Mean Square**	**F**	**Sig.**
Between Groups	0.345	2	0.173	9.068	.000
Within Groups	0.876	46	0.019		
Total	1.221	48			

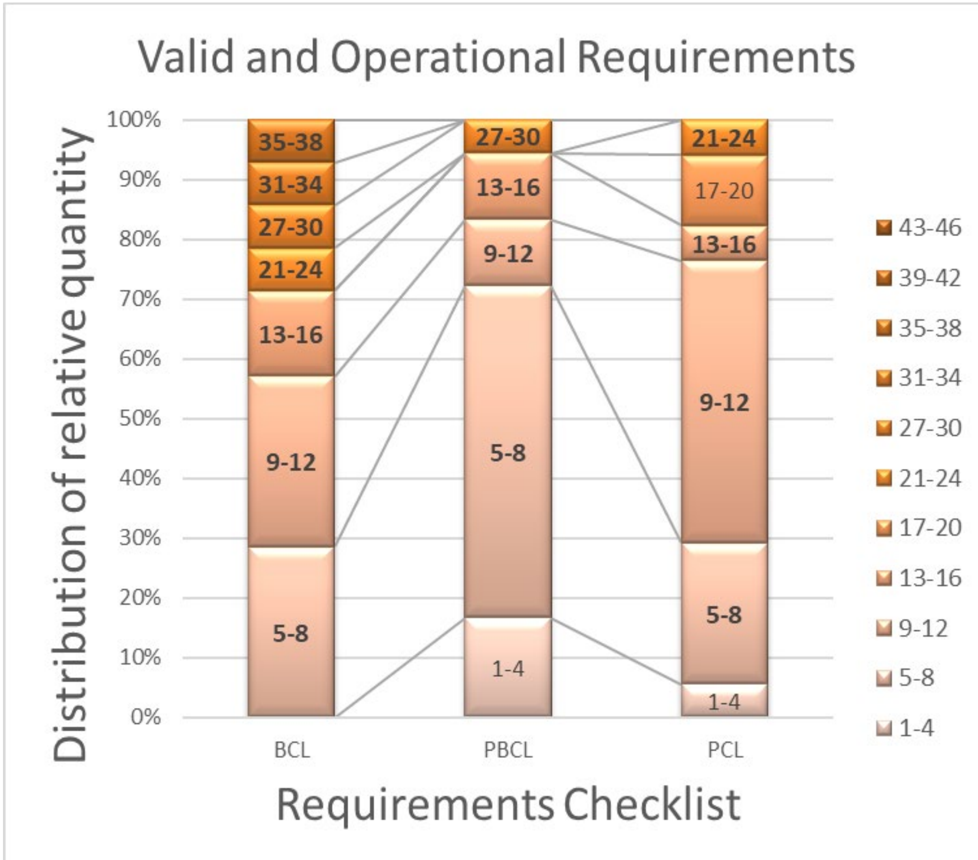


Figure 6. Normalised Data for Clusters There Represented to Clarify the Probability Mass Functions, One Per Each of the Experimental Treatments.

Table 12 Summary of the Pairwise Comparisons Between the Three Checklists Using the Percentage of Valid and Operational Requirements (I – rows; J – columns)

		J				
		BCL	**PBCL**		**PCL**	
I	**BCL**	-	Mean diff. (I-J)	0.17	Mean diff. (I-J)	-0.08
			Std. Error	0.05	Std. Error	0.05
			Sig.	0.06	Sig.	0.189
	PBCL	-	-		Mean diff. (I-J)	-0.20
					Std. Error	0.04
					Sig.	**0.00**
	PCL	-	-		-	

The difference between the PCL, PBCL and BCL groups of generated requirements is significant (F=9.068***, p-value=0.000), as reported in Table 11. Looking at Table 12, it is possible to observe that the statistically significant difference is again between the PCL and the PBCL (p-value= 0.000), where PCL perform better than PBCL in generating both valid and operational requirement of about 20%. As for the previous check, also this case confirms the assumption of the normality in data distribution (p-value>0.05 for the Shapiro-Wilk test).

4.2.3. Non-redundancy

Table 13 shows the results for non-redundancy by subject and requirements checklist considered for this benchmarking study.

Table 13. Amount of Total and Valid, Operational and Non-Redundant Requirements, by Subject and By Checklist.

	Number of requirements for each subject																		Tot.	AVG.	St. dev.
Checklist	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18			
BCL																					
a)	31	17	28	43	15	8	13	10	12	38	11	11	14	18					269	19.2	11.2
b)	16	7	19	7	4	5	8	6	8	28	8	7	7	8					138	9.9	6.6
c)	.5	.4	.7	.2	.3	.6	.6	.6	.7	.7	.7	.6	.5	.4					-	0.5	0.2
PBCL																					
a)	13	15	7	6	17	17	20	9	28	10	21	15	17	16	11	11	5	9	247	13.7	5.9
b)	5	5	4	1	6	4	12	5	6	4	5	5	5	7	5	5	4	3	90	5.0	2.0
c)	.4	.3	.6	.2	.4	.2	.6	.6	.2	.4	.2	.3	.3	.4	.5	.5	.8	.3	-	0.4	0.2
PCL																					
a)	21	13	29	10	12	15	-	23	10	17	13	11	9	4	11	10	9	17	234	13.8	6.1
b)	9	8	21	7	7	8	-	17	8	7	12	8	6	4	10	6	6	6	150	8.8	4.3
c)	.4	.6	.7	.7	.6	.5	-	.7	.8	.4	.9	.7	.7	1	.9	.6	.7	.4	-	0.7	0.2

Based on the descriptive analyses shown in Table 13, BCL and PCL are the checklists that have similar performance and, in percentage, generate more valid, operational and non-redundant requirements (rows "c" of Table 13). The sample of subjects exposed to PBCL provided on average the lowest percentage of valid, operative and non-redundant requirements compared to the total generated by each student. The average, set to specifications having a size of 5 requirements, is less than the number of attributes used as input for requirements identification and formalisation.

Figure 7 presents the distribution of specification sizes, with the same logic used for Figure 6 and Figure 5.

The distributions of probabilities are now substantially different from the ones considered in the above diagrams. The highest threshold in terms of the final size of the specification is achieved using BCL, but more than the 60% of the collected specifications did not reach a size larger than eight requirements.

The ANOVA test, reported in Table 14, still shows a statistically significant difference among the three groups (F=11.608*, p-value=0.000). Looking at the results in Table 15, the difference is significant in the pairwise comparisons between PBCL vs BCL (p-value=0.044) and PBCL vs PCL (p-value=0.000). The results are in line with the descriptive analysis reported in Table 13: there are no significant differences between BCL and PCL (p-value= 0.110), as they generate on average the same percentage of non-redundancy. However, BCL produces on average 15% more valid, operational and non-redundant requirements than PBCL. PBCL, on the other hand, continues to perform worse than PCL, producing on average 28% less non-redundant requirements. The Shapiro-Wilk test confirms the null hypothesis of normality (p-value>0.05).

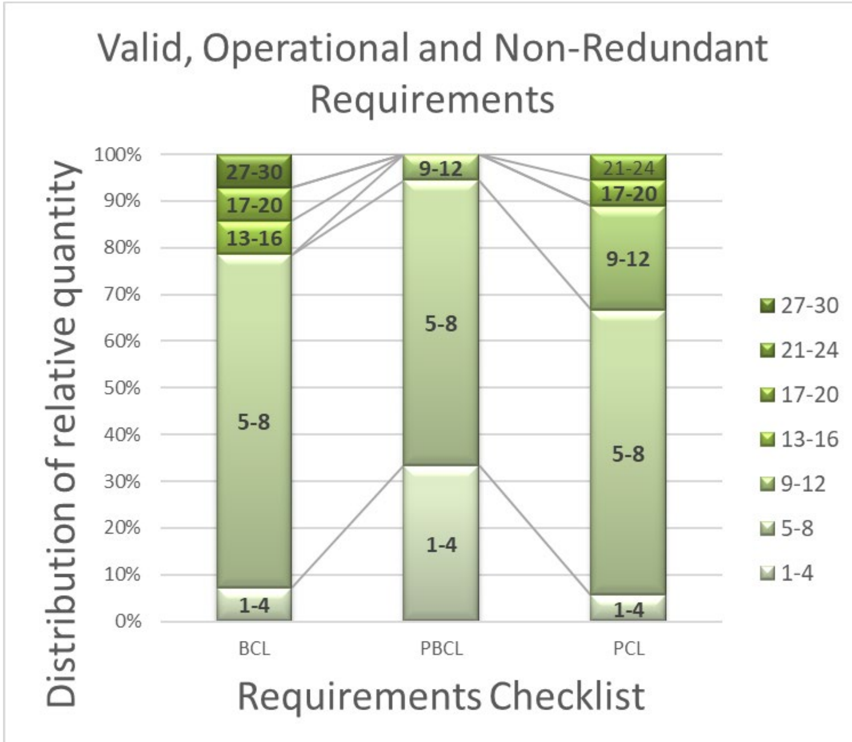


Figure 7. Distribution of Specification Sizes in Clusters of 4 Requirements, with Probabilities Which Are Proportional to Bar Heights and Colours Whose Intensity Grows with The Size of the Design Specification

Table 14 Summary of the ANOVA Test Using the Percentage of Valid, Operational and Non-Redundant Requirements (df – degrees of freedom; F – Fisher Snedecor statistics; Sig – Significance)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	0.666	2	0.333	11.608	0.000
Within Groups	1.320	46	0.029		
Total	1.986	48			

Table 15 Summary of the Pairwise Comparisons Between the Three Checklists Using the Percentage of Valid, Operational And Non-Redundant Requirements (I – rows; J – columns)

		J				
		BCL	**PBCL**		**PCL**	
I	**BCL**	-	Mean diff. (I-J)	0.15*	Mean diff. (I-J)	-0.13
			Std. Error	0.06	Std. Error	0.06
			Sig.	**0.044**	Sig.	0.110
	PBCL	-	-		Mean diff. (I-J)	-0.28***
					Std. Error	0.06
					Sig.	**0.00**
	PCL	-	-		-	

The mean difference is significant at the * p <0.05, ** p <0.01, *** p <0.001

4.2.4. Completeness

The results about the completeness metric follow a requirement-centred perspective, in place of subject-centred. It means that results are still organised by checklist, but they refer to requirements as the collection of all the requirements produced by the whole set of subjects participating in the study.

Figure 8 shows a radar plot of the distribution of requirements, consistently with the class of stakeholders /solution's lifecycle stage they refer to, except for use stage. In fact, the requirements generated with all the three checklists appear to focus particularly on the use stage, i.e. when the solution works to fulfil the need of a beneficiary (which is not necessarily the user of the solution itself). For this reason, the values have been separated in a histogram chart. Despite small differences, all of them trigger the identification of these requirements in 50-60% of cases.

One of the most evident results is that all the three checklists do not trigger any particular stimulus towards the definition of requirements concerning the transportation phase, despite there is an explicit reference to this phase both in PBCL and in PCL. On the other hand, it is also worth noticing that PBCL allows for a wider exploration of requirements classified under the label "other", at least with reference to the whole set of requirements, when compared to the results obtained with BCL and PCL.

The results, then, show that the three checklists almost equally span the remaining classes considered here, with no notable differences except for BCL in the "sale" category.

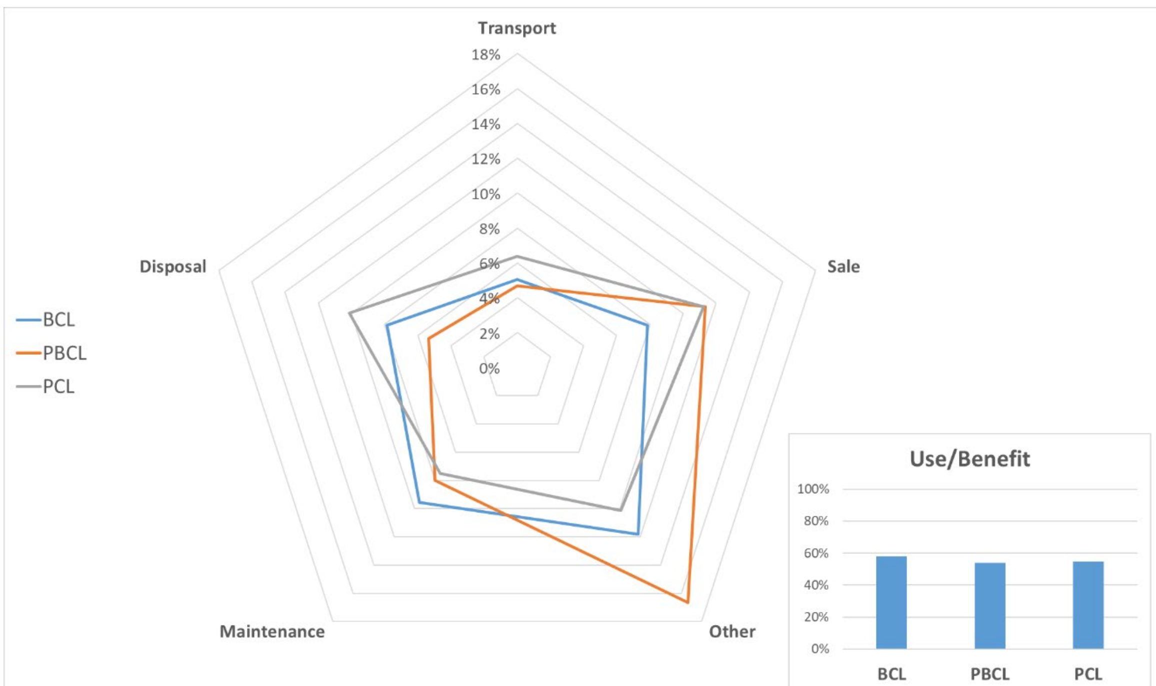


Figure 8. Distribution of Requirements in Terms of Spanned Life Cycle Phases

5. Discussion

5.1. Obtained results

Concerning productivity, the three checklists present some statistically significant differences, highlighted in the summary of the results provided in Table 16 below.

Table 16 Summary of the Pairwise Comparisons

Metrics	Results of the pairwise comparisons
Overall productivity	BCL>PBCL; BCL>PCL; PBCL<PCL
Validity	BCL>PBCL; BCL<PCL; PBCL<PCL*
Validity and Operationality	BCL>PBCL; BCL<PCL; PBCL<PCL***
Validity, Operationality and Non-Redundancy	BCL>PBCL*; BCL<PCL; PBCL<PCL***

The mean difference is significant at the * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Looking at the total amount of generated requirements, the difference between BCL, PBCL and PCL is not significant, according to the high standard deviation of the three distributions. It means that although some of the checklists provide different types of stimuli (see Figure 1), there is a high level of subjectivity in their interpretation and exploitation. Nevertheless, as shown in Figure 4, BCL presents higher potentialities for generating more populated checklists, suggesting that it has a (slightly) stronger effectiveness in stimulating designer's formalization of design constraints.

When applying the filter of "validity", both PCL and BCL show a higher percentage of valid requirements generated by students. According to Figure 5, the three checklists have (approximately) the same probabilities to generate requirements in the clusters 8-11 and 12-15, but PBCL is characterised by a higher probability to obtain the cluster 4-7, together with the absence of any subject capable to obtain more than 27 valid requirements. It implies that PBCL appears to be the less performant in terms of validity, also confirmed by the ANOVA test (see Tables 8 and 9). However, the significant difference is verified only between the PBCL and PCL. Very similar considerations can be made when applying the further filter of "operationality", where PBCL still appears as the less performant (p-value<0.05) compared to PCL. These results can be explained by the extreme conciseness and simplifications characterising PBCL. If this can speed up the process; conversely, it can lead to misleading outcomes, especially if used by inexperienced users (as for the students participating at the experiment). In fact, in the context of this study, the guidance provided by PCL to inexperienced users resulted in a generation of a higher amount of valid and operational requirements, if compared to PBCL. When considering the additional non-redundancy filter, the worst performances of PBCL are even more extreme. The difference of performance resulted statistically significant also in comparison with BCL (see Table 15). Therefore, when dealing with a set of product attributes, PBCL demonstrated to be not particularly effective in the generation of a comprehensive design specification. This is consistent with previous results, which showed that the general productivity of PBCL is not particularly effective when used in conceptual design (Becattini et al., 2015). In general, the three checklists show a sensible and comparable reduction rate. PBCL, then, demonstrated to be the least capable of supporting the designer in the formalization of design constraints. The limited number of requirements in the related design specifications suggests that PBCL is also not particularly effective in supporting the externalization of internally perceived design constraints, despite a specific investigation of differences between internal and external constraints is left for future investigations.

However, BCL and PCL are the most affected by the non-redundancy filter. Indeed, they have similar performances also confirmed by the absence of statistically significant differences in their comparison (see Table 15). This appears to be controversial, as these checklists have opposite characteristics (Figure 1). The results potentially show that there is an undisclosed variable (beyond the human factor) which is affecting the phenomenon and that still needs to be explored.

Resuming, PBCL triggered the definition of 247 items to populate the design specification. Once skimmed, just 90 of these items could be considered requirements by validity, operationality and non-redundancy (thus design task-related constraints). The ratio between the final and the initial set of requirements provides a preliminary, but quantitative, measure of the effectiveness of the checklist. The average efficiency of PBCL, therefore, is approximately 36% (90/247). BCL and PCL, instead, produce a much more comparable result: the efficiency of BCL is approximately 50% (138/269), while for PCL is 64% (150/234). From this perspective, the best performance is achieved by PCL. As most of these efficiency drops depend on the non-redundancy filter, it appears that in different ways these checklists can trigger some fixation in the designer. This kind of fixation could be probably reflected on the solutions generated to address the requirement list; nevertheless, these experimental data just show that the subjects were focusing their attention on the same few problems.

For what concerns the completeness metric, Figure 8 shows that all the three checklists led the students to generate requirements that focus on the same life cycle phase, i.e. the use of the product. This is a further confirmation that the novice designers participating in the study paid particular attention to specific objectives, potentially neglecting some others that are relevant and that facilitate the variety and the novelty of the generated ideas.

The results of this experiments show that unexperienced designers should preferably use PCL, since it has, overall, the best performance score. However, BCL is a potential alternative based on a completely different strategy (see Figure 1), which also shows a good performance. Nonetheless, it is strongly based on a specific definition of a function, which needs to be introduced to users before BCL use. In general, under the limitations summarised in the following section, the investigation has shown that, at least for novice designers, a checklist whose stimulation strategy uses examples and is too much dependent on the context, seems to perform worse than other strategies that exploit sets of more general questions or abstract stimuli. Notwithstanding, the investigated checklists organise the stimuli into categories that take into consideration the several conditions and scenarios under which the system could operate, the lack of completeness appears to be the negative aspect shared by all the considered approaches. Accordingly, Figure 9 summarises the findings above discussed.

Figure 9. Summary of the Performance for The Checklists Considered in The Study, According to The Metrics Defined for The Assessment. The Checklists Are Distinguished by Their Capability to Elicit Requirements as A Podium Ranking (Column: Quantity). An Efficiency Diagram, Described As 4 Bars of Increasing Height for Increasing Performance Displays the Results for Validity, Operationality and Non-Redundancy. The Last Column Shows the General Performance of The Checklists as An Overall Indicator of Efficiency (ratio of "survived" requirements in the specification after skimming).

5.2. Limitations and future developments

A non-negligible limitation of the work is the limited number of checklists that have been experimented, which inevitably reduces the general validity of the considerations presented in the previous section. Widening the sample of tested checklists could strengthen the validity of the results, especially with reference to the categories of strategies shown in Figure 1. Therefore, a first research work for the future should repeat the same study with an enlarged sample of checklists. This can be done by taking into account also contributions coming from other fields like software engineering, which is a sector very sensitive to engineering requirements. This would also complement the viewpoint on tools to support designers in externalizing internally perceived design constraints, which is what checklists aim at doing. The cross-domain comparisons of checklists would also help understanding the differences between different levels of abstraction for design constraints, beyond the subtle differences emerged through the comparison of a checklist based on abstract stimulation (BCL) and one based on concrete examples (PBL).

Another limitation of the work comes from the predefined input delivered to students that introduced a common set of external constraints for all the participants (that could probably be the reasons of the shared focus observed in Figure 8). This need arose to ensure a common vision of the task, but it inevitably led to a condition that can sensibly differ from a real case. In addition, the considered sample of participants (engineering students from the same institution) implies non-negligible limitations. Indeed, besides the effects of their limited expertise (which led to results that could not be applicable to more experienced practitioners), it is also not clear to what extent the results are affected by ethnological aspects. Consequently, it is necessary to perform additional experiments extended to students from different institutions, disciplines and/or countries. This could be useful to obtain statistically significant evaluations, and, then, more robust indications for selecting the most suited checklist, according to the user needs.

Furthermore, besides the generic indications provided in this paper, it is still unclear whether more detailed and statistically significant correlations can be obtained between each metric presented in Section 3 and parameters like the complexity of the task, the expertise of the designer and/or the design team and the type of product. Moreover, it is also unclear whether a specific checklist can be more suited for analysis, synthesis or evaluation purposes. This kind of information could pave the way for the development of criteria to identify the most effective requirement checklist for the design context that supports the related creative design activity.

Further development of the work could be the evaluation of the "quality" of the obtained specifications in terms of possible help or hindrance in exploring the design space, which is crucial to complement the missing step of this research: how a more or less complete design specification stimulates the generation of more creative ideas. Indeed, if on the one hand abstract and generic requirements may be useful in order to avoid undesired fixations (Vasconcelos et al., 2018; Jansson & Smith 1991) on specific designs; on the other hand, they can lead to several design iterations. However, excessively detailed specifications can lead to the opposite effects. The reduction of the efficiency for the whole design process due to fixation or iterations would require attention on the side of organizational (time-related) constraints. Therefore, future studies should perform additional investigations on design outcomes coming from design tasks where the requirements emerge from the use of specific checklists. To this purpose, well known creativity or idea generation effectiveness metrics can be successfully adopted (e.g. Shah at al., 2003; Sarkar & Chakrabarti 2011; Nelson et al., 2009).

Eventually, another future development concerns the usability and the perceived onerousness of the checklists. To this purpose, future experiments could consider the use of simple surveys to be administered to participants, in order to extract the required information. Predefined frameworks like the NASA TLX (Hart, 2006; Sandra & Staveland, 1988) can address this objective.

5.3. Expected impact

The comparison performed in this work led to important indications about what checklist novice designers or students should use to start their design process, so that they have a clear guidance for the analysis, synthesis and evaluation of problems and solutions, despite this can be potentially boosting or undermining their creativity. Additionally, the experimental approach used for the work presented in this paper can be repeated with any design checklist. Moreover, the obtained numerical results might represent the first reference to run comparisons with results coming from further experiments using the same checklists.

Therefore, this work paves the way for future studies on requirement checklists, where one of the potential outcomes is the definition of a framework to start linking the design phases of product planning and conceptual design. To that purpose, the matrix shown in Figure 3 constitutes a preliminary tool proposal to help novices in better exploiting the potentialities of checklists.

Concluding, the experimental approach, the obtained results and the research hints provided in this section are expected to promote new research about how to support designers in defining comprehensive sets of requirements.

6. Conclusion

The paper aims at unveiling potential differences between requirements checklists, as they help the designer to set the initial constraints to steer the whole design process across its cycles of analysis, synthesis and evaluation. The identification and formalization of these design constraints help the designer to decompose the problem and reduce its complexity with a more efficient allocation of cognitive resources. The comparison among checklists for their applicability and related outcomes is here measured in terms of the metrics proposed by Roozenburg and Eekels (1991). More specifically, the metrics adopted in this paper were Validity, Operationality, Non-redundacy and Completeness. The group of 50 MS Mechanical Engineering students that took part in this study worked individually to generate a meaningful tentative design checklist. The outcomes of the experiment showed that there are differences among the three checklists here considered, i.e. Pugh's checklists (PCL), Pahl and Beitz's checklist for conceptual design (PBCL), and a recent design specification checklist as proposed in Becattini & Cascini (2013) (BCL). The benchmark between the checklists is consistent, as all the subjects participating in the study address the same design brief with the same set of proposed product attributes and translate it into a set of engineering requirements.

The obtained numerical results are, to the knowledge of the authors, one of the few quantitative examples of checklist application, despite the paper does not focus on how these checklists can stimulate a creative design process and/or outcomes. Nevertheless, the results showed that PBCL is not particularly effective in the generation of a sufficient comprehensive design specification as they collect a limited set of elements, that are poorly representative of the external and the designer's constraints. In other words, PBCL does not effectively support the formalization of design constraints, thus it probably poorly supports creative behaviour and outcomes for novice designers.

PCL and BCL, on the contrary, showed very similar performance overall, despite they have very different characteristics. PCL presents a set of questions that span the various life cycle stages of the solution to be designed (thus context specific categories). BCL, conversely, proposes a set of more abstract concepts, displayed as examples, which the checklist user has to contextualize to properly define requirements. BCL allowed the generation of a higher number of requirements per person than PCL, but the number of redundant items in the specification align the performance of the two checklists, whose results are, eventually, comparable. From the perspective of completeness, all the checklists appear to be mostly focused on the use stage of the solution lifecycle, while there are opportunities for the combined use of PCL and BCL in order to cover different moments of the solution life cycle and overcome potential lacks that might emerge in case just one of these checklists is used.

The results presented in the paper might represent the first benchmark to run comparisons with results coming from further experiments using the same checklists. Moreover, they could be useful for future activities focused on the development of guidelines for the selection of the most suitable design checklist.

Notwithstanding the above-mentioned results, some limitations might be ascribed to this work. Accordingly, the experimental subjects are students, thus not representative of the actual industrial perspective. Moreover, the initial design brief is quite simple, for a product with a relatively low degree of complexity (toothbrush). Therefore, the applicability of these results to the industrial context is not immediate. This requires additional investigations: with new experiments could be tailored to the specific engineering domain at hand as well as involve subjects with a higher level of expertise. Beyond what suggested here, the authors expect that peers interested in performing additional analysis and/or at validating the results presented in this paper, might reuse the presented research approach and the experimental protocol. Additionally, several hints for further research activities are provided in this paper, among which, that of investigating around the actual support that the obtained requirement lists can provide in terms of design space exploration.

References

Akao, Y. (1990). Quality Function Deployment: Integrating Customer Requirements into Product Design. Cambridge: Productivity Press.

Altshuller, G. S. (1984). Creativity As An Exact Science. Amsterdam: Gordon and Breach Science.

Arrighi, P.-A., Le Masson, P., & Weil, B. (2015). Addressing Constraints Creatively: How New Design Software Helps Solve the Dilemma of Originality and Feasibility. Creativity and Innovation Management, 24(2).

Bacciotti, D., Borgianni, Y., Cascini, G., & Rotini, F. (2016). Product Planning Techniques: Investigating The Differences Between Research Trajectories And Industry Expectations. Research in Engineering Design, 27(4), 367–389. https://doi.org/10.1007/s00163-016-0223-6

Baldussu, A., Becattini, N., & Cascini, G. (2011). Network of Contradictions Analysis And Structured Identification Of Critical Control Parameters. Procedia Engineering, 9(2011), 3–17.

Becattini, N., & Cascini, G. (2014). General-Purpose Requirements Checklist For Improving The Completeness Of A Design Specification. In Proceedings of International Design Conference, DESIGN (Vol. 2014-Janua, pp. 111–120).

Becattini, N, & Cascini, G. (2013). Mapping Causal Relationships And Conflicts Among Design Parameters And System Requirements. Computer-Aided Design and Applications, 10(4), 643–662.

Becattini, N, Cascini, G., & Rotini, F. (2015). Requirements Checklists : Benchmarking the Comprehensiveness of the Design Specification. Proceedings of the 20th International Conference on Engineering Design (ICED 15), Vol. 1: Design for Life, (July), 1–10.

Boden, M. A. (2009). Computer Models of Creativity. AI Magazine, 30(3), 23–34.

Bonnardel, N., & Bouchard, C. (2017). Creativity in Design. In J. C. Kaufman, J. Baer, & V. P. Glăveanu (Eds.), The Cambridge Handbook of Creativity across Domains (pp. 403–427). Cambridge: Cambridge University Press. https://doi.org/DOI: 10.1017/9781316274385.022

Brace, W., & Cheutet, V. (2012). A Framework to Support Requirements Analysis in Engineering Design. Journal of Engineering Design, 23(12), 876–904. https://doi.org/10.1080/09544828.2011.636735

Brglez, Š., & Dolšak, B. (2016). A Necessity-Based Method for Product Requirement Elicitation and Classification. Transactions of FAMENA, 40(1), 87–110.

Caniëls, M. C. J., & Rietzschel, E. F. (2015). Organizing Creativity: Creativity and Innovation under Constraints. Creativity and Innovation Management, 24(2), 184–196. https://doi.org/10.1111/caim.12123

Carrizo, D., Ortiz, C., & Aguirre, L. (2016). What Do Researchers Mean by "The Right Requirements Elicitation Techniques." Ingeniare. Revista Chilena de Ingeniería, 24(2).

Cascini, G., Fantoni, G., & Montagna, F. (2013). Situating Needs and Requirements in the FBS Framework. Design Studies, 34(5), 636–662. https://doi.org/10.1016/j.destud.2012.12.001

Cavallucci, D., & Khomenko, N. (2005). From TRIZ to OTSM-TRIZ : Addressing Complexity Challenges in Inventive Design Denis Cavallucci , Nikolaï Khomenko. Int. J. Product Development, X(X), 1–17.

Cross, N. (2000). Engineering Design Methods Strategies for Product Design. Third Ed. (3rd ed.). Chichester: Wiley, John.

Dick, J., Hull, E., & Jackson, K. (2017). Requirements Engineering. Springer.

Eckert, C. (2013). That Which Is Not Form: The Practical Challenges in Using Functional Concepts in Design. Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 27(03), 217–231. https://doi.org/10.1017/S089006041300022X

Eckert, C., Alink, T., & Ruckpaul, A. (2011). Different Notions of Function : Results From An Experiment on the Analysis of An Existing Product. Journal of Engineering Design, 22(11–12), 811–837. https://doi.org/http://dx.doi.org/10.1080/09544828.2011.603297

Eckert, C. M., Stacey, M., Wyatt, D., & Garthwaite, P. (2011). Change As Little As Possible: Creativity in Design by Modification,. Journal of Engineering Design, 23(4), 337–360. https://doi.org/10.1080/09544828.2011.639299

Eder, W., & Hosnedl, S. (2008). Design Engineering - A Manual for Enhanced Creativity -. CRC Press.

Evbuomwan, N. F. O., Sivaloganathan, S., & Jebb, A. (1996). A survey of design philosophies, models, methods and systems. Journal of Engineering Manufacture, 210, 301–320.

Finke, R. A. (1990). Creative Imagery. Discoveries and Inventions in Visualization. Hillsdale, NJ: Lawrence Erlbaum Associates. https://doi.org/10.1080/03060497.2001.11086090

Fiorineschi, L., Becattini, N., Borgianni, Y., & Rotini, F. (2020). Testing a New Structured Tool for Supporting Requirements ' Formulation and Decomposition. Applied Sciences (Switzerland), 10, 3259.

Fiorineschi, L., Rotini, F., & Rissone, P. (2016). A New Conceptual Design Approach for Overcoming the Flaws of Functional Decomposition And Morphology. Journal of Engineering Design, 27(7), 438–468. https://doi.org/10.1080/09544828.2016.1160275

Gadd, K. (2011). TRIZ For Engineers : Enabling Inventive Problem Solving. Chichester.

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50(9), 904–908. https://doi.org/10.1177/154193120605000909

Hart, Sandra G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. Advances in Psychology, 52(C), 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Hayes, A. F., & Krippendorff, K. (2007). Answering the Call for A Standard Reliability Measure for Coding Data. Communication Methods and Measures, 1(1), 77–89.

Herrmann, M. E., Goldschmidt, G., & Miron-Spektor, E. (2018). The Ins And Outs of the Constraint-Creativity Relationship. In DS 89: Proceedings of The Fifth International Conference on Design Creativity (ICDC 2018) (pp. 160–167).

Horkoff, J., Ersare, J., Kahler, J., & Jörundsson, T.D. Hammouda, I. (2018). Efficiency and Effectiveness of Requirements Elicitation Techniques for Children. In IEEE 26th International Requirements Engineering Conference (pp. 194–204).

Jansson, D. G., & Smith, S. M. (1991). Design Fixation. Design Studies, 12(1), 3–11. https://doi.org/10.1016/0142-694X(91)90003-F

Johnson-Laird, P. (1988). Freedom and Constraint in Creativity. In S. J. Sternberg (Ed.), The Nature of Creativity. New York: Cambridge University press.

Jones, I. A. and Kyoung-Yun, K. (2015). Systematic Service Product Requirement Analysis with Online Customer Review Data. Journal of Integrated Design and Process Science, 19(2), 25-48.

Kano, N., Seraku, N., Takahashi, F., & Tsuji, S. (1984). Attractive Quality and Must-Be Quality. Journal of the Japanese Society for Quality Control, 14(2), 39–48.

Matzler, K., & Hintertuber, H. H. (1998). How to Make Product Development Projects More Successful by Integrating Kano's Model of Customer Satisfaction into Quality Function Deployment. Technovation, 18(1), 25–38.

Michelin, F., Reyes, T., Vallet, F., Eynard, B., & Duong, V. L. (2015). Improving the Management of Environmental Requirements in Clients/Suppliers Co-Design Process. In 20th International Conference on Engineering Design (ICED 15) Vol 1 (pp. 31–42). Milan.

Mokhov, S. A., Song, M., Chilkaka, S., Das, Z., Zhang, J., Llewellyn, J., & Mudur, S. P. (2016). Agile Forward-Reverse Requirements Elicitation As a Creative Design Process: A Case Study of Illimitable Space System V2. Journal of Integrated Design and Process Science, 20(3), 3-37.

Nelson, B. A., Wilson, J. O., Rosen, D., & Yen, J. (2009). Refined Metrics for Measuring Ideation Effectiveness. Design Studies, 30(6), 737–743. https://doi.org/10.1016/j.destud.2009.07.002

Nikulin, C., Cascini, G., Viveros, P., R., S., & Barbera, L. (2014). An Algorithm for Root Cause Analysis Integration Based on OTSM-TRIZ: Complex Problem Analysis. In 22nd Annual Conference on European Safety and Reliability (ESREL 2013) (pp. 643–650). Amsterdam: CRC Press.

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

Peterson, D. R., Barrett, J. D., Hester, K. S., Robledo, I. C., Hougen, D. F., Day, E. A., & Mumford, M. D. (2013). Teaching People to Manage Constraints: Effects on Creative Problem-Solving. Creativity Research Journal, 25(3), 335–347. https://doi.org/10.1080/10400419.2013.813809

Pugh, S. (1991). Total Design. Integrated Methods for Succesfull Product Engineering. Reading, Massachusetts: Addison Wesley Publishing Company.

Roozenburg, & Eekels. (1991). Product Design: Fundamentals and Methods. John Wiley and sons, Inc.

Roskes, M. (2015). Constraints that Help or Hinder Creative Performance:A Motivational Approach. Creativity and Innovation Management, 24(2).

Salamatov, Y. (1999). TRIZ: The Right Solution at the Right Time: A Guide to Innovative Problem Solving. Insytec B.V.

Sarkar, P., & Chakrabarti, A. (2011). Assessing Design Creativity. Design Studies, 32(4), 348–383. https://doi.org/10.1016/j.destud.2011.01.002

Schwertman, N. C., Owens, M., & Adnan, R. (2004). A Simple More General Boxplot Method for Identifying Outliers. Comput. Stat. Data Anal., 47, 165–174.

Shah, J. J., Vargas-Hernandez, N., & Smith, S. M. (2003). Metrics for Measuring Ideation Effectiveness. Design Studies, 24(2), 111–134. https://doi.org/10.1016/S0142-694X(02)00034-0

Shu, M., Tan, S., Fu, L., Zeng, Y., Cao, X., & Zeng, Y. (2017). Application of Design Methodologies to Web System Design: A Case Study of JIDPS Editorial System. Journal of Integrated Design and Process Science, 21(4), 79-112.

Stokes, P. D. (2001). Variability, Constraints, and Creativity: Shedding Light on Claude Monet. American Psychologist, 56(4).

Sumesh, S., & Krishna, A. (2020). Hybrid Analytic Hierarchy Process-Based Quantitative Satisfaction Propagation in Goal-Oriented Requirements Engineering Through Sensitivity Analysis. Multiagent and Grid Systems, 16(4), 433-462.

Tompkins, M., Iammartino, R., & Fossaceca, J. (2018). Multiattribute Framework for Requirements Elicitation in Phased Array Radar Systems. IEEE Transactions on Engineering Management, PP, 1–18. https://doi.org/10.1109/TEM.2018.2878688

Ullman, D. G. (2010). The Mechanical Design Process 4th ed. New York, USA: Mc Graw HIll.

Ulrich, K. T., & Eppinger, S. D. (2012). Product Design and Development (5th ed.). New York: Mc Graw Hill Irwin. https://doi.org/10.1016/B978-0-7506-8985-4.00002-4

Vasconcelos, L. A., Neroni, M. A., & Crilly, N. (2018). The Effect of Explicit Instructions in Idea Generation Studies. Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM, 32(3), 308–320. https://doi.org/10.1017/S0890060417000658

Vermaas, P. E., & Eckert, C. (2013). My Functional Description Is Better! Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 27(03), 187–190. https://doi.org/10.1017/S089006041300019X

Wellsandt, S., Hribernik, K. ., & Thoben, K. D. (2014). Qualitative Comparison of Requirements Elicitation Techniques That Are Used to Collect Feedback Information About Product Use. Procedia CIRP, 21, 212–217.

Worinkeng, E., Joshi, S., & Summers, J. D. (2015). An Experimental Study: Analyzing Requirement Type Influence on Novelty And Variety of Generated Solutions. International Journal of Design Creativity and Innovation, 3(2), 61–77. https://doi.org/10.1080/21650349.2014.909294

Author Biographies

Stefania Altavilla is a research fellow at the Department of Faculty of Science and Technology at the Free University of Bozen-Bolzano. Her research activities focus on the exploration of methods for product cost estimation, looking at the evaluation of the effects of design choices on product life cycle costs. More in general, her research interest focus on the design and management of innovation, with a specific focus on the methods and tools for the modelling of product development processes, platform design and associated trade-offs.

Niccolò Becattini is Assistant Professor at the Department of Mechanical Engineering of Politecnico di Milano, where he has been working since 2009. His researches focus on Design Methods and Tools, with a specific focus on Systematic Innovation and Design creativity. Within this thread, he developed Computer-Aided systems for (Co-)Design by studying the inherent role of cognition. He is currently working on national and EU funded projects. In 2014 he was awarded by ASME – CIE division with the Best PhD Thesis/Dissertation award. He is author of more than 60 publications on authoritative journals and conference proceedings.

Lorenzo Fiorineschi obtained the Master degree in Mechanical Engineering at the University of Florence, Italy (2006) and the PhD in Industrial Engineering at the same Institution (2015). He is currently Research Fellow at the Industrial Engineering Department of the University of Florence, Italy. His research interests include engineering design methods, mechanical design, creativity assessment, product architecture, prototyping, problem-solving techniques, and intellectual property. He actively participated as lead designer to several multidisciplinary projects and collaborated to many design-related academic courses. He authored more than 20 publications in authoritative journals and international conferences.

Federico Rotini took the master degree in Mechanical Engineering at Florence University, Italy. He is Ph.D. in Machine Design and works as Associate Professor since 2017 at the School of Engineering of the Florence University. His main topics of research concern the development of methodologies to support Creativity in product and process innovation, Business Process Re-engineering activities and the Product Design Cycle. He is author of more than 90 scientific papers presented in International Conferences and published in authoritative Journals, one book and four patents. He is involved in many research projects having national and international relevance. He holds courses at the Schools of Engineering and Architecture, both belonging to Florence University.