Journal of Building Engineering 51 (2022) 104223

Contents lists available at ScienceDirect

Journal of Building Engineering

journal homepage: www.elsevier.com/locate/jobe

Using machine learning techniques for architectural design tracking: An experimental study of the design of a shelter

Eva Millán [a], María-Victoria Belmonte [a,*], Francisco-Javier Boned [b], Juan Gavilanes [b], José-Luis Pérez-de-la-Cruz [a], Carmen Díaz-López [c]

[a] Departamento de Lenguajes y Ciencias de la Computación. Escuela Técnica Superior de Ingeniería Informática, Universidad de Málaga, Málaga, Spain

[b] Departamento de Arte y Arquitectura. Escuela Técnica Superior de Arquitectura, Universidad de Málaga, Málaga, Spain

[c] Departamento de Construcciones Arquitectónicas I, Universidad de Sevilla, Sevilla, Spain

ARTICLE INFO

Keywords:
Machine learning
Design theory
Architectural projects
Experimental study

ABSTRACT

In this paper, we present a study aimed at tracking and analysing the design process. More concretely, we intend to explore whether some elements of the conceptual design stage in architecture might have an influence on the quality of the final project and to find and assess common solution pathways in problem-solving behaviour. In this sense, we propose a new methodology for design tracking, based on the application of data analysis and machine learning techniques to data obtained in snapshots of selected design instants. This methodology has been applied in an experimental study, in which fifty-two novice designers were required to design a shelter with the help of a specifically developed computer tool that allowed collecting snapshots of the project at six selected design instants. The snapshots were described according to nine variables. Data analysis and machine learning techniques were then used to extract the knowledge contained in the data. More concretely, supervised learning techniques (decision trees) were used to find strategies employed in higher-quality designs, while unsupervised learning techniques (clustering) were used to find common solution pathways. Results provide evidence that supervised learning techniques allow elucidating the class of the best projects by considering the order of some of the decisions taken. Also, unsupervised learning techniques can find several common problem-solving pathways by grouping projects into clusters that use similar strategies. In this way, our work suggests a novel approach to design tracking, using quantitative analysis methods that can complement and enrich the traditional qualitative approach.

1. Introduction

The study of real design processes to understand the way people design can generate knowledge that, in turn, could be used for the development of Computer-Aided Design (CAD) tools to support designers and to guide instructors. One aspect of this study is '*design tracking*'. By this, we mean tracing the design process (understood as the sequence of actions and decisions selected and executed by the designer) and then analysing the information obtained to achieve relevant conclusions. The main research goal of the work presented in this paper is to use design tracking to discover 'good' design practices (i.e., practices that lead to higher-quality solutions) and common pathways (similar paths in the solutions provided by different designers). More concretely, this goal can be described by two

* Corresponding author.
E-mail address: mavi@lcc.uma.es (M.-V. Belmonte).

https://doi.org/10.1016/j.jobe.2022.104223
Received 25 November 2021; Received in revised form 11 February 2022; Accepted 12 February 2022
Available online 17 February 2022
2352-7102/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

research questions:

RQ1. Is it possible to predict the quality of a final design by tracking the design process at selected moments of completion?

RQ2. Which are the common problem-solving functional and spatial solution pathways that designers follow in simple design problems?

To this end, we present an empirical study in which a design process is tracked and analysed using a novel approach. The design task consisted of designing a shelter, which is a simple design problem and therefore allows reducing the multiple variables of the architectural design process but maintains its essence. The study was carried out with a group of 52 architecture students, who used a specific computational design tool, MG-Shade. During the design process, each student had to provide snapshots of his/her design project at six selected moments. In every snapshot, certain variables were collected. In this way, it was possible to analyse and code the evolution of the design between two selected instants. Then, machine learning techniques were applied to the dataset made up of the variables extracted from the designs at the six selected moments of completion.

The rest of the paper is organized as follows: Section 2 presents some relevant antecedents. Section 3 describes the materials and methods used. Section 4 shows the data analysis and results, which are discussed in Section 5, and Section 6 finalizes the paper with some concluding remarks.

2. Background and related work

Tracing the path of decisions and actions taken by the designer is a problem that has been studied by many researchers. In our study, we focus on the conceptual stage. According to Ref. [1], 'in the architectural design, a designer focuses on the conceptual stage. In this stage, the designer generates and explores new ideas based on analysing the problem'. At this stage, designers usually use sketches (either freehand or digital) to express their ideas, to explore and recognize conflicts and possibilities, and to revise and refine their solutions, generating new ones [2].

Design tracking in the conceptual design stage is usually based on a detailed analysis of the sketches produced, using different methodologies. On the other hand, some studies are mainly qualitative (QL), while others use some qualitative techniques (S), and others are mainly quantitative (QT). Table 1 summarizes the more relevant experimental studies found, classified according to the methodology used. Each of these studies is also labelled (QL, S, or T) to account for the nature of the analysis techniques used.

The first studies analysed freehand sketches (upper rows of Table 1). In these sketches, the designer captures design ideas and proposals directly in a paper. Later studies also considered digital sketches (lower rows of Table 1), i.e., sketches generated by the mediation of a computer tool, even in a virtual reality environment [18].

It is usually acknowledged that protocol analysis has been the major technique to examine cognitive processes in design [6], so it is the first and most used methodology for both kinds of sketching, as can be seen in Table 1. Sometimes, semi-structured interviews [12] or self-reports of the design activity [11] are also used. Some of these studies rely on Goel's concepts of vertical and lateral transformations [19] or in the function-behaviour-structure paradigm [17]. Linkography [20] is a usual graphical formalism for representing the path of design decisions. A more formal alternative for the analysis of a sequence of sketches is given by shape grammars, as in Prats et al. [8].

When digital sketching is used, the problem of recording all the elementary interactions of the designer with the sketch is somehow simplified, since we have access to the log history [13].

A common characteristic of these studies is that, due to the difficulty of involving human subjects to participate in the experiments and the high cost of the methodologies used, most of them have been carried out using a small sample of designs.

Our approach is based on the use of design tracking techniques to find common problem-solving strategies and design pathways. In this sense, Clarkson and Eckert [21] interpret that a design pathway is something that can be recognized as a recurring event in the

Table 1
Experimental design tracking studies.

	Methodology	Studies
Freehand sketching	Protocol analysis	Brösamle & Hölscher 2018 (QL) [3]
		Kavakli, Suwa et al., 1999 (QL) [4]
		Schön & Wiggings 1992 (QL) [5]
		Suwa, Purcel et al., 1998 (QL) [6]
		Suwa & Tversky 1997 (QL) [7]
	Shape grammars	Prats, Lim et al., 2009 (QL) [8]
	Goel methodology (Vertical and lateral transformations)	McGown, Green et al., 1998 (QL) [9]
		Rodgers, Green et al., 2000 (QT) [10]
	Analysis of open-ended design briefs	Goldschmidt & Rodgers 2013 (S) [11]
	Analysis of semi-structured interviews	Serrano Lasa et al., 2021 (S) [12]
Digital sketching	Classification of logs of design behaviour	Zhu, Zeng et al., 2021 (QL) [13]
	Protocol analysis	Bilda & Demirkan 2003 (QL) [2]
		Gül, 2014 (QL) [14]
		Salman & Laing 2014 (S) [15]
		Tang, Lee et al., 2009 (S) [16]
	Segmentation and function-behaviour-structure coding	Tang, Lee et al., 2011 (QT) [17]
	Linkography	Heidari and Polatoğlu, 2019 (QT) [1]
	Virtual Reality	Neroni et al., 2021(QT) [18]

design process and detected and described at any level of detail. The identification of this kind of design pathways can be useful to conceptualize strategies in problem-solving [22–27]. Such strategies can be useful to help designers face new problems and overcome the nebulous nature of the design process [21,28].

Our methodology is different to the one used in the studies presented in Table 1. In the first place, we have used a computer tool specifically designed to track the design process. In our experiment, designers use MG-Shade to produce the digital sketches to be then analysed. Second, our design tracking is based on the analysis of a dataset obtained by collecting the values of certain variables at selected moments of the design process. In this way, we address one of the main difficulties when evaluating the quality of design projects as described in Refs. [22,29], namely that the quality of the project should not only be scored not only by looking at the final design, but also by considering the student's design process. Third, the dataset obtained is analysed using machine learning techniques. Machine learning has been used successfully for other design-related issues (see for example Sun et al. [30], Chang et al. [31] and Loche et al. [26]) but, as far as we know, not for design tracking.

The use of machine learning for design tracking may suggest a move towards an intensification of the quantitative approach in the field of design studies. Hopefully, it can reduce the time spent on data analysis [32] and provide an in-depth understanding of the design processes students have gone through [33]. On the other hand, quantitative studies present some technical difficulties, for example, the problem of controlling variables, and some conceptual drawbacks, like the difficulty to quantify abstract ideas or infer meaning beyond the raw results given by statistical analysis [34].

In summary, and to the best of our knowledge, our methodology for design tracking is novel in several regards: a) the use of a specifically designed software tool (MG-Shade); b) the nature of the dataset used; and c) the use of machine learning techniques.

3. Materials and methods

3.1. Materials

3.1.1. Participants and environment

Our study was carried out with students from the course *Composición Arquitectónica I (Architectural Composition I)*. This course is compulsory for students in their seventh semester at the University of Málaga School of Architecture (mostly 21-years-olds). The course provides an introduction to contemporary architecture (the development of modernity from the early twentieth century avant-garde) and is conceived to provide the students with a useful basis for their subsequent development as architects.

A total of 52 students (26 females and 26 males) participated in the study. This group of students was selected because, being in the fourth year of their architectural degree, they were advanced enough in their curriculum to accomplish this design task and develop interesting projects.

The learning objective of *Composición Arquitectónica I* is to know and handle the main underlying concepts in architectural theory from the Renaissance to the present. One of the tasks assigned to students in this subject is the design of a shelter using a software tool, which we describe next.

3.1.2. Software tool: MG-Shade

The use of the same computer tool was mandatory for all projects. Ideally, the tool should be known by the users and powerful enough to easily modify and display the designs interactively. To achieve these goals, we used Trimble SketchUp, a well-known 3D modelling program that combines a toolset with an interactive drawing system.[1]

However, to trace the design paths in a uniform way, certain additional features must be provided, and the full drawing capabilities of SketchUp should be restricted. For this reason, plain SketchUp was not used in our research; instead, we used a SketchUp plugin (implemented in Ruby) called MG-Shade (MeGastructures SHApe DEsign). MG-Shade was originally developed for a prior project [35] to interactively design megastructures and, with some minor modifications, was well suited to our present purpose (interactive design of shelters). The main idea underlying MG-Shade is that the user interacts with the systems in a structured way, adding or deleting elements by the application of a rule selected from a predetermined set (in fact, MG-Shade is based on the formalism of shape grammars [36]). In this way, it is possible to take snapshots of the state of the design project at selected times of completion and at the desired level of granularity. Notice that MG-Shade is not intended to be 'intelligent'; it is just a tool that allows users to design the shelter progressively, dividing the design process into stages. This distinctive feature of MG-Shade provides adequate capabilities for the task of design tracking.

Fig. 1 shows a screenshot of the graphical interface of MG-Shade. Rules of the selected grammar appear on the left-hand side of the canvas. The first rule adds an adjacent cube to the initial shape and the second one adds a cube that shares an edge with the initial shape.

MG-Shade provides a set of operations (activating the corresponding icon in the toolbox) which allow the user to control the generation/design process:

- Establish a maximum number of pieces for the design.
- Select the basic piece or cell for the design (cube, triangular prism, or hexagonal prism).
- Activate/deactivate a volume constraint. If activated, all the cells must lay inside a cube with a given maximum size.
- Place a piece anywhere on the design canvas (at the start or at any step).
- Delete a piece present in the design.

[1] www.sketchup.com.

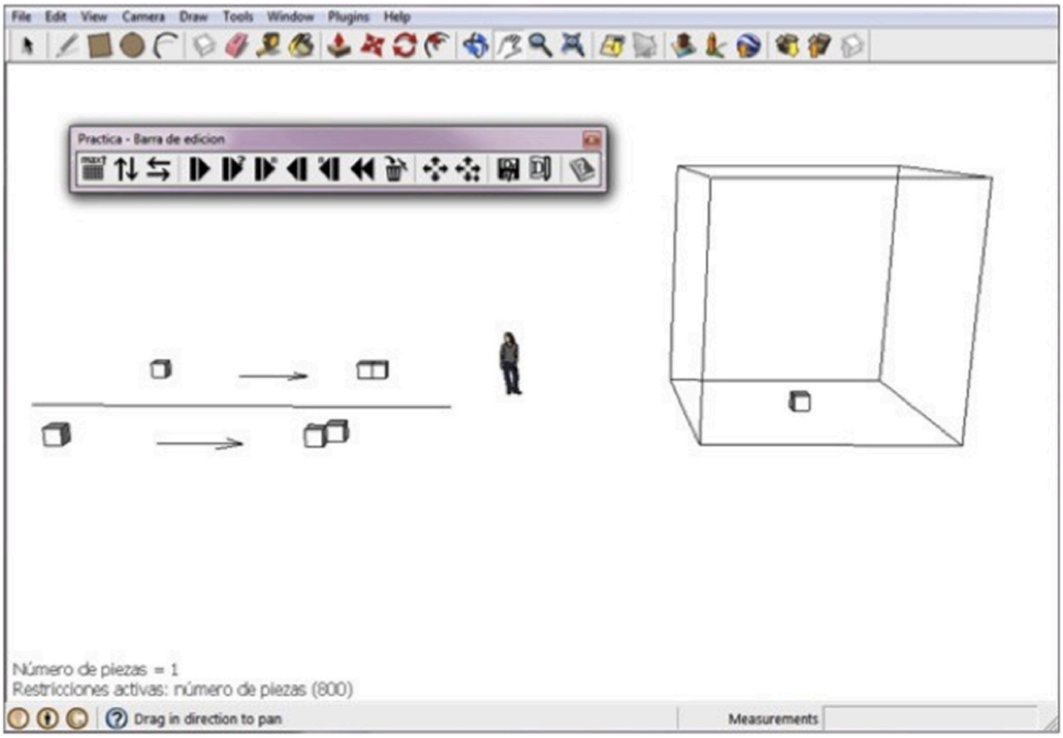

Fig. 1. The graphical interface of MG-Shade. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

- Select the cell to which another cell will be adjoined at each step.
- Select the rule(s) that will be applied to add new cells or pieces.
- Select the number of consecutive applications of the selected rule(s).
- Select rules that produce only horizontal or vertical growth.

3.1.3. Design task

For the individual assignment, each student had to design a shelter using the MG-Shade software tool.

There are some basic reasons for choosing the design of a shelter for the task of the research work presented in this paper. This choice has allowed a certain simplification of the multiple variables of the architectural design process without losing its essence.

The choice of the shelter for the design task can be interpreted as a return to the origins of humanity, not only because of its closeness to the concept of a 'house' as a place to live but also due to the simplicity of the set of needs of such a piece. In the same way that a house is more than a roof and walls, the shelter is also more but on a smaller scale [37]. A refuge means shelter and abode, but it is also a representation, a symbol, a dream, or a place of happiness [38]. The refuge-house offers the first protection to the body [39]. It also involves the use of very few resources to achieve maximum protection and well-being, an economy of means that creates a need to sharpen one's wits. In this sense, the minimum requirements to provide shelter from the weather and still qualify to be considered as architecture are; a boundary with a well-defined exterior, an access from the outside, and an interior space that has natural light and is stable, safe, and habitable.

The task was to design a shelter in compliance with the following requirements:

- It should have an exterior access and natural lighting.
- It should be composed of pieces of 40 × 40 × 40 cm^3 (cubes).
- It should be contained in an enveloping cube of 6 × 6 × 6 m^3.
- For economic efficiency, the number of pieces used for the design should be between 1688, and 2025 pieces.[2]

The students were asked to provide: a) six snapshots of their creative process, that we call *design instants*, at selected moments of the design process (10%, 25%, 50%, 75%, 90%, and 100% of pieces), and b) a written report about their final design, including front, lateral, and top views, axonometric projections, linear perspectives, and sections.

Fig. 2 and Fig. 3 show the work presented by one of the participants, whose goal was to produce a regular shelter with two lounges

[2] This number of pieces represents between 50% (useable minimum) and 60% (useable maximum) of the total number of cells that would result from filling the volume of the enveloping cube.

(one dark, one with natural light).

3.2. Methods

3.2.1. Students' projects processing

To assess the overall quality of the final designs, an independent expert (an architecture lecturer with more than 20 years of experience both in teaching and in professional practice) evaluated them. He graded the projects according to the perceived quality in four categories: aesthetics (Qa), functionality (Qf), technical aspects (Qt), and economical aspects (Qe). These criteria were developed by adding the economical aspect to the Vitruvian triad (*firmitas, utilitas, venustas*). Each of these aspects was given a value from 1 (worst value) to 5 (best value). The quality of each project (Q) was then computed as the sum of the values of these four variables. Therefore, Q takes values from 4 (worst) to 20 (best). This expert was one of the teachers of this group of students. His evaluation of this project was used as part of their final grade and discussed in the oral presentation session of the projects, and both other teachers and students agreed that the grades were adequate.

Additionally, two of the authors of this paper (both architects with extensive teaching and professional practice experience) analysed every design instant of every project. This analysis was blind concerning the quality and authorship of the projects. For the analysis, instants I1 to I6 of each design were considered. Each design instant Ii was compared with the previous design instant Ii – 1, and the evolution between the two instants was coded according to a set of variables. These variables were defined according to ideas taken from contemporary architectural theory, which we will explain next.

The first variable is related to the prevailing process used at that instant. In general problem solving, it is possible to reach a solution by adding new elements to the present configuration or by subtracting elements from it [40]. This is what, in the architectural field, is referred to as the contraposition of tectonic vs stereotomic reasoning [41], where the former works on structural elements to configure a whole, while the latter works by 'subtracting' elements from an initial mass.

The type of growth exhibited in each step must be also recorded. Following [42], it is important to detect laminar and volumetric growth, given by the accumulation of planes and volumes, respectively. The spatial growth [43,44] is given by the definition of interior spaces. On the other hand, the design can grow in a regular/irregular way. Architectural shapes can also be transparent (open to the exterior space) or rather closed and opaque [45].

Finally, since the work of A. Klein, the importance of function and functional variables is widely acknowledged in the conception of a building [46]. Therefore, the entry to a building and the inner paths to the specialized spaces must be considered in the architectural design, independently of the size of the project [44]. So, it is relevant to record the instant in which each functionality appears in the design path. In our case (shelters), three functionalities must be provided: to get into the shelter (access), to move inside the shelter (track), and to stay inside the shelter (lounge). To sum up, the following variables and values were recorded for each instant of the design:

- *Addition vs Subtraction of shapes*, AS (A, S). Addition is the operation by which some pieces are placed on others. New pieces can be either stacked or attached to the existing ones. On the contrary, subtraction is the operation that involves the extraction of pieces, generating holes in the form.
- *Laminar Growth*, LG (yes, no). Laminar growth is based on the preferential use of horizontal, vertical, oblique planes.

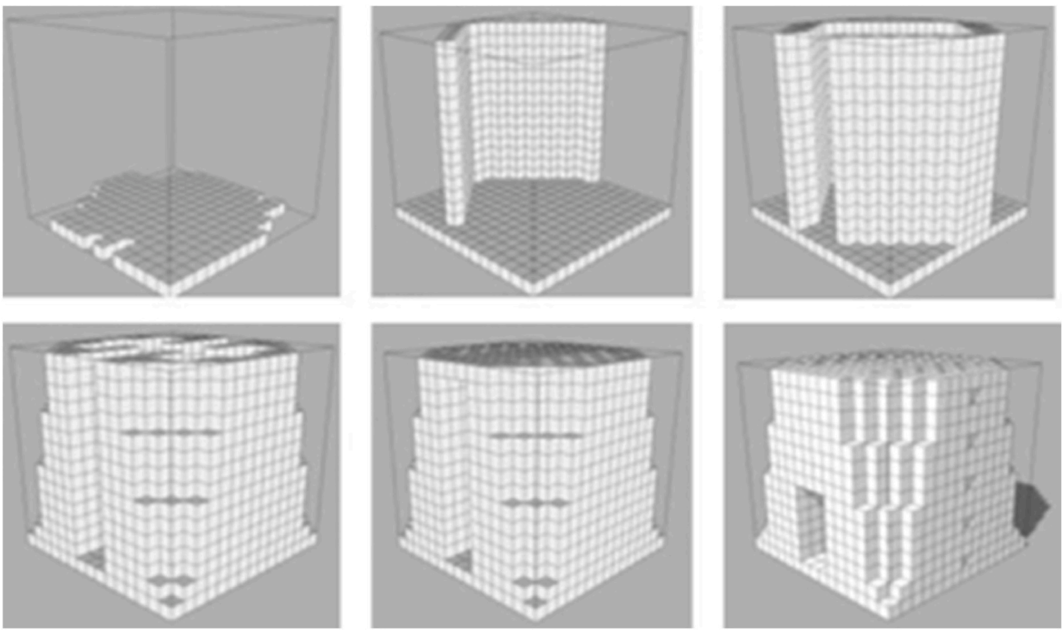

Fig. 2. Example of the six selected design instants.

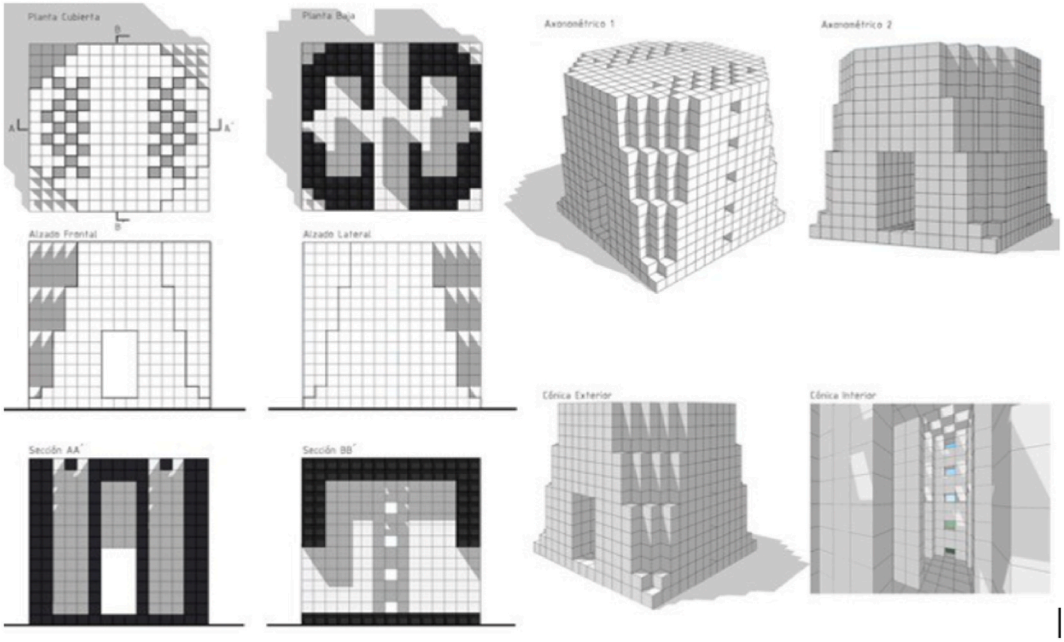

Fig. 3. Example of a final project.

- *Volumetric Growth*, VG (yes, no). Volumetric growth prioritizes the generation of volumes. The three dimensions are more balanced. It originates forms.
- *Spatial Growth*, SG (yes, no). Spatial growth causes the appearance of interior spaces, e.g., lounges and/or rooms that have floors, walls, and ceilings.
- *Regular/Irregular/Partial Growth*, G (R, I, P). In regular growth, regular forms, parallelisms, defined edges, and similar dimensions prevail. In irregular growth, irregular boundaries (with less clear or defined geometries) arise: scaling, edges, irregular curvilinear paths, jumps, and imbalances. Finally, where there is no clear prevalence of regular or irregular forms, we have used the label P (partial growth).
- *Opacity vs Transparency*, OT (O, T). In opaque designs, there is a prevalence of closed forms, wall-like, and a low proportion of gaps. On the contrary, transparency refers to shapes in which the proportion of holes is high, and it is possible to see through the shape. In transparent shapes, there is a stronger relation between inner and outer spaces.
- *Dimensional Function 1: Shelter's Access*, DFA (yes, no). It refers to the instant in which the access from the exterior to the interior of the shelter is designed.
- *Dimensional Function 2: Shelter's Track/Path*, DFT (yes, no). It refers to the instant in which the shelter's entrance is connected to the interior rooms. The connecting path can be horizontal, vertical, or oblique.
- *Dimensional Function 3: Shelter's Lounges*, DFL (yes, no). The lounges in the shelter are the places of maximum protection in the design. A lounge complies with the conditions to be considered a space, and it is possible to stay in it.

Each researcher independently analysed and coded the evolution between the selected design instants of each of the 52 projects for the analysis. Where there was disagreement, the researchers discussed the data and revised their coding accordingly. Fig. 4 shows examples of such coding (in the snapshots taken at five selected completion stages) for four sample design projects presented by the students.

3.2.2. Dataset pre-processing

Considering the six instants and the nine variables, there are a total number of $6 \times 9 = 54$ variables. For the *yes/no* binary variables, the *yes* value was coded as 1, while the *no* value was coded as 0.

Some of the participants did not present all the six instants, so there were missing values. According to Acuña and Rodríguez [47], the use of attributes or variables with more than 15% of 'lost' values may be questionable when data analysis is performed. For this reason, six participants and the design instant I5 were removed from our analysis. This reduced our dataset to 46 participants and 45 (9×5) variables.

In addition to these data, we used the evaluation of the quality of the final designs, Q. Those designs with a score greater than or equal to 18 points were classified as 'best', and the remaining ones were classified as 'rest'. As a result, from the 46 projects, 7 (15,22%) were assigned to the 'best' class and 39 (84.78%). to the 'rest' class (Examples of such classification can be seen in Fig. 4: according to their score, projects 1 and 2 were classified as 'best', while projects 3 and 4 were classified as 'rest'.

Fig. 4. Examples of the coding of the defined variables.

3.2.3. Dataset processing concerning RQ1

As aforementioned, one of the goals of our research is to discover 'good' design practices (i.e., practices that lead to higher-quality solutions. This goal was expressed as our first *research question:*

RQ1. Is it possible to predict the quality of a final design by tracking the design process at selected moments of completion?

Therefore, in this case, our target variable is the concept or class of the 'best' projects. Variables used for the characterization (i.e., the predictor variables) are the 45 variables described above.

This is a problem of supervised learning, since we want to characterize or describe a concept based on instances of it. That is, a descriptive model will be constructed from the set of instances (participants), which contain both examples and non-examples of the concept (good project) to be learned. This set of instances is called the training data.

The statistical analysis was performed using the WEKA software [48]. We used classifiers based on Decision Trees (J48 and Random Forest), rules (JRip), and Bayesian classifiers (Naïve Bayes and Bayesian Network). All these techniques are intended to discover a model capable of predicting the target variable (being/not being in the class of best designs).

To evaluate the accuracy of supervised classification models, the confusion matrix is constructed. It relates the predictive power of the model to the data. The elements of the confusion matrix are:

- TP (True Positives), i.e., cases are positive instances (class 'yes') that were correctly classified as positive.
- TN (True Negatives) are negative instances (class 'no') that were correctly classified as negative.
- FP (False Positives) are negative instances that were incorrectly classified as positive.
- FN (False Negatives) are positive instances that were incorrectly classified as negative.

The confusion matrix is shown in Table 2.

Table 3 summarizes some of the indicators more commonly used in the validation phase of the models, namely: sensitivity, specificity, and precision values, which are between 0 and 1 (the higher, the better). It also presents another commonly used concept to evaluate the performance of a binary classification model, the Area under the ROC curve (AUC). The Receiver Operating Characteristic (ROC) curve plots the true positive rate and false positive rate as their discrimination threshold is varied. The AUC value is commonly used by the machine learning community to compare the predictive capacity of binary classifiers. AUC values over 0.8 are good, and over 0.9, excellent.

To validate the models, we have used k-fold cross-validation: data are divided into k subsets, and then one of these subsets is used for the test data and the remaining (k – 1) as training data. The validation process is repeated for k iterations, with each one of the possible k subsets playing the role of test data, and the average of all iterations is calculated. A usual value is $k = 10$. We have used 10-fold-cross-validation.

The results of the application of these techniques are shown in Section 4.1.

3.2.4. Dataset processing concerning RQ2

In the research reported here, we intend to find common solution paths of problem-solving behaviours corresponding to classes or groups of participants. This was verbalized as our second research question:

RQ2. Which are the common problem-solving functional and spatial solution pathways that designers follow in simple design problems?

In this case, this is an unsupervised learning problem. In unsupervised learning, the goal is to characterize a new or unknown concept from instances of it. There are no predefined classes, and algorithms try to create them from the available data. Several variants of clustering are typically used to achieve this goal.

The input of clustering algorithms is a set of objects. The output is a partition of the set in several groups (or clusters) in such a way that objects belonging to the same group are 'similar', and objects belonging to different groups are 'dissimilar'.

In our study, we have used hierarchical clustering analysis. The main output of this technique is a dendrogram [49], which shows the hierarchy of the clusters produced by the hierarchical clustering algorithm: each number at the lowest level represents an instance (a participant), and each fork in the dendrogram represents a cluster. The vertical axis measures the dissimilarity between clusters. At each step, the two clusters (or individuals) that are most similar are joined into a single new cluster. Their fusion is represented by a new node: the two joined clusters are connected by vertical lines descending from the junction, and the length of the horizontal line measures the dissimilarity between them. The process goes on until all instances are descendants of a single node (see Fig. 7 for an example of a dendrogram). Algorithms for hierarchical clustering are agglomerative (bottom-up) or divisive (top-down). The Ward algorithm is a popular agglomerative algorithm. As for the quality of the results, in general, a clustering is 'good' when instances inside the same cluster are close, and instances belonging to different clusters lie far away. Several measures or indicators of clustering quality have been defined. We consider the following in our study:

- *Connectivity*. It indicates the degree of connectedness of the clusters, computed by considering the distances between k-nearest neighbours. Connectivity should be minimized.
- *Dunn index*. It is the ratio of the smallest distance between instances belonging to different clusters, to the largest distance between instances belonging to the same cluster. This index should be maximized.

Table 2
Confusion matrix.

	Reality	
Prediction	yes	no
yes	True Positive (TP)	False Positive (FP)
no	False Negative (FN)	True Negative (TN)

Table 3
Quality indicators.

Indicator	Formula	Definition
Sensitivity TP Rate	TP/(TP + FN)	Number of positive instances correctly classified divided by total number of positive instances
Specificity TN Rate	TN/(TN + FP)	Number of negative instances correctly classified divided by the total number of negative instances
Precision	TP/(TP + FP)	Proportion of positive instances that are true positives
AUC		Area Under ROC curve

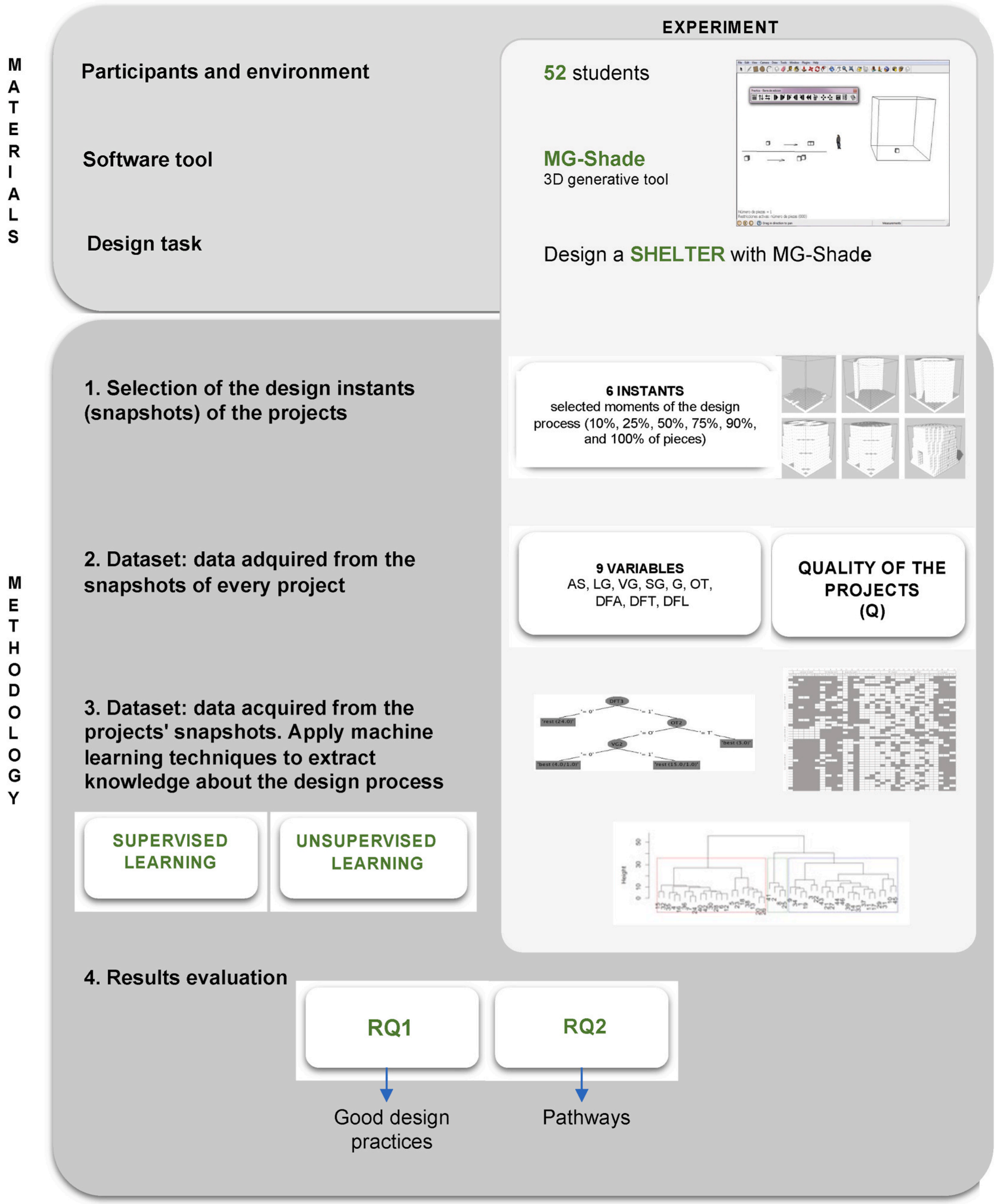

Fig. 5. Materials and methods.

- *Silhouette width.* It tries to measure the difference between the average intra-cluster dissimilarity and the average dissimilarity for two neighbour clusters. This index should be maximized.

In our experimental data, not all variables in the dataset should be considered for the analysis. For example, the class variable is not considered in unsupervised learning, so in our case, the variable quality Q has been removed. We have also discarded variable G because it is the only one that is not binary. Finally, since our goal is to find differences among design tracks, variables that have the same value for all the participants are not relevant. For this reason, variables whose value is the same for 90% or more of the participants have also been removed from the clustering analysis, which is the case of variables AS1-4, VG4, OT1, SG3, and SG4. In this way, the 32 binary variables used for this analysis are: AS5, LG1, LG2, LG3, LG4, LG5, VG1, VG2, VG3, VG5, SG1, SG2, SG5, OT2, OT3, OT4, OT5, DFA1, DFA2, DFA3, DFA4, DFA5, DFT1, DFT2, DFT3, DFT4, DFT5, DFL1, DFL2, DFL3, DFL4, and DFL5.

The analysis was performed by using the R language with additional packages: cluster [50] and clValid [51] for clustering. The Manhattan distance between instances was considered, and the Ward algorithm (variant Ward.D) was applied.

Fig. 5 shows a graphical summary of the materials and methods used.

4. Experimental results

4.1. Results concerning RQ1

Table 4 and Table 5 show the performance results of the different classification models mentioned in Section 3.2.3.

The best results are obtained by the J48 decision tree, which shows good values for all the indicators. Its AUC is 0.848, which is considered a good value. As for the other indicators, all of them are above 0.8 (except for the TP rate of the 'best' class, which is 0.714). The J48 decision tree model correctly classifies 89.1% of the instances. Table 6 shows the confusion matrix of the J48 decision tree model. It correctly classifies 5 out of the 7 instances of the 'best' class, and 36 of the 39 instances of the 'rest' class.

Fig. 6 shows the decision tree produced by the J48 classifier. To classify a new example using this decision tree, we should look at the value of DFT3 (variable DFT in the design instant I3). If it is 0, the example will be classified as 'rest'; if not, we check the value of OT2 (variable OT in the design instant I2). If it is T, the example is classified as 'best'; if not, we look at the value of VG2 (variable VG in the design instant I2). If it is 0, the example belongs to the class 'best', or to 'rest' otherwise. Therefore, the more discriminant variables are DFT3, OT2, and VG2.[3] In Section 5.1, we will discuss how this tree structure, in terms of the most discriminatory conditions, is supported by the analysis of the 'best' projects.

In the decision tree, we can also appreciate that, within the training set, only two instances are incorrectly classified. A 'best' project is classified as 'rest' (leaf node best (4.0/1.0)), and a 'rest' project is classified as 'best' (leaf node rest (15.0/1.0)).

4.2. Results concerning RQ2

Fig. 7 shows the dendrograms generated by the Ward algorithm (the raw output of the algorithm in the upper half and the three selected clusters in the lower half).

The quality of tentative clustering was computed for cutting the dendrogram in such a way that n clusters should be generated ($n =$ 2, 3, 4, 5). The corresponding values for the measures are shown in Table 7.

Optimal scores were obtained for $n = 2$ (connectivity), $n = 3$ (Silhouette), and $n = 4$ or $n = 5$ (Dunn). Recall that connectivity should be minimized, while Dunn and Silhouette should be maximized. Therefore, the solutions obtained are non-dominated. We decided to choose $n = 3$ clusters because it maximises the Silhouette value while still providing good values for Dunn and Connectivity (second-best values, in both cases).

More explicitly, the clusters or classes are as follows (each number represents a particular participant):

- C1 = {1,3,9,10,11,14,17,19,21,22,27,29,31,33,34,37,39,43,44,45,46}
- C2 = {2,8,25,41}
- C3 = {4,5,6,7,12,13,15,16,18,20,23,24,26,28,30,32,35,36,38,40,42}

For each variable V and cluster C, the percentage of instances with positive (1 or yes[4]) values is shown in Table 8.

On the other hand, average values (scale from 4 to 20) for the quality grades (Q) are 12.1 for C1, 11 for C2, and 11,14 for C3, and the global average is 11.56.

Fig. 8 graphically displays the output of the clustering. Each column represents a variable, and each row represents an instance (participant). White cells stand for 0 (no) values and black cells for 1 (yes) value. The findings of the clustering algorithm suggest that the designers have used three different problem-solving pathways to design the shelter. Each of these pathways is represented by a cluster. Cluster 1 comprises rows 1–21; cluster 2, rows 22–25; and cluster 3, rows 23–46. Most of the designers belong to clusters 1 and 3 (25 are in cluster 1 and 22 in cluster 3), while 4 participants are in cluster 2.

In Section 5.2 we will describe these classes qualitatively and characterize the detected styles from an intuitive point of view.

[3] Recall that DFT, OT, and VG describe the shelter's track, opacity vs transparency, and vertical growth respectively.

[4] For this analysis, values T for variable OT and S for variable AS were coded as 1 and their opposites as 0.

Table 4
Performance average results of classification algorithms and TP Rate of 'best' and 'rest' classes obtained with 10-fold-cross-validation.

Model	TP Rate best	TP Rate rest	Precision	Sensitivity	AUC
J48	0.714	0.923	0.898	0.891	0.848
Naïve Bayes	0.286	0.949	0.823	0.848	0.736
Bayes Net	0.286	0.949	0.823	0.848	0.769
Random Forest	0	1	0.719	0.848	0.718
JRip	0.143	0.897	0.7	0.783	0.548

Table 5
Instances classification with 10-fold-cross-validation.

Model	% Correctly Classified	% Incorrectly Classified
J48	89.1	10.8
Naïve Bayes	84.78	15.21
Bayes Net	84.78	15.21
Random Forest	84.78	15.21
JRip	78.0	22.0

Table 6
Confusion matrix of J48 decision tree.

		Prediction	
		best	rest
Reality	best	5 (TP)	2 (FN)
	rest	3 (FP)	36 (TN)

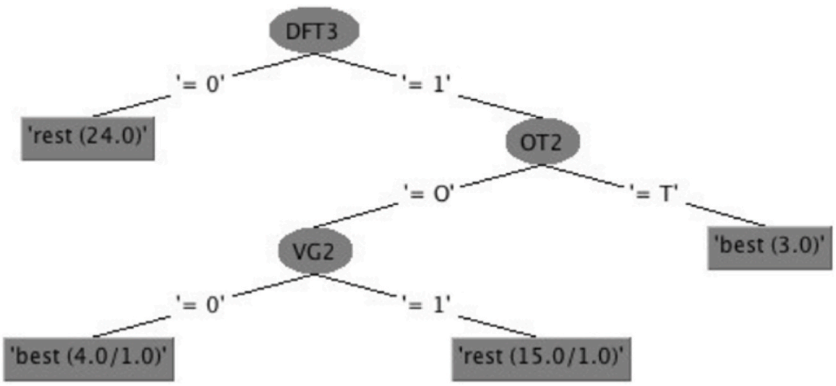

Fig. 6. Decision tree after applying J48 Classifier.

5. Discussion

5.1. Research question RQ1

RQ1. Is it possible to predict the quality of a final design by tracking the design process at selected moments of completion?

The decision tree resulting from the study shows that the most discriminant variable when classifying the 'best' projects is the DFT variable at design instant I3. If DFT3 = 0 (i.e., the internal route has not been considered at instant I3), the decision tree classifies the project as 'rest'. On the contrary, all the 'best' projects have DFT3 = 1, that is, all of them built the connection of the shelter's entrance to the interior rooms at instant I3 (which represents an intermediate point of the design process). Therefore, either because of the need for access or because of a need for spatial qualification, it seems that the creation of the route acts as a substantial operation in the project and as a key point of the architectural design process.

Analysing the best projects, most of the designers have used three different ways (strategies) to implement the shelter's track:

(1) The creation of the access (transition between exterior and interior).
(2) The design of the staircase.
(3) The creation of the main space, as the organizer of the design process, which must be reached as an internal destination.

Fig. 9 shows examples of the 'best' final designs, that use strategies one, two, and three, respectively. The first project (Fig. 9, first

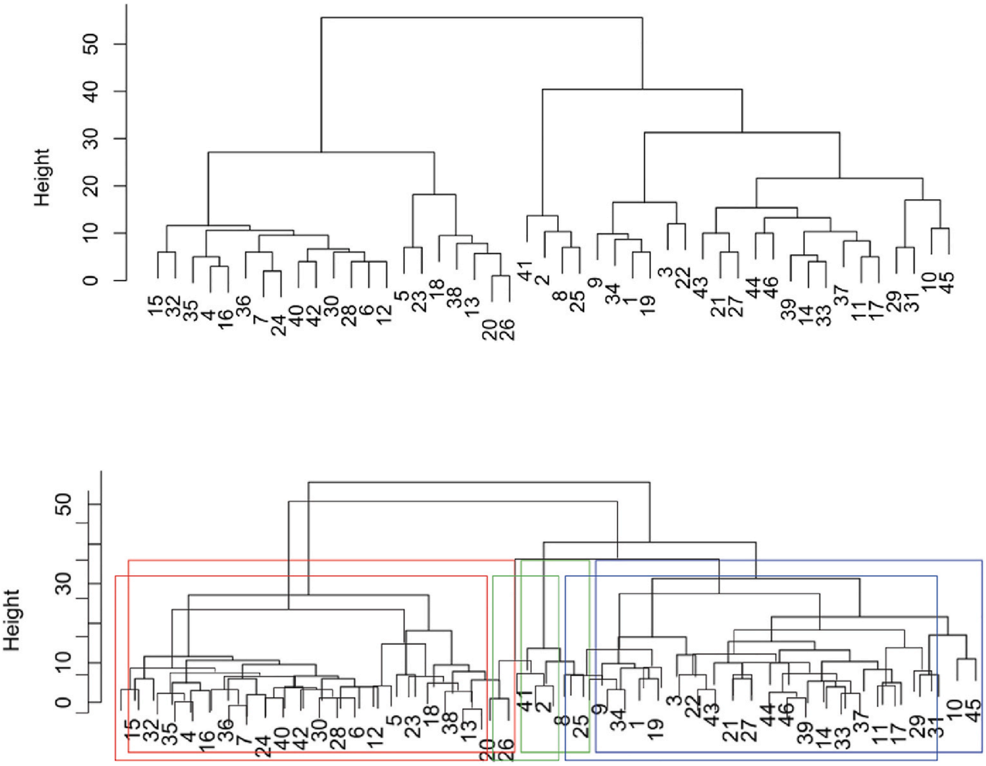

Fig. 7. Raw dendrogram (upper half) and dendrogram with the 3 selected clusters (lower half).

Table 7
Indicator values for several numbers of clusters (*n*).

n	2	3	4	5
Connectivity	31.0611	36.2631	42.2750	52.9103
Dunn	0.1818	0.2000	0.2222	0.2222
Silhouette	0.1423	0.1584	0.1562	0.1360

Table 8
Percentage of positive values for each variable V and cluster C.

V	AS5	LG1	LG2	LG3	LG4	LG5	VG1	VG2
C1	23.8	95.2	81.0	90.5	95.2	71.4	33.3	76.2
C2	0.0	0.0	0.0	0.0	0.0	0.0	100	100
C3	14.3	95.2	100	100	100	90.5	14.3	76.2

V	VG3	VG5	SG1	SG2	SG5	OT2	OT3	OT4
C1	76.2	52.4	14.3	95.2	76.2	23.8	28.6	23.8
C2	100	100	75.0	100	100	25.0	0.0	25.0
C3	95.2	81.0	0.0	14.3	95.2	0.0	4.8	9.5

V	OT5	DFA1	DFA2	DFA3	DFA4	DFA5	DFT1	DFT2
C1	23.8	19.0	47.6	38.1	23.8	23.8	23.8	33.3
C2	25.0	75.0	50.0	25.0	25.0	25.0	50.0	75.0
C3	23.8	19.0	9.5	38.1	57.1	47.6	19.0	4.8

V	DFT3	DFT4	DFT5	DFL1	DFL2	DFL3	DFL4	DFL5
C1	52.4	38.1	9.5	14.3	71.4	81.0	81.0	76.2
C2	75.0	50.0	25.0	50.0	75.0	75.0	75.0	50.0
C3	38.1	52.4	33.3	9.5	0.0	71.4	85.7	81.0

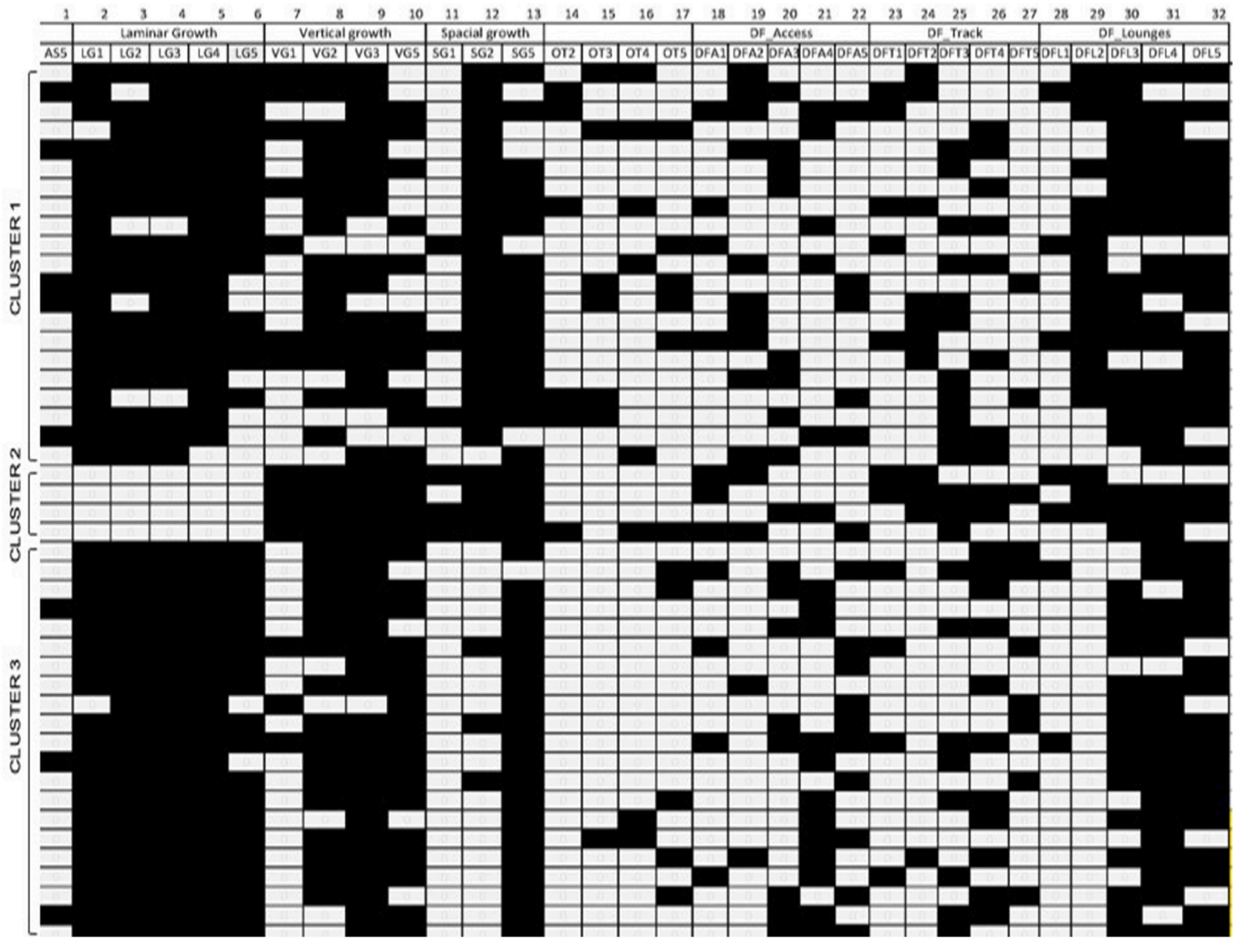

Fig. 8. Visual representation of clusters.

row) received a grade of 19 in quality (Q), while the other two received 18 (out of 20 points). For each project, we show design instants I2, I3, I4, and I6 (recall that the more discriminant variable is DFT3, so we want to see what happens exactly at instant I3) and the final design.

For the project shown in the first row in Fig. 9, the main spatial element is introduced in instant I3, and this spatial element creates and shapes the path (DFT3 = 1). The interior space is designed simultaneously with the element to access it. In instant I2, the shelter is wall-like, and the proportion of gaps is low (OT2 = O). So, the decision tree shown in Fig. 6 will correctly classify this example as 'best'.

Concerning the project shown in Fig. 9, second row, the access is created at instant I3 (DFT3 = 1). This allows the interior to be divided into two rooms (which implies an inevitable internal route). Later, the staircase is created as a vertical communication element to access the upper floor. In this way, a relationship between the access and the internal route is established in the early stages of the project.

Finally, in the project shown in the third row of Fig. 9, there is a main space that acts as the organizer of the different ways to access the shelter. This space is conceived from the mobility created by the structural support and its transparency.

Therefore, regarding the access, its character as a transition element bears relevance, especially in the single-level designs. The 'entering and reaching the space' is manifested as an indissoluble operation once the initial geometric basis has been laid according to the given rules.

In this way, we can observe that the functional fact (in this case, a space to provide shelter) is incorporated into the idea of design using several principles to organize the interior space. These principles are considered substantial for the definition of the architectural object. For this reason, the functional fact tends to appear at the central moment of the architectural design process. This suggests that, if it is not considered in the key moment, it might be difficult to complete the spatial relationships that determine the design project.

Therefore, and concerning RQ1, we can conclude that it is possible to predict the final quality of the design by tracking the design process at specific instants. The supervised learning classifier shows that the more discriminant variable when classifying the 'best' projects is the DFT variable at design instant I3, i.e., to connect the shelter's entrance to the interior rooms at the central moment of the design process. This knowledge could be used by instructors to guide novice students when solving analogous tasks.

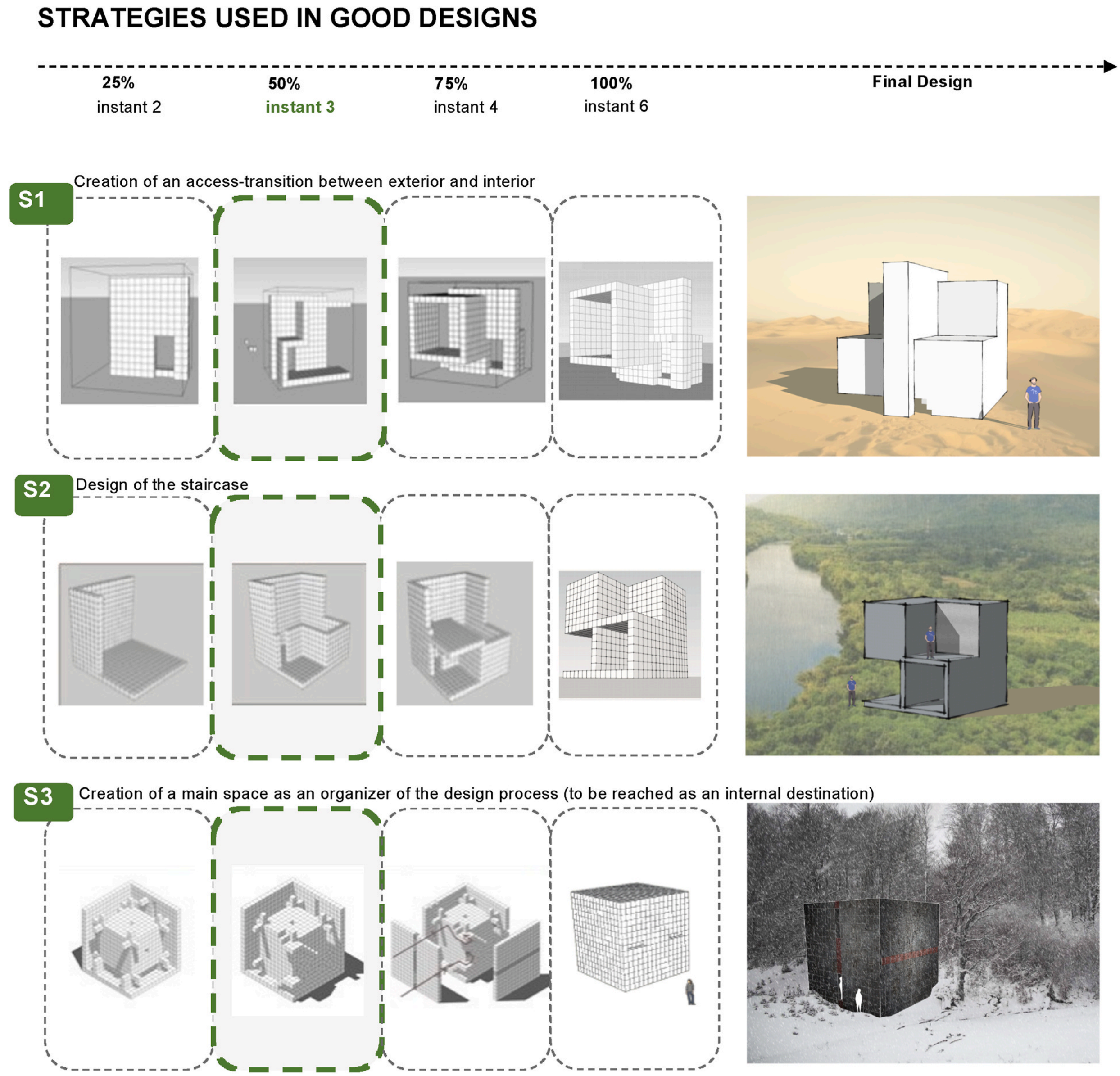

Fig. 9. Examples of the 'best' final designs.

5.2. Research question RQ2

RQ2. Which are the common problem-solving functional and spatial solution pathways that designers follow in simple design problems?

To answer this question, let us discuss the results displayed in Fig. 8. We have three clusters, each of them corresponding to a pathway.

- Cluster 1. Most of the participants in this cluster start their designs (instant I1) with laminar growth (LG1), i.e., they introduce horizontal, vertical, or oblique planes. This strategy is then kept throughout the design process. Between instants I1 and I2, most of the participants use volumetric growth (VG), which decreases in the final instants. The access to the shelter (DFA) is mostly built in the intermediate instants I2 (12 out of 25 participants) and I3 (9 out of 25 participants), and the connection of this with the inner rooms is established mainly in instants I2 and I3. Finally, and with some exceptions, the rooms (DFL) are placed from instant I2 and until the end of the design. These results are consistent with those of the SG (spatial growth) variable, which situates the appearance of interior spaces at instant I2.
- Cluster 2. Designers begin directly with the use of volumetric (VG) and spatial growth (SG), strategies that are kept during their design process. Their designs are characterized by the superposition of volumes to generate the shelter. The access (DFA) is mostly

worked at instant I1, and the interior spaces (DFL) begin to appear in instant I2 and are finished at instant I3. Finally, 3 out of 4 designs establish the connection of the spaces (DFT) at instant I2.

- Cluster 3. The strategy is like the one used in cluster 1, starting with laminar growth (LG) and postponing volumetric growth (VG). The main difference between the two clusters is that in cluster 3, volumetric growth starts at instant I2 (VG2). This delays the appearance of the rooms (DFL), which is postponed to instant I3. Consequently, there is no spatial growth until instant I3. This fact is ratified by the non-use of spatial growth (SG) in instants I1 and I2. Consequently, the values of the DFA and DFT variables of the pathway corresponding to this cluster are also displaced in time in relation to route 1. The access (DFA) and the connection with the rooms (DFT) are mostly built between instants I3 and I5.

We can also see that some of the variables considered do not have an important role in the definition of the clusters. For example, in terms of the variable OT (opacity/transparency), no significant differences appear in the three clusters. In all the designs there is a prevalence of closed forms, and a low proportion of holes (which makes sense when designing shelters). Holes are introduced sparingly, either to provide natural light or to be able to see through the opening.

Next, we present some examples of each cluster (Fig. 10) to further clarify the different pathways used in each one of them.

Fig. 10 (first row) shows an example of the first pathway. The pathway followed by the designer represents a spatial growth from

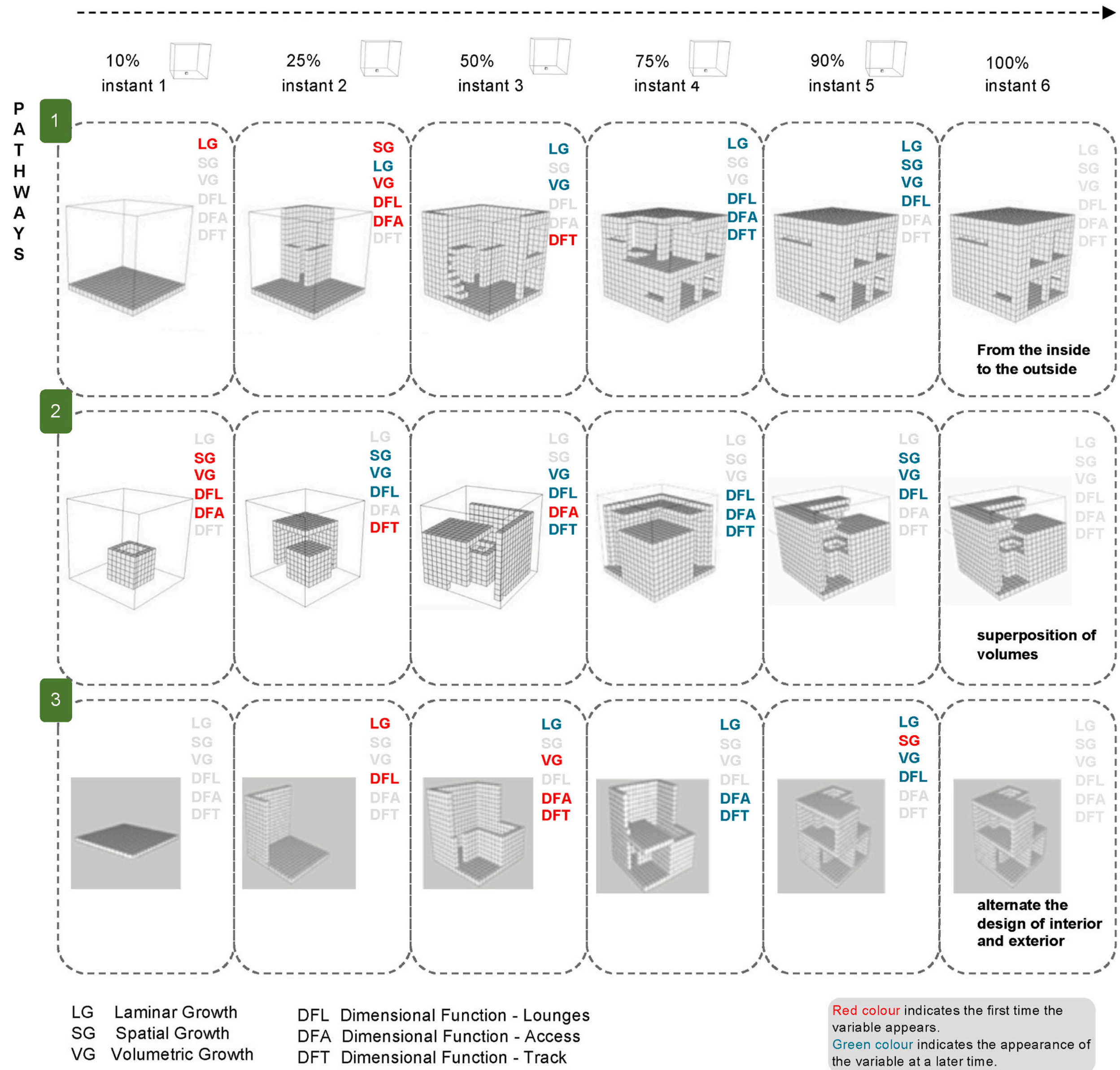

Fig. 10. Pathways' examples.

the inside to the outside, starting from laminar growth (LG) in instant I1 and followed by a small interior space (VG) in instant I2. Access (DFA) is started in instant I2 and determined in instant I4 and the connection of this with the interior spaces (DFT) in instants I3 and I4. The final volume is formed when the interior space (DFL) and path (DFT) are resolved (instant I5). The final opacity/transparency emerges from the gaps that arise as the direct result of the previous placement of the rooms.

Fig. 10 (second row) represents an example of pathway 2. The design starts at instant I1 with volumetric growth (VG), which is also used in the rest of the design and the subsequent creation of the first room (DFL). Then, in instants I2 to I4, the routes (DFT) and the access (DFA) are created as a natural result of the spaces that arise between the volumes. The opacity/transparency of the shelter is also determined by the separations between volumes. In this way, natural light is introduced in the shelter without needing to create dedicated holes.

Fig. 10 (third row) shows an example of pathway 3. In this case, laminar growth (LG) is shaping the design starting from the beginning in instants I1 and I2 (like the example of pathway 1). Volumetric growth (VG) is postponed to instant I3. The access (DFA) and the connection with the other spaces (DFL) are designed.

In instants I3 and I4, therefore, the participant alternates the design of the interior and exterior. As a result of using this strategy based on laminar growth, the shelter shows a high degree of transparency.

In summary, most of the participants (clusters 1 and 3) use a design route that begins with a laminar growth (based on planes with which they create walls or floors), that allows them to have greater control of their design. Differently, the participants in cluster 2 start with volumetric growth, which is more complex to control. In both cases, when volumetric growth begins, the interior spaces arise, and the designers need to work on the access and its connection with the interior spaces. These events take place mostly at intermediate moments. It seems that if the functional fact of access and connection to interior spaces is not addressed at intermediate moments, it could be difficult to complete the spatial relationships that determine the design project.

Concerning the process for architectural design, common solution paths can serve as useful guidelines. In our study, the discovered pathways allow differentiating the functional relationships (in this case, access to the shelter and an inner route that serves the different rooms) from the spatial ones, something that is always difficult in the design process. In addition, they also allow the designers to differentiate and establish the design instants, being aware of which elements appear in the process and when and promoting a certain control of the object that is being designed or projected.

To sum up, three different problem-solving pathways have been identified inside the 52 projects developed by novice designers. These pathways or clusters can be described qualitatively in terms of design decisions taken along the design process. Therefore, and concerning RQ2, we have been able to find common pathways or typologies between the design tracks of the projects. As we mentioned above, such pathways can help the designers to face the complex nature of the design problems.

6. Conclusions

The goal of this paper was to develop and test a new methodology for design tracking to allow for a better understanding of the design processes from their early stages.

To this end, an empirical study with fourth-year architecture students was conducted. Each student had to design a shelter using a design software tool. The participants had to provide snapshots of their projects at six design moments of their creative process. Tracking of the projects was based on the analysis of certain variables or criteria, which were collected from the snapshots taken in the six selected design instants. Machine learning techniques were then applied the values collected.

With these techniques, two analyses were carried out. In the first one, supervised classification models were used, and a decision tree was constructed. This decision tree was able to characterize the class of the 'best' projects, i.e., projects that obtained the best grades. It shows that the most discriminating variable was the creation of the route from the shelter's entrance to the interior rooms at the intermediate instant of the design process. So, it seems that the consideration of this functional fact in the central moment of the design process is key to completing spatial relationships coherently.

In the second analysis, unsupervised learning techniques were used with the objective of grouping the projects into clusters that share typologies or common problem-solving pathways. After applying hierarchical clustering algorithms, three pathways or clusters were identified, in which the designers demonstrated some common characteristics in their design processes. These pathways can help designers establish different and specific behaviour patterns which allow the designers to clarify the process rationally, from the moment they establish and make visible a certain order of action. In addition, they also foster analytical reasoning in that they allow structuring the design process, always establishing the final object as a sum of parts and partial aspects.

The findings of this study must be seen considering some limitations and difficulties, which we discuss next.

First, the sample size in our study is 52 (fourth-year architecture students). For sure, it would be desirable to have a larger set of data, but organizational and human resources constraints make it difficult to obtain such a set. Nevertheless, the same problem is also present in related studies of design tracking. For example, the sample sizes of Bilda and Demirkan [2], Rodgers et al. [10], Suwa and Tversky [7], and Jain and Sobek [33] were 6, 3, 9, and 14 respectively. In comparison, it could be argued that our sample size is not so small, given its research context.

Second, although the analysis of the collected data suggests that our methodology can be used to predict the quality of designs and to identify common pathways, the design scenario addresses a unique and simple design task that deals with a small set of well-defined requirements and variables. In real design scenarios, where design variables and requirements are often multiple and sometimes ill-defined, performing the same analyses would require a more complex methodology and involve more sophisticated tools, and it will probably result in less structured data that will need to be processed with alternative machine learning techniques which might produce different findings.

Third, we must acknowledge the main difficulties found in our experiment and discuss its reproducibility for other interested researchers. We think that our methodology could be suitable to analyse other design processes, as long as they can be decomposed into stages that can be described as variables (to allow for their quantitative analysis). In this sense, the main difficulties we have found in our experiment are a) the sample size, and b) the tracking design method. As for the sample size, the design process is complex and takes time, and it can be hard to find participants willing to take part in the experiment. Possible alternatives include finding ways to compensate participants or using available datasets. As for the tracking design method, it should provide the support needed in the tracking process. In our case, we have implemented a SketchUp plugin (as explained in Section 3.1.2), but alternative solutions could be using snapshots taken during freehand/digital sketching and/or using elements like think-aloud transcripts or video records (as discussed in some related works cited in Section 2) or developing targeted software tools to better suit the needs of the new studies.

From our point of view, this study makes at least two important contributions. First, it describes a novel approach to design tracking. We use an interactive design software tool that allows collecting information at selected design instants and then extracting the knowledge contained at the snapshots of the designs in those instants, using data analysis and machine learning techniques. An additional advantage is that these analysis techniques may suggest a quantitative approach in the design field instead of the traditional qualitative approach. We hope our study can serve as an inspiration for further studies on the topic dealing with more complex and realistic design environments. Second, we provide evidence that these techniques can identify 'good' design processes and pathways, that can help designers to face the complexity of these processes and achieve better results. In summary, our work illustrates how powerful statistical techniques can be used to elucidate strategies and relationships in design processes. Our findings could also be of interest when implementing new computer tools to support the design process and contribute to the possibility of using data analysis techniques in other design contexts.

Concerning future work, and departing from the results reported here, we intend to continue our research and achieve more substantiated results by performing further experiments and studies that involve more complex experiments, including different requirements, variables and/or analysis tools.

CRediT authorship contribution statement

Eva Millán: Conceptualization, Methodology, Software, Writing – review & editing, Writing – original draft, Formal analysis, Validation. **María-Victoria Belmonte:** Conceptualization, Methodology, Software, Writing – review & editing, Writing – original draft, Formal analysis, Validation. **Francisco-Javier Boned:** Conceptualization, Methodology, Writing – review & editing, Writing – original draft, Resources. **Juan Gavilanes:** Conceptualization, Methodology, Writing – review & editing, Writing – original draft, Resources. **José-Luis Pérez-de-la-Cruz:** Conceptualization, Methodology, Software, Writing – review & editing, Writing – original draft, Funding acquisition, Formal analysis, Supervision, Validation. **Carmen Díaz-López:** Conceptualization, Methodology, Writing – review & editing, Writing – original draft, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Manuela Ruiz-Montiel for the implementation of the software tool used in this research work. This work has been partially funded by the Spanish Government, Agencia Estatal de Investigación (AEI), and the European Union, Fondo Europeo de Desarrollo Regional (FEDER), grant TIN2016-80774-R (AEI/FEDER, UE). Funding for open access charge: Universidad de Málaga/ CBUA.

References

[1] P. Heidari, Ç. Polatoğlu, Pen-and-paper versus digital sketching in architectural design education, Int. J. Architect. Comput. 17 (2019) 284–302, https://doi.org/10.1177/1478077119834694.

[2] Z. Bilda, H. Demirkan, An insight on designers' sketching activities in traditional versus digital media, Des. Stud. 24 (2003) 27–50, https://doi.org/10.1016/S0142-694X(02)00032-7.

[3] M. Brösamle, C. Hölscher, Approaching the architectural native: a graphical transcription method to capture sketching and gesture activity, Des. Stud. 56 (2018) 1–27, https://doi.org/10.1016/J.DESTUD.2018.01.002.

[4] M. Kavakli, M. Suwa, J. Gero, T. Purcell, Sketching Interpretation in Novice and Expert Designers, Visual and Spatial Reasoning in Design, 1999, pp. 209–220.

[5] D.A. Schön, G.E. Wiggins, Kinds of seeing and their functions in designing, Des. Stud. 13 (1992) 135–156.

[6] M. Suwa, T. Purcell, J. Gero, Macroscopic analysis of design processes based on a scheme for coding designers' cognitive actions, Des. Stud. 19 (1998) 455–483, https://doi.org/10.1016/S0142-694X(98)00016-7.

[7] M. Suwa, B. Tversky, What do architects and students perceive in their design sketches? A protocol analysis, Des. Stud. 18 (1997) 385–403, https://doi.org/10.1016/S0142-694X(97)00008-2.

[8] M. Prats, S. Lim, I. Jowers, S.W. Garner, S. Chase, Transforming shape in design: observations from studies of sketching, Des. Stud. 30 (2009) 503–520, https://doi.org/10.1016/J.DESTUD.2009.04.002.

[9] A. McGown, G. Green, P.A. Rodgers, Visible ideas: information patterns of conceptual sketch activity, Des. Stud. 19 (1998) 431–453. https://www.academia.edu/3773031/Visible_ideas_information_patterns_of_conceptual_sketch_activity. (Accessed 25 November 2021).

[10] P.A. Rodgers, G. Green, A. McGown, Using concept sketches to track design progress, Des. Stud. 21 (2000) 451–464.

[11] G. Goldschmidt, P.A. Rodgers, The design thinking approaches of three different groups of designers based on self-reports, Des. Stud. 34 (2013) 454–471, https://doi.org/10.1016/J.DESTUD.2013.01.004.

[12] I. Serrano Lasa, M. Mazmela Etxabe, I. Ricondo Iriondo, New application for sketching in a machine tool company, J. Eng. Des. (2021) 1–31, https://doi.org/10.1080/09544828.2021.1976398.
[13] G. Zhu, Y. Zeng, W. Xing, H. Du, C. Xie, G. Zhu, Y. Zeng, W. Xing, H. Du, C. Xie, Reciprocal relations between students' evaluation, reformulation behaviors, and engineering design performance over time, JSEdT 30 (2021) 595–607, https://doi.org/10.1007/S10956-021-09906-3.
[14] L.F. Gül, The impact of digital design representations on synchronous collaborative behaviour, J. Inf. Technol. Construct. 19 (2014) 47–71.
[15] H.S. Salman, R. Laing, A. Conniff, The impact of computer aided architectural design programs on conceptual design in an educational context, Des. Stud. 35 (2014) 412–439, https://doi.org/10.1016/J.DESTUD.2014.02.002.
[16] H.-H. Tang, Y.-Y. Lee, W.-K. Chiou, C.M. Herr, H.-H. Chiu, H.-J. Hu, J. Yen, Interface design (6a) Is an on-virtu digital sketching environment cognitively identical to in-situ free-hand sketching? 473 appropriating Donald Schön 483 research on hybrid tectonic methodologies for responsive architecture 493 conceptual model for design team toward website, Construction 503 (2009).
[17] H.H. Tang, Y.Y. Lee, J.S. Gero, Comparing collaborative co-located and distributed design processes in digital and traditional sketching environments: a protocol study using the function-behaviour-structure coding scheme, Des. Stud. 32 (2011) 1–29, https://doi.org/10.1016/j.destud.2010.06.004.
[18] M.A. Neroni, A. Oti, N. Crilly, Virtual Reality design-build-test games with physics simulation: opportunities for researching design cognition, Int. J. Design Creativity Innovat. 9 (2021) 139–173, https://doi.org/10.1080/21650349.2021.1929500.
[19] V. Goel, Sketches of thought, sketches of thought. https://doi.org/10.7551/MITPRESS/6270.001.0001, 1995.
[20] G. Goldschmidt, Linkography: unfolding the design process, Linkography (2014), https://doi.org/10.7551/MITPRESS/9455.001.0001.
[21] J. Clarkson, C. Eckert, Design Process Improvement: A Review of Current Practice, Design Process Improvement: A Review of Current Practice, 2005, pp. 1–560, https://doi.org/10.1007/978-1-84628-061-0.
[22] L. Zhang, L. Chen, Z. Wu, S. Zhang, H. Song, Investigating young consumers' purchasing intention of green housing in China, Sustainability 10 (2018) 1044, https://doi.org/10.3390/su10041044.
[23] E. Sung, T.R. Kelley, Identifying design process patterns: a sequential analysis study of design thinking, Int. J. Technol. Des. Educ. 29 (2019) 283–302.
[24] T.F. Schubert, F.G. Jacobitz, E.M. Kim, Student perceptions and learning of the engineering design process: an assessment at the freshmen level, Res. Eng. Des. 23 (2012) 177–190, https://doi.org/10.1007/S00163-011-0121-X/FIGURES/9.
[25] M. Zandvoort, N. Kooijmans, P. Kirshen, A. van den Brink, Designing with pathways: a spatial design approach for adaptive and sustainable landscapes, Sustainability 2019 11 (2019) 565, https://doi.org/10.3390/SU11030565, 565.
[26] I. Loche, C. Bleil de Souza, A.B. Spaeth, L.O. Neves, Decision-making pathways to daylight efficiency for office buildings with balconies in the tropics, J. Build. Eng. 43 (2021) 102596, https://doi.org/10.1016/J.JOBE.2021.102596.
[27] T. Palmieri, L. Huybrechts, O. Devisch, Co-producing, curating and reconfiguring dwelling patterns: a design anthropological approach for sustainable dwelling futures in residential suburbs, Des. Stud. 74 (2021) 101011, https://doi.org/10.1016/J.DESTUD.2021.101011.
[28] B. Reimlinger, Q. Lohmeyer, R. Moryson, M. Meboldt, A comparison of how novice and experienced design engineers benefit from design guidelines, Des. Stud. (2019).
[29] R. Bailey, Z. Szabo, Assessing Engineering Design Process Knowledge*, (n.d.).
[30] H. Sun, H.v. Burton, H. Huang, Machine learning applications for building structural design and performance assessment: state-of-the-art review, J. Build. Eng. 33 (2021) 101816, https://doi.org/10.1016/J.JOBE.2020.101816.
[31] S. Chang, W. Dong, H. Jun, Use of electroencephalogram and long short-term memory networks to recognize design preferences of users toward architectural design alternatives, J. Comput. Des. Eng. 7 (2020) 551–562, https://doi.org/10.1093/JCDE/QWAA045.
[32] H.Z. Zhang, C. Xie, S. Nourian, Are their designs iterative or fixated? Investigating design patterns from student digital footprints in computer-aided design software, Int. J. Technol. Des. Educ. 28 (2018) 819–841, https://doi.org/10.1007/S10798-017-9408-1.
[33] V.K. Jain, D.K. Sobek, Linking design process to customer satisfaction through virtual design of experiments, Res. Eng. Des. 17 (2006) 59–71, https://doi.org/10.1007/S00163-006-0018-2/TABLES/7.
[34] C.M. Castellan, Quantitative and qualitative research: a view for clarity, Int. J. Educ. 2 (2010) 1, https://doi.org/10.5296/IJE.V2I2.446.
[35] M.V. Belmonte, E. Millán, M. Ruiz-Montiel, R. Badillo, J. Boned, L. Mandow, J.L. Pérez-De-La-Cruz, Randomness and control in design processes: an empirical study with architecture students, Des. Stud. 4 (2014) 392–411, https://doi.org/10.1016/J.DESTUD.2014.01.002.
[36] G. Stiny, Introduction to shape and shape grammars, Environ. Plann. Plann. Des. 7 (1980) 343–351, https://doi.org/10.1068/b070343.
[37] J. Rykwert, On Adam's House in Paradise: the Idea of the Primitive Hut in Architectural History, The Museum of Modern Art, 1972.
[38] Gustau Gili Calfetti, Casas Refugios = Private Retreats, Gustavo Gili, Barcelona, 1996.
[39] J. Pallasmaa, Identity, Intimacy and Domicile - notes on the phenomenology of home, in: N.D. Benjamin (Ed.), The Home; Words, Interpretations, Meanings and Environments. Avebury, Aldershot, 1995, pp. 131–147.
[40] G.S. Adams, B.A. Converse, A.H. Hales, L.E. Klotz, People systematically overlook subtractive changes, Nature 2021 592:7853 592 (2021) 258–261, https://doi.org/10.1038/s41586-021-03380-y.
[41] G. Semper, The Four Elements of Architecture and Other Writings, Cambridge University Press, Cambridge [England]; New York, 1989.
[42] B. Zevi, The Modern Language of Architecture, first ed., University of Washington Press, 1978. https://openresearch-repository.anu.edu.au/handle/1885/115190?mode=full. (Accessed 25 November 2021).
[43] J.M. Montaner Martorell, Las formas del siglo XX, 2002. https://dialnet.unirioja.es/servlet/libro?codigo=219497. (Accessed 25 November 2021).
[44] 1887-1965. le Corbusier, J. Goodman, Toward an architecture, Frances Lincoln, 2008.
[45] S.E. Rasmussen, Experiencing Architecture, second ed., The MIT Press, 1964.
[46] M.G. Bevilacqua, A. Klein, Alexander klein and the existenzminimum: a 'scientific' approach to design techniques, Nexus Network J. 2011 13:2 13 (2011) 297–313, https://doi.org/10.1007/S00004-011-0080-6.
[47] E. Acuña, C. Rodriguez, The treatment of missing values and its effect on classifier accuracy, classification, clustering, and data mining applications. https://doi.org/10.1007/978-3-642-17103-1_60, 2004, 639-647.
[48] Witten, Frank, Eibe, Data Mining: Practical Machine Learning Tools and Techniques, Second Edition, (n.d.).
[49] T. Hastie, R. Tibshirani, J. Friedman, The elements of statistical learning. https://doi.org/10.1007/978-0-387-84858-7, 2009.
[50] M. Mächler, P. Rousseeuw, A. Struyf, M. Hubert, K. Hornik, Cluster: cluster Analysis basics and extensions. R package version 2.1.2. https://CRAN.R-project.org/package=cluster, 2021. (Accessed 25 November 2021).
[51] G. Brock, V. Pihur, S. Datta, S. Datta, clValid: an R package for cluster validation, J. Stat. Software 25 (2008) 1–22, https://doi.org/10.18637/jss.v025.i04.