Received: 23 November 2023 | Revised: 1 March 2024 | Accepted: 14 April 2024

DOI: 10.1002/cae.22749

RESEARCH ARTICLE

WILEY

Beyond analytics: Using computer-aided methods in educational research to extend qualitative data analysis

Camilo Vieira[1] | **Juan D. Ortega-Alvarez**[2,3] | **Alejandra J. Magana**[4] | **Mireille Boutin**[5,6]

[1]Department of Education, Universidad del Norte, Barranquilla, Colombia

[2]Department of Engineering Education, Virginia Tech, Blacksburg, Virginia, USA

[3]Escuela de Ciencias Aplicadas e Ingeniería, Universidad EAFIT, Medellín, Colombia

[4]Department of Computer and Information Technology and School of Engineering Education, Purdue University, West Lafayette, Indiana, USA

[5]Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, Netherlands

[6]Department of Mathematics, Purdue University, West Lafayette, Indiana, USA

Correspondence
Camilo Vieira, Department of Education, Universidad del Norte, Barranquilla 080020, Colombia.
Email: cvieira@uninorte.edu.co

Funding information
National Science Foundation

Abstract

This study proposes and demonstrates how computer-aided methods can be used to extend qualitative data analysis by quantifying qualitative data, and then through exploration, categorization, grouping, and validation. Computer-aided approaches to inquiry have gained important ground in educational research, mostly through data analytics and large data set processing. We argue that qualitative data analysis methods can also be supported and extended by computer-aided methods. In particular, we posit that computing capacities rationally applied can expand the innate human ability to recognize patterns and group qualitative information based on similarities. We propose a principled approach to using machine learning in qualitative education research based on the three interrelated elements of the assessment triangle: cognition, observation, and interpretation. Through the lens of the assessment triangle, the study presents three examples of qualitative studies in engineering education that have used computer-aided methods for visualization and grouping. The first study focuses on characterizing students' written explanations of programming code, using tile plots and hierarchical clustering with binary distances to identify the different approaches that students used to self-explain. The second study looks into students' modeling and simulation process and elicits the types of knowledge that they used in each step through a think-aloud protocol. For this purpose, we used a bubble plot and a k-means clustering algorithm. The third and final study explores engineering faculty's conceptions of teaching, using data from semi-structured interviews. We grouped these conceptions based on coding similarities, using Jaccard's similarity coefficient, and visualized them using a treemap. We conclude this manuscript by discussing some implications for engineering education qualitative research.

KEYWORDS

assessment triangle, clustering algorithms, computer-aided methods, data analysis, education research, machine learning, qualitative methods, visualization

1 | INTRODUCTION

New inquiry fields in education have called for the creation or extension of formal models and methodologies for education research and evaluation. Computer-aided methods present a significant potential to supplement, extend, or strengthen traditional methods of education research [13, 18, 36]. Over the last two decades, computation has become an increasingly important tool in scientific inquiry and engineering innovation. Some have denominated computational science as the third pillar of science along with the theoretical and experimental approaches [43]. Computation allows researchers to process large amounts of data, create visual representations of phenomena, and represent complex phenomena in the form of models and simulations [33, 42].

With the large amount of data that we collect today and the complex problems that we need to solve in a globalized world, many disciplines have seen the emergence of subdisciplines with the adjective "computational" in front of them. Computational biology (or Bioinformatics), Computational Materials Science, and Computational Social Sciences are some examples of these new disciplines [20, 33]. Engineering education and the field of educational research have also been part of this trend. An example of this phenomenon in education is the emergence of specialized academic conferences and journals in learning analytics, computer applications, and educational data mining [38].

The use of computation in educational research, however, is not limited to learning analytics or educational data mining [26]. The increasingly large computing capacity and the development of computational techniques such as automatic classification methods, natural language processing, and pattern recognition techniques offer new approaches to analyzing the data that have been traditionally collected by educational researchers (e.g., performance, perceptions, transcripts) [1]. However, an important consideration to implement novel research approaches is that these need to be scientifically based. In particular, (a) selecting appropriate and effective methods for addressing research questions and (b) providing a clear explanation of procedures and valid conclusions that confirm scientific inferences via explicit chains of reasoning are two important elements when it comes to methodological considerations [23]. This paper proposes a rationale and guidelines to use computer-aided methods in qualitative research and provides three examples of studies where different approaches of visualization and clustering have been used for analyzing qualitative data in engineering education research. This paper contributes to the body of knowledge of computer-aided qualitative research methods by proposing a conceptual framework that may guide researchers in selecting, using, and interpreting the outcomes of a principle-based method for a specific data set and purpose. The manuscript demonstrates different affordances of these methods to support qualitative data analysis using three case studies.

2 | BACKGROUND

In 2017, 2019, and 2020, two of the authors of this paper prepared and taught special sessions related to the use of pattern recognition techniques to analyze educational data at engineering education conferences [35, 37]. The sessions were designed to help educational researchers familiarize themselves with emergent computer-aided methods that could provide different affordances to analyze educational data. Specifically, we suggested that the pattern recognition process and machine learning methods could support educational research through four main approaches: exploration (visualization), classification, grouping, and validation.

A common qualitative data analysis process starts from raw data such as text (e.g., transcripts or written documents), images (e.g., drawings), or audio and video recordings (e.g., interviews, focus groups). The researcher will often assign categories or codes that will help them to start characterizing the phenomenon. This leads to a high-dimensional data set where each item in the data set has several categories assigned to it, and it is often difficult to interpret. The next step is to find themes from this high-dimensional data set to describe and characterize how participants experience/describe/conduct a given phenomenon. This process often groups participants within these themes and the resulting groups must be validated and characterized.

In parallel, a pattern recognition process starts from raw data in the form of pictures, text, audio, or video. For example, a picture is represented as a high-dimensional data point where each pixel is represented by three colors: red, green, and blue. In order to recognize patterns (e.g., the difference between a face and the rest of the body), a set of feature vectors (i.e., vectors with relevant information for the task) are identified to find groups in the data set. These groups are later validated and characterized with additional data or with experts analyzing the results in comparison with the original data set. Clustering algorithms that generate these groups are part of the family of unsupervised machine learning algorithms. Figure 1 illustrates the commonalities and complementarity of the data analysis and the pattern recognition processes.

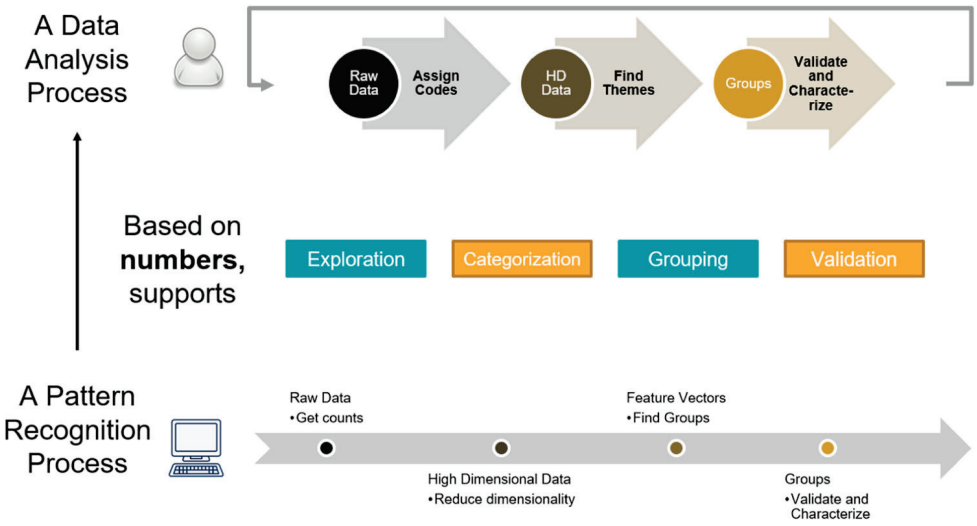


FIGURE 1 A parallel between a pattern recognition process and a qualitative data analysis process.

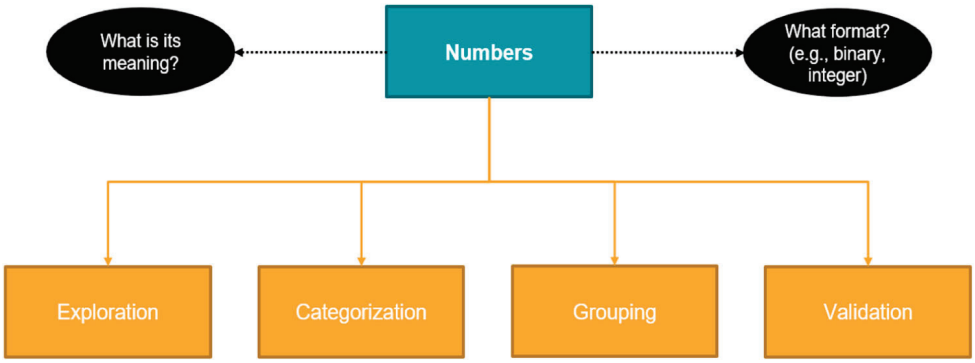


FIGURE 2 Affordances of using pattern recognition techniques to analyze quantified qualitative data.

The key to success in this collaboration between the pattern recognition process and the qualitative data analysis process lies in transforming attributes or relations within the data set into numbers. There are different ways in which one can transform qualitative data into numbers, and this choice will have an effect on the methods that one will be able to use. Figure 2 highlights key questions that need to be addressed to decide what visualization technique or clustering algorithm can be used for a given purpose. For instance, one can count the number of times that a given category is present in a transcript using content analysis and would end with positive integers. This would require a specific type of visualization that enables the researcher to compare the number of occurrences for a given category (e.g., a bar plot or bubble plot [5]). Alternatively, one can analyze whether a category is present (1) or not (0), which will make the variable binary. In this case, rather than comparing magnitudes, the researcher may prefer to use a visualization that shows what categories emerge for each participant and what pattern may emerge with these categories among participants. Likewise, natural language processing techniques use different approaches to turn text into numbers (e.g., the bag of words counts the number of times each word appears in a document, while the term frequency–inverse document frequency compares among documents).

The meaning and the format of the resulting number will have an implication on the methods that one can use for exploration (i.e., what approach to visualization would be most meaningful for exploration), for categorization (e.g., whether we want to understand the most popular word in a set of tweets or we want to understand the quality of an argument in students' reflection), for grouping (i.e., different clustering methods use different metrics of distance), and for validation (e.g., the underlying assumptions for a permutation test).

2.1 | Conceptual framework

The conceptual framework for our study is the assessment triangle [11]. The assessment triangle has traditionally been used in education to design valid assessments [21], such as the design of concept inventories [31]. However, the assessment triangle has also been used as a conceptual framework for guiding investigations [17].

A recent systematic literature review investigating how machine learning has been used to supplement science assessment identified existing research gaps in considering or making such alignment explicit [44]. For this study, as Magana and Boutin [18] proposed, the assessment triangle is used as a framework to follow a principle-based method for integrating computer-aided methods into qualitative research methods. By adopting a principle-based method facilitated by the assessment triangle, investigators can make explicit connections between their study's theoretical or pedagogical foundations with the construct being measured and the observations and interpretations of their findings under their selected theoretical lens.

The assessment triangle consists of an approach where three constructs need to be explicitly aligned in the process of creating assessments or, similar to this case, the process of analyzing and interpreting educational data. The triangle comprises three corners: cognition, observation, and interpretation [25]. Specifically, the cognition corner suggests making explicit the theoretical foundations of the phenomenon being measured or characterized. This involves aligning a theory or conceptual framework that describes or hypothesizes how learners develop knowledge or behaviors conducive to learning or other desired outcomes. For instance, when applying the assessment triangle to inquiries not related to learning, the cognition corner can also include theories supporting people's conceptualizations or perceptions of educational issues. Once constructs are grounded in the proper learning theory or a theoretical or conceptual framework, the next step involves identifying the type of overt activities or artifacts that constitute evidence of the learning, performance, behavior, or conceptualization, leading us to the observation corner. The observation corner allows us to operationalize or characterize performances, behaviors, conceptualizations, or evidence of learning. The chosen mechanism of evidence should be aligned with the cognition corner. This step also involves the identification of the collection and interpretation of the data. The data can take any form, including text, video, or interview transcripts. Once the data have been collected in the corresponding format, the next step is to characterize that evidence in terms of the construct being investigated, thus transitioning to the interpretation corner. The interpretation corner deals with how results derived from the observation corner are analyzed, validated, and interpreted.

The following sections will present three studies previously conducted by the authors that used visualization and clustering methods to analyze qualitative data in engineering education research. In all these three studies, visualization and clustering techniques contributed to providing insights and managing high-dimensional data sets derived from an initial manual qualitative coding process. The first study explored the characteristics of students' written explanations of programming worked-examples, and used binary numbers to represent whether a category was present (1) or not (0) for a given explanation. Thus, this study used a tile plot and hierarchical clustering with binary distances [32] for visualization and grouping purposes. The second study used content analysis [14] to explore how students used different types of knowledge during a think-aloud protocol where they were completing a computational modeling task. As content analysis usually counts the number of instances that each category appears, this study used the bubble plot [5] and the k-means clustering method [12] for exploration and grouping. These two studies used the R programming language to create visualizations, preprocess the data (e.g., turning categories into binary numbers), and implement the clustering algorithms. The third study aimed at grouping engineering faculty into clusters based on their conceptions of teaching (CoT). CoT were determined from the transcripts of semi-structured interviews focused mainly on participants' views of multiple dimensions including the role of the teacher, the role of the student, the purpose of assessment, and the nature of knowledge. In this case, the research team used the software Nvivo® to form clusters based on coding similarities by computing Jaccard's similarity coefficient [22] and to create hierarchy tables to determine the major commonalities across participants clustered together. Examples of automatic categorization and validation are beyond the scope of this manuscript, but can be found in [4, 10, 36]. For instance, Gillies and colleagues [10] suggest that natural language processing techniques such as topic modeling may support qualitative data analysis by providing early categories that researchers may interpret and refine, while initial qualitative data analysis may improve topic modeling outcomes. This paper focuses on demonstrating how computer-aided visualization and grouping (i.e., clustering) may support qualitative data analyses in educational research.

2.2 | Study one: Analyzing students' written explanations of programming worked-examples

This study took place in the context of a Computation and Programming for Materials Science and Engineering course. In this course, first-year undergraduate students learned computation and programming skills situated in the context of their engineering discipline. Twenty-four students in this course participated in this study. The goal of this first study was to understand what approaches undergraduate students used to self-explain a set of computational modeling worked examples in the context of a computational science course [40, 41]. This represents the cognition corner, where we looked for instances of declarative, procedural, schematic, and strategic knowledge, adapting Shavelson and colleagues' [29] types of knowledge framework for assessing learning. The worked examples included a problem statement and a step-by-step solution [2], including the programming code in MATLAB®. We collected data asking students to write in-code comments to self-explain how the code was solving the problem, the observation corner. The first four self-explaining activities of the semester were graded, while nine additional activities were proposed for extra credit. In total, 24 undergraduate students enrolled in this course and completed the self-explaining activities analyzed in this study.

We hypothesized that students' explanations would include different components for different parts of the code. For instance, when the code defines a function, this function has a goal and it often includes one or more input parameters and one or more outputs. Conversely, when the program defines a loop statement, it includes a condition that has to be met in order to continue executing the code inside. Hence, we hypothesized that students would discuss issues like the input parameters and outputs in the part of the code where a function is defined, and the stopping condition of a loop statement. To account for this, we analyzed each section of the worked example separately from each other and assigned categories to the explanations of each section. For instance, if the explanation describes the consequences of executing a section, we would assign the code Consequences of Actions (COA). Likewise, if it describes the parameters of a function, we would assign the code Parameters (PAR). The full coding scheme can be accessed at [41].

An example of the resulting data set is presented in Figure 3, where each row represents a student and each column represents a section. These data are difficult to understand in this format, so we created a binary table as shown in Figure 4 for each column from Figure 3 (i.e., each section). Figure 4 only shows Section 1 as an example of how we processed this information. We then created a visualization for each section using a tile plot (Figure 5). Figure 5 depicts a "Creating a Function" section and an "Iterating" section (i.e., Sections 1 and 3), which show that students are actually writing different types of explanations for different types of sections. This also shows that some types of explanations are common to some sections, but different students used different approaches to self-explain. We used R programming language for transforming and visualizing the data.

The next step was to group students based on their different approaches to self-explaining. We used a hierarchical clustering method with binary distance [32] to identify students who used similar approaches and group them. The binary distance compares where two students coincide and represents that with a one

```
% Section 1 - Creating the Function
function bondmat = atomicbonds(pos, cutoff)
    % Section 2 - Setting up Problem Parameters
    N = size(pos,1);
    % Section 3 - Setting up Supporting Variables
    bondmat=sparse(N,N);
    % Section 4 - Iterating
    for n = 1:N-1
        for m = n+1:N
            % Section 5 - Setting up Problem Parameters
            dist = pos(m,:)-pos(n,:);
            len = norm(dist);
            % Section 6 - Validation
            if len < cutoff
                bondmat(n,m)=len;
            end
        % Section 7 - End of the Function
        end
    end
end
```

Std	Section 1	Section 2	Section 3
S1	COA, PAR	COA, PRO, HOW, VAR	VAR, BGK
S2	COA, PAR, VAR, PRO, WHY, BGK	COA, RAG, HOW, BGK, VAR, WHY, PRO	COA, GOA, PRO, HOW, BGK, VAR
S3	COA	HOW, COA, PRO	COA, WHY, VAR
S4	COA	HOW, COA, PRO	COA, CON, RAG
S5	COA, PAR	PHR, EXE, HOW, CON	COA, VAR, EXE, BGK
S6	COA, PAR, GOA, PRO	COA, VAR, PRO	COA, VAR, INC
S7	COA, PAR, GOA, PRO	COA, PAR, PRO	COA, CHK, VAR, CON, BGK
S8	COD, EXE, GOA, PRO	COA, PRO, WHY, HOW	VAR, WHY
S9	COA	COA, VAR, PRO, HOW	COA, VAR
S10	COA, GOA	HOW, PAR, WHY	RAG, BGK, COA, CON, WHY, VAR
S11	OWN, EXE	OWN, COA, HOW	OWN, COA, VAR
S12	COA, GOA, PRO, PAR, VAR	COA, PRO, HOW, BGK, WHY, VAR	COA, PRO, WHY, VAR, CHK, RAG
S13	COA, PAR, VAR	COA, PRO, HOW, WHY	COA, RAG, BGK, VAR
S14	GOA, PAR	COA, VAR, PRO	COA, PAR,
S15	COA	COA, PRO, VAR	COA, CHK, RAG
S16	COA, GOA, BGK, VAR, PAR, WHY, PRO	COA, PRO, VAR, WHY, BGK, HOW, RAG	PRO, RAG, BGK, VAR, PAR, HOW
S17	OWN, PRO, VAR	OWN, HOW, VAR, WHY	OWN, COA, VAR, BGK
S18	OWN, COA, PAR, PRO	OWN, COA, VAR, HOW, PRO	OWN, COA, VAR
S19	OWN, INS, COA	OWN, COA, HOW, PAR	OWN, RAG, BGK, VAR, WHY
S20	COA, PRO, HOW, WHY	VAR, HOW	VAR, BGK
S21	PAR, VAR, HOW	COA, HOW	HOW, COA, CON, WHY, BGK
S22	PAR	COA, VAR, HOW, PRO	COA, HOW, CHK, PRO
S23		COA, PRO, HOW, WHY	COA, RAG, BGK, CON, WHY
S24	OWN, COA, PAR	OWN, COA, HOW, PRO	OWN, VAR, BGK

FIGURE 3 Dividing the code by section to assign categories.

STD	SIM	INC	LIM	PHR	COA	VAR	DAT	PAR	COD	HOW	EXE	PRO	GOA	BGK	WHY	RAG	INS	CON	MON	CHK	OWN
S15	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S19	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S14	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0
S1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
S3	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
S7	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
S8	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
S11	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
S18	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
S25	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
S5	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
S12	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
S2	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
S16	0	0	0	0	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0
S10	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
S17	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
S20	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
S22	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
S24	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1
S9	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
S21	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
S23	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
S4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

FIGURE 4 From categorization of sections to binary tables.

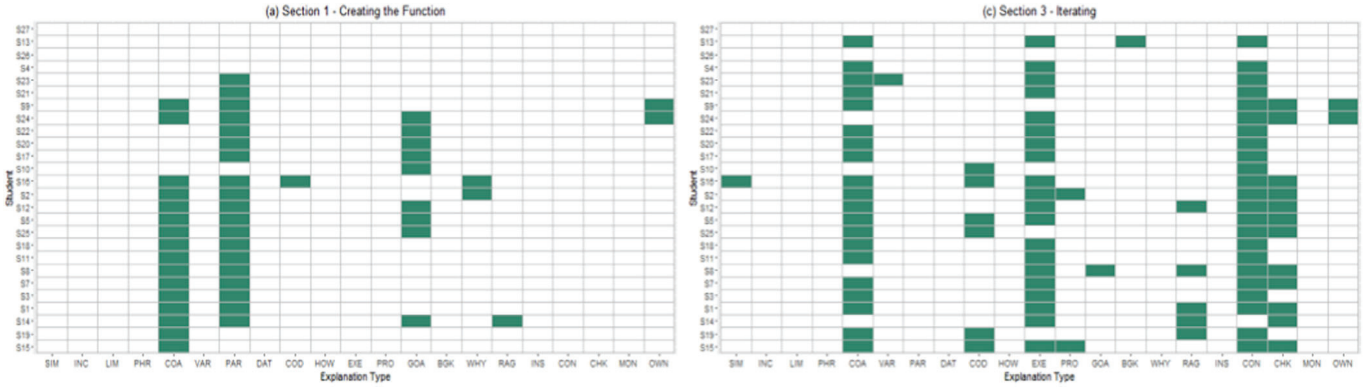


FIGURE 5 Visualization patterns from binary tables.

(1) or a zero (0), where no coincidence is found. In this case, a coincidence was found where two students had similar categories assigned for their written self-explanations. We used these groups to re-create the tile plot, reorganizing students and using colors to identify them, as depicted in Figure 6. We used the R package *ggplot2* to create these tile plots and the *hclust* function to compute the clusters.

Of course, all students in a group did not write exactly the same explanation, but this approach helped us manage the high dimensionality in the data set (i.e., number of categories multiplied by number of sections) and find clear patterns in the data. The next step in the process was to identify whether these groups are meaningful, going back to the data and to existing literature to characterize students' approaches to self-explaining. Our interpretation corner includes (1) assigning categories to students' self-explanations in a binary form; (2) creating tile plots to visualize the characteristics of students' explanations; (3) using hierarchical clustering with binary distance to group students; and (4) connecting back to literature and theories of self-explanations to make sense of the findings. This last step of the process is discussed in detail in Vieira et al. [41].

2.3 | Study two: Understanding the use of different types of knowledge in a computational modeling task

This study took place in the same context of the first study, a Computation and Programming for Materials Science and Engineering course for first-year undergraduate students. This study focused on understanding how undergraduate students developed a computational modeling activity while engaging in a retrospective think-aloud protocol [34] (observation corner). Specifically, we aimed at characterizing the computational cognitive and metacognitive knowledge that they used in the context of modeling and simulation practices. This represents our cognition corner for our conceptual framework. This study used an adapted version of the modeling and simulation process described by Shiflet and Shiflet [30] to identify instances of different types of knowledge in each step of students' problem-solving process, namely, (1) analyze the problem; (2) solve the model; (3) verify the solution; and (4) interpret, report, and maintain the solution. We also used Shavelson and colleagues' [29] types of knowledge framework to characterize (1) declarative knowledge—knowing that; (2) procedural knowledge—knowing how; (3) schematic knowledge—knowing why; and (4) strategic knowledge—knowing when, where, and under what conditions. Hence, when analyzing the think-aloud transcript for each student, we identified what stage of the modeling and simulation process students were working on, what type of knowledge they were using, and how many times (i.e., content analysis). In total, 11 students who were enrolled in this course in one of two different semesters participated in this study. Given that the methods consisted of a one-on-one think-aloud protocol, each lasting 1 h, 11 students were selected to participate in the study. The retrospective think-aloud protocol took place towards the end of the semester, once students had submitted the fourth (i.e., out of five) computational modeling challenge. As commonly practiced in qualitative studies, the small sample was deemed adequate for this study. The participating students were purposefully selected to maximize variability based on their earlier performances in the course. Five participants were female and six participants were male.

Figure 7 shows a sample data set for four of the students who participated in this study. Each table shows the number of instances they used each type of knowledge at each step of the modeling and simulation process. While some ideas may be drawn from this data set, it is difficult to understand and compare among students, steps, and types of knowledge. Instead, Figure 8 shows a bubble plot [5], where students correspond to the *y*-axis, the steps of the modeling and simulation process are crossed with the types of knowledge in the *x*-axis, and the size of each dot (i.e., bubbles) corresponds to the number of instances. This visualization shows much clearer patterns; for instance, students did not discuss much in the Maintain step, most of them did not use strategic knowledge in the Analyze step, and many did not use schematic knowledge in the Solve step.

This visualization also shows that not all students used the same types of knowledge during the modeling and simulation activity. Thus, the next step was to identify similar approaches to create groups (Figure 9). In this case, since our numbers are integers, we used the clustering algorithm k-means, which uses a Euclidean distance to compute the distance between a set of centroids [12]. These sets of centroids are generated at random at the beginning and iteratively refined based on how far/close all points are (i.e., updating the centroids to be located on the average location of the data points assigned to each centroid). Our interpretation corner for this study involved (1) using content analysis to analyze think-aloud data; (2) using bubble plots to visualize it and gain insights; (3) using k-means to group students

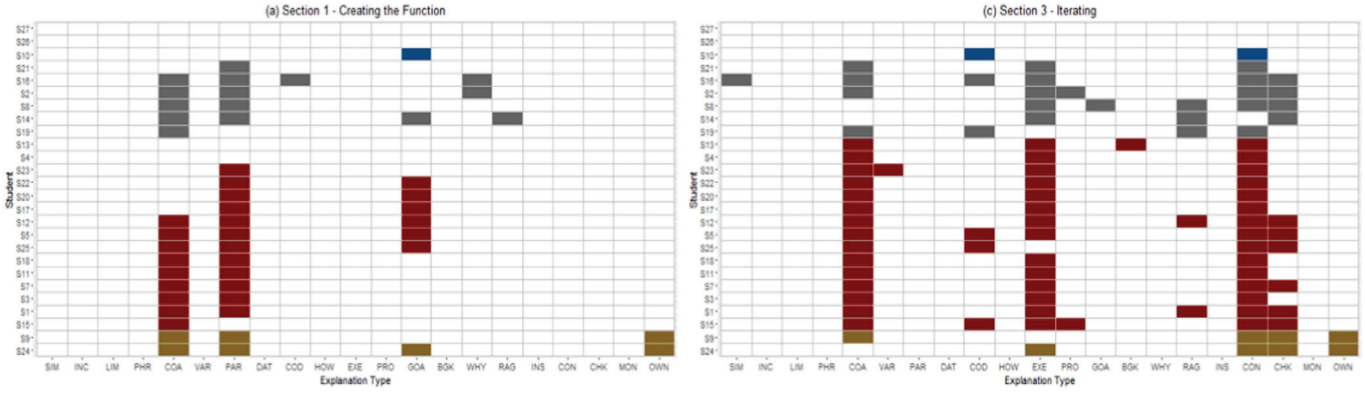


FIGURE 6 Visualization of clusters.

2014				
Student 1 (Milan_L)				
	Analyze	Solve	Verify	Maintain
Declarative	7	0	1	0
Procedural	1	5	1	0
Schematic	1	1	1	0
Strategic	1	4	1	0

2015				
(Student 1, Logan_L)				
	Analyze	Solve	Verify	Maintain
Declarative	10	7	2	0
Procedural	0	13	0	1
Schematic	0	0	1	0
Strategic	1	6	1	3

Student 2 (Lennon_L)				
	Analyze	Solve	Verify	Maintain
Declarative	9	3	2	1
Procedural	1	8	2	0
Schematic	2	0	0	1
Strategic	4	6	5	2

(Student 2, Armani_H)				
	Analyze	Solve	Verify	Maintain
Declarative	3	2	4	0
Procedural	2	10	2	3
Schematic	2	0	2	1
Strategic	2	6	4	1

FIGURE 7 Counts per student from the content analysis.

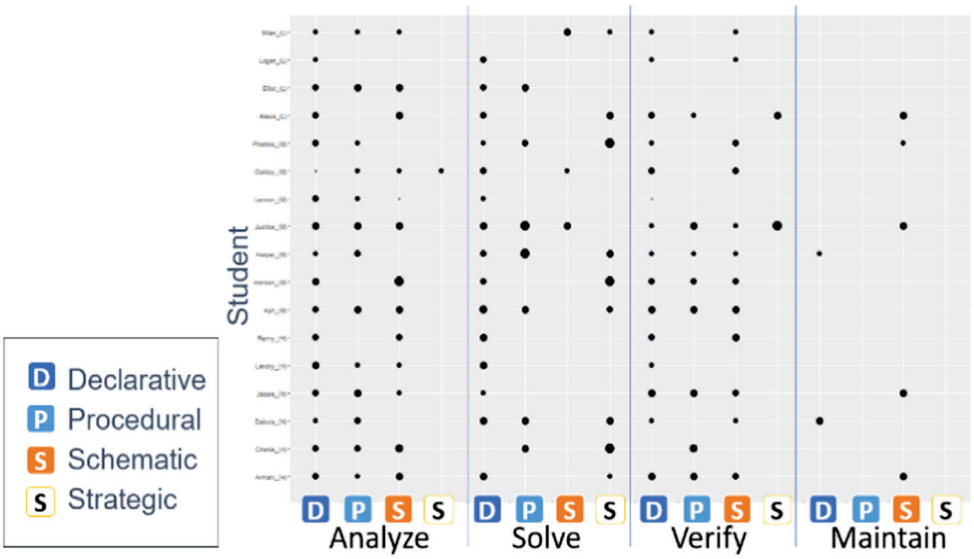


FIGURE 8 Bubble plot depicting counts per student (rows) from the content analysis.

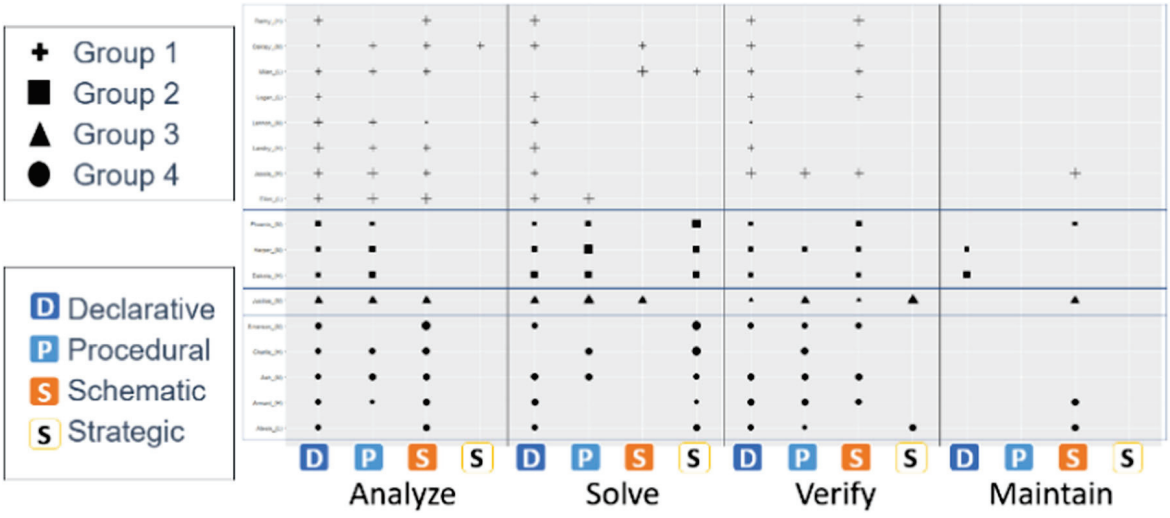


FIGURE 9 Clustered students using k-means.

based on the types of knowledge that they used; and (4) connecting back to the literature in the use of cognitive and metacognitive knowledge for modeling and simulation in engineering. In this case, however, we used this process to validate the already conducted qualitative data analysis, where clusters were created manually, and yielded an accuracy rate greater than 95% [19]. As in study one, we used ggplot2 from R programming language to create the visualizations and, in this case, we used the *kmeans* function to identify the clusters in the data set.

2.4 | Study three: Exploring conceptions of teaching among Colombian engineering faculty

As a part of a larger study, the transcripts of semi-structured interviews with 20 Colombian engineering faculty were coded to identify variations along different dimensions of their conception of teaching [24]. The participants came from two medium-sized and one large institution in Colombia, both private and public, with varying degrees of research activity. They voluntarily enrolled in the study after participating in faculty development workshops offered at the three institutions.

The cognition corner of the study is supported by a few theoretical frameworks that describe the multiple dimensions across which instructors make sense of their teaching duty [27, 28]. Drawing from these frameworks, the dimensions of teaching in this study included the role of the teacher, the role of the students, the purpose of assessment, the expected outcome (of teaching), and the teaching–research nexus as viewed and expressed by each participant. Transcripts of the semi-structured interviews mentioned before and summary narratives crafted by a researcher and validated by participants constitute the cognition corner. Inductive thematic analysis of the transcripts [7] yielded multiple codes within each dimension and provided initial support for the interpretation corner. However, thematic analysis across multiple dimensions proved insufficient to readily identify conceptions of teaching shared by participants as determined by similarities and differences across the high-dimensional data. To strengthen the interpretation,

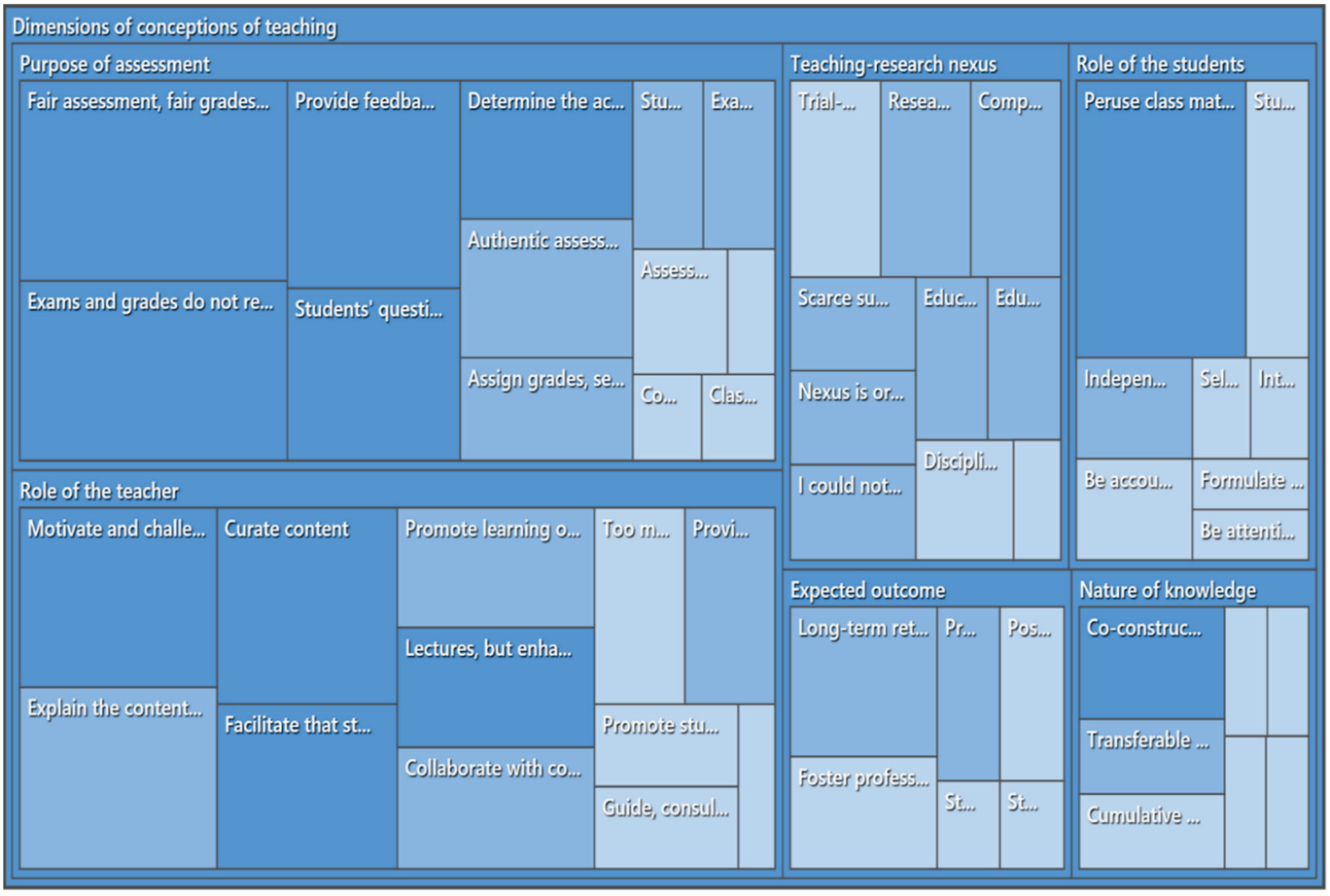


FIGURE 10 Dendrogram of clustering results based on faculty's conceptions of teaching.

the portion of the study described in this paper aimed at identifying groups of participants sharing similar views and potentially similar conceptions of teaching in a way more systematic and efficient than manually processing thematic analysis results. Twelve out of the 20 transcripts were selected due to their richness and variability. The software Nvivo® was used to cluster the transcripts on the basis of coding similarities using Jaccard's similarity coefficient [22].

When applied to coding, Jaccard's coefficient returns the number of instances when two participants A and B are assigned the same code, divided by the total number of instances when either or both are assigned any code. Calculated this way, the similarity coefficient ranges between 0 (no similarity) and 1 (complete similarity). Nvivo® calculates Jaccard's coefficient for every possible pair of participants in the sample. The resulting similarity coefficients for this study ranged from 0.15 to 0.44, with a median of 0.28. Nvivo® uses these coefficients as distances to apply the complete-linkage (farthest neighbor) clustering algorithm [9]. The output of the complete linkage algorithm can be represented as a dendrogram. In this example, as presented in Figure 10, clustering yielded four distinct clusters (A, B, C, and D) at the third level. To improve the consistency of the clustering, cluster D was split at the fourth level due to its heterogeneity. This yielded a total of five clusters: two comprising three participants (Clusters 1 and 2), two comprising four participants (Clusters 4 and 5), and one comprising six participants (Cluster 3).

To further characterize the clusters and identify the commonalities across the participants clustered together, we used treemaps [16] (called hierarchy charts in Nvivo®) and commonality tables. Hierarchy charts like the one in Figure 11 allow visual identification of the codes common to participants within a cluster. Darker areas in the chart represent a higher number of participants sharing a code and larger areas represent a higher number of passages coded, even within the same transcript. For instance, in the dimension of "Purpose of Assessment," there are multiple codes shared by all participants in this cluster, like their views on fair assessment, feedback as the main purpose of assessment, and the limitation of exams and grades in capturing student learning. In contrast, in the "Teaching–Research Nexus," there is less agreement in how participants in this cluster discussed this dimension, as indicated by the lighter colors of the codes within this category.

The information in the chart can be summarized in a commonality table. Commonality tables, like Table 1, list the codes assigned to the highest number of participants within each category or codes assigned

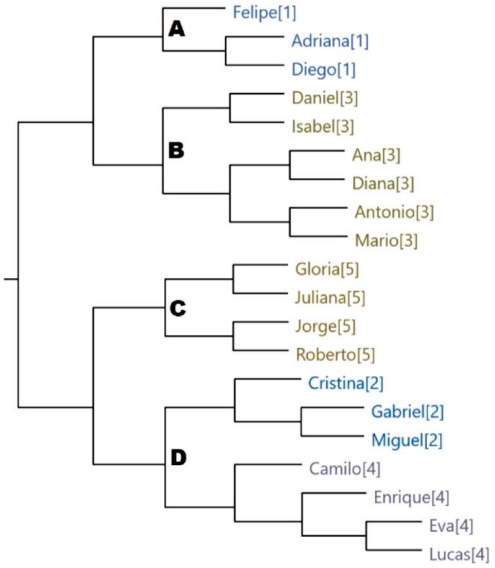


FIGURE 11 Hierarchy chart of participants in Cluster 1.

to a majority of participants including how many references. Codes in the table are sorted first in descending order of the number of participants coded with the same code and then in descending order of the total number of references. Numbers in parentheses in front of a code indicate the number of participants who referred to that code, if less than the number of participants in the cluster. This way, clusters can be formed, and the common characteristics highlighted for further analysis and interpretation.

2.5 | Discussion and implications

This paper proposed a rationale and guidelines to use computer-aided methods for data processing, analysis, and visualizations for complementing and supporting qualitative education research. Furthermore, it exemplified this approach in the context of three different studies in engineering education ranging from student learning to faculty development. The emergence of powerful computational technologies and methods has turned computation into an important tool for inquiry. In education, computational tools and methods have been mostly used for collecting students' interactions with tools and resources (i.e., learning analytics), for processing large data sets with artificial intelligence algorithms [3], and for intelligent tutoring systems that provide personalized experiences and feedback [15].

We argue, however, that these methods can be equally useful with smaller data sets, including transcripts from a small number of interviews or any other

TABLE 1 Commonalities of Cluster 1 (three participants).

Quantity	Conversion from Gaussian and CGS EMU to SI[a]
Role of the teacher	Curate content Motivate and engage students Facilitate students to learn at their own pace Lectures, but enhanced
Role of the student	Peruse class materials beforehand outside the class Independently browse multiple sources for content
Nature of knowledge	Co-constructed with the teacher
Purpose of assessment	Fair assessment, fair grades Provide feedback to the students Determine the achievement of learning objectives Exams do not reflect learning Students' questions and class attitude evidence learning
Expected outcome	Long-term retention of basic concepts (2) Prepare students for future duties and job (2)
Teaching-research nexus	Research as a tool for students (2) Competing time and resources (2) Scarce support for Scholarly teaching (2) Educational research must be applicable in context (2) Educational literature supports innovation (2) Nexus is or could be constructive (2) I could not assert that I do educational research as such (2)

qualitative data set, as the richness of qualitative data is often complex but difficult to report by using traditional descriptions. Learning analytics, for example, does not necessarily require large data sets because it essentially means collecting data on the interaction of the user/student with an educational tool. For example, Vieira et al. [39] collected students' interaction with an educational CAD tool to characterize students' engineering design process by analyzing their decisions and actions on building an energy-efficient house. These text files were relatively small (i.e., kilobytes) but provided rich information about students' processes. On the other hand, some machine learning methods, especially classifiers, typically require large data sets for data training and prediction. It is possible that given this general belief that computer-aided methods require large data sets, qualitative researchers may choose not to use computer-aided methods for education data sets. Nonetheless, they can be at least as good as human judgment for grouping small but highly dimensional qualitative data sets as we described in the studies about the types of knowledge that students used during a modeling and simulation process and the multiple dimensions that shape faculty's conceptions of teaching.

The main contribution of this study relates to the integration of guidelines from the assessment triangle to make explicit how theory, methods, and findings align with each other when using computer-aided methods in qualitative education research [18]. By means of the cognition corner, we made explicit the theoretical foundations of each of the three studies. Aligned with the purpose of the observation corner, we then operationalized constructs associated with each theoretical foundation to specific performances or behaviors, along with the ways in which data would be interpreted. Finally, by means of the interpretation corner, we interpreted the results derived from each corresponding study under the lens of the theoretical foundation stipulated in the cognition corner.

Qualitative researchers often rely on the human brain's innate ability to recognize patterns, but automatic clustering approaches may help them deal with high dimensionality in a consistent way, as we demonstrate in our case studies and discuss below. Key questions that researchers and educators should consider when using these methods to support their inquiry process include (1) How does it make sense to transform the qualitative categories into numbers? Are we interested in the presence/absence of a category (i.e., binary) or in the frequency of appearance (i.e., binary)? (2) What visualization strategy may be effective for the specific format of these numbers? [8] (3) Are the clusters meaningful? Clustering algorithms will often return groups for any data set [9, 12], even if the groups are meaningless. Thus, supporting qualitative data analysis using computer-aided methods does not liberate the researchers from evaluating the outcome, and even

deciding the number of clusters. While there are automatic methods to make an informed decision on the number of clusters, educational researchers should also go back to the data and existing theories to interpret the resulting groups. Likewise, a recent review of visual learning analytics [38] showed that it is uncommon to find studies that connect educational theories to make sense of an educational phenomenon with visualization theories to inform their selection of a visualization strategy. This results in inconclusive findings and limited theoretical contributions as (1) specific visualization strategies must be used for specific data sets and purposes [8] and (2) theoretically grounded research has the potential to enhance our understanding of educational phenomena.

In the first study, we analyzed students' written explanations of programming examples (observation) to characterize the approaches that they use to self-explain (cognition). We then used a tile plot and a hierarchical clustering approach with binary distance [32] to analyze these explanations, for example, interpretation; [6] and connect back to the theory of the self-explanation effect to make sense out of these results. In the second study, we characterized the types of knowledge [29] that students used during a modeling and simulation process (cognition). We collected the data using think-aloud protocols (observation) and used content analysis together with a bubble plot [5] and the k-means [12] clustering algorithm (interpretation) to understand how students approached this differently. In the third study, we explored engineering faculty conceptions of teaching (cognition) using a semi-structured interview protocol (observation). We analyzed this data set using Jaccard's similarity coefficient [22] based on coding similarities from thematic analysis, and visualized them using a treemap [16]. We connected back these groups using the theories of conceptions of teaching (interpretation [27, 28]).

Computer-aided methods, particularly pattern recognition, and qualitative analysis methods share similar processes for understanding messy and unstructured data (e.g., documents and videos). After familiarizing themselves with the raw data, researchers begin a qualitative data analysis by assigning categories or codes that could describe the phenomenon under study. Although it is rarely thought of this way, this process often results in a high-dimensional data set (i.e., a matrix), where each data point (e.g., interview transcript, document, video; represented as rows in the matrix) is assigned a large number of codes/categories (i.e., dimensions; columns in the matrix). A qualitative researcher would try to make sense of these highly dimensional data sets to group participants/documents based on similarities and would try to identify and describe encompassing themes. In comparison, a pattern recognition process starts by turning the unstructured data into numbers. The next step is to identify the key features (i.e., feature vectors) that best describe the data for a given purpose and to find groups/categories based on these feature vectors. One key difference is that the pattern recognition process always transforms the unstructured data into numbers, while qualitative analysis only does this when explicitly using content analysis. However, as we saw in the first example study, simply assigning categories to a data point can result in a number (i.e., a binary number).

3 | CONCLUSION

Computer-aided methods can be extremely useful to boost the analysis of qualitative data in educational research when rationally applied. In this paper, we provided three examples of studies where different computational approaches of visualization and clustering have been applied to support the analysis of qualitative data in engineering education research, while using the assessment triangle framework to guide that application.

Specifically, we proposed using the three corners of the assessment triangle [25] to guide a principle-based approach to integrate these methods in a meaningful manner [18]. Specifically, the cognition corner would inform the phenomenon that we are observing, whether learning, perceptions, or conceptualizations. The cognition corner also presents an opportunity to ground the study in a theoretical or conceptual framework. The observation corner describes how elicited students' behaviors and resulting artifacts can be used for analysis and visualization of the data are elicited, and the interpretation corner describes how we process, analyze, and interpret the data. In this last corner, it is imperative to go back to the data, the literature, and the theoretical framework (when available) to make sense of the outcomes. The alignment between the three corners of the assessment triangle also emphasizes the importance of rigor and alignment. For instance, the fact that visualization and clustering methods may provide insights and suggest approaches to grouping does not mean that these results are necessarily meaningful. It is the duty of the researcher to always go back to the data, to the literature, and to the theory to validate and make sense of these findings. In conjunction, these three studies show how to use a principle-based approach to extend qualitative research abilities by using computer-aided methods with small data sets.

ACKNOWLEDGMENTS

This research was supported in part by the U.S. National Science Foundation under awards No. DGE-2219271 and EEC-1826099. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

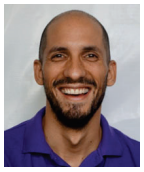
Camilo Vieira http://orcid.org/0000-0001-8720-0002
Juan D. Ortega-Alvarez http://orcid.org/0000-0001-6110-0791
Alejandra J. Magana http://orcid.org/0000-0001-6117-7502
Mireille Boutin http://orcid.org/0000-0002-0837-6577

REFERENCES

1. M. Arboleda, C. Vieira, and J. L. Chiu, *Opening the Machine Learning Black Box for Multidisciplinary Students: Scaffolding from GUI to Coding*, 2023 IEEE Front. Education Conference (FIE), 2023, pp. 1–5. https://doi.org/10.1109/FIE58773.2023.10343043
2. R. K. Atkinson, S. J. Derry, A. Renkl, and D. Wortham, *Learning from examples: instructional principles from the worked examples research*, Rev. Educ. Res. **70** (2000), 181–214.
3. A. Bakharia, *Towards cross-domain MOOC forum post classification*, Proc. Third (2016) ACM Conf. Learning@ Scale, 2016, pp. 253–256.
4. C. G. Berdanier, C. M. McComb, and W. Zhu, *Natural language processing for theoretical framework selection in engineering education research*, 2020 IEEE Front. Educ. Conf. (FIE), IEEE, 2020, pp. 1–7.
5. L. Bessler, *Bubble Plots*, Visual Data insights using SAS ODS graphics: a guide to communication-effective data visualization (L. Bessler, ed.), Apress, Berkeley, CA, 2023, pp. 263–281. https://doi.org/10.1007/978-1-4842-8609-8_7
6. J. L. Chiu, and M. T. Chi, *Supporting self-explanation in the classroom*, Applying science of learning in education: infusing psychological science into the curriculum (V. A. Benassi, C. E. Overson, and C. M. Hakala, eds.), Society for the Teaching of Psychology, 2014, pp. 91–103.
7. V. Clarke, and V. Braun, *Teaching thematic analysis: overcoming challenges and developing strategies for effective learning*, Psychologist **26** (2013), 120–123.
8. S. D. Evergreen, Effective data visualization: the right chart for the right data, SAGE Publications. https://www.google.com.co/books/edition/Effective_Data_Visualization/ilWKDwAAQBAJ?hl=en&gbpv=0, 2019.
9. B. S. Everitt, S. Landau, M. Leese, and D. Stahl, Cluster analysis: Wiley series in probability and statistics, John Wiley & Sons, Southern Gate, Chichester, West Sussex, United Kingdom, 2011.
10. M. Gillies, D. Murthy, H. Brenton, and R. Olaniyan, *Theme and topic: how qualitative research and topic modeling can be brought together*, ArXiv **2210** (2022), 00707.
11. R. Glaser, N. Chudowsky, J. W. Pellegrino, Knowing what students know: the science and design of educational assessment, National Academies Press, 2001.
12. A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, *K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data*, Inf. Sci. (Ny). **622** (2023), 178–210.
13. A. Klašnja-Milićević, M. Ivanović, and Z. Budimac, *Data science in education: Big data and learning analytics*, Comput. Appl. Eng. Educ. **25** (2017), 1066–1078. https://doi.org/10.1002/cae.21844
14. A. J. Kleinheksel, N. Rockich-Winston, H. Tawfik, and T. R. Wyatt, *Demystifying content analysis*, Am. J. Pharm. Educ. **84** (2020), 7113. https://doi.org/10.5688/ajpe7113
15. E. Kochmar, D. D. Vu, R. Belfer, V. Gupta, I. V. Serban, and J. Pineau, Automated personalized feedback improves learning gains in an intelligent tutoring system, Int. Conf. Artif. Intell. Educ., Springer, 2020, pp. 140–146.
16. N. Kong, J. Heer, and M. Agrawala, *Perceptual guidelines for creating rectangular treemaps*, IEEE Trans. Vis. Comput. Graph. **16** (2010) 990–998. https://doi.org/10.1109/TVCG.2010.186
17. E. G. Lyon, *Beliefs, practices, and reflection: exploring a science teacher's classroom assessment through the assessment triangle model*, J. Sci. Teach. Educ. **22** (2011), 417–435.
18. A. J. Magana and M. Boutin, *A principled approach to using machine learning in qualitative education research*, 2018 IEEE Front. Educ. Conf. (FIE), IEEE, 2018, pp. 1–7.
19. A. J. Magana, H. W. Fennell, C. Vieira, and M. L. Falk, *Characterizing the interplay of cognitive and metacognitive knowledge in computational modeling and simulation practices*, J. Eng. Educ. **108** (2019), 276–303.
20. S. F. Marion and J. W. Pellegrino, *A validity framework for evaluating the technical quality of alternate assessments*, Educ. Meas. Issues Pract. **25** (2006), 47–57.
21. National Academies of Sciences Engineering Medicine, Promising practices for strengthening the regional STEM workforce development ecosystem, National Academies Press, Washington, DC, 2016.
22. S. Niwattanakul, J. Singthongchai, E. Naenudorn, and S. Wanapu, *Using of Jaccard coefficient for keywords similarity*, Proc. Int. Multiconf. Eng. Comput. Sci., 2013, pp. 380–384.
23. NRC, Scientific research in education, Committee on Scientific Principles for Education Research, 2002.
24. J. D. Ortega-Alvarez, Conceptions of teaching among Colombian engineering faculty: an exploratory study, Ph.D. Thesis, Purdue University Graduate School, West Lafayette, IN, 2019.
25. J. W. Pellegrino, *Important considerations for assessment to function in the service of education*, Educ. Meas. Issues Pract. **39** (2020), 81–85.
26. A. G. Picciano, *The evolution of big data and learning analytics in American higher education*, Online Learn. **16** (2012), 9–20.

27. D. D. Pratt, *Conceptions of teaching*, Adult Educ. Q. **42** (1992), 203–220.
28. K. Samuelowicz and J. D. Bain, *Conceptions of teaching held by academic teachers*, High. Educ. **24** (1992), 93–111.
29. R. Shavelson, M. A. Ruiz-Primo, M. Li, and C. C. Ayala, Evaluating new approaches to assessing learning, National Center for Reserach on Evaluation, Standard and Student Testing California University, Los Angeles, 2003.
30. A. B. Shiflet and G. W. Shiflet, Introduction to computational science: modeling and simulation for the sciences, Princeton University Press, Princeton, NJ, 2014.
31. R. A. Streveler, R. L. Miller, A. I. Santiago-Román, M. A. Nelson, M. R. Geist, and B. M. Olds, *Rigorous methodology for concept inventory development: using the "assessment triangle" to develop and test the thermal and transport science concept inventory (TTCI)*, Int. J. Eng. Educ. **27** (2011), 968–984.
32. D. Tamasauskas, V. Sakalauskas, and D. Kriksciuniene, *Evaluation framework of hierarchical clustering methods for binary data*, 2012 12th Int. Conf. Hybrid Intell. Syst. (HIS), IEEE, 2012. 421–426.
33. R. Taub, M. Armoni, E. Bagno, and M. (Moti) Ben-Ari, *The effect of computer science on physics learning in a computational science environment*, Comput. Educ. **87** (2015), 10–23. https://doi.org/10.1016/j.compedu.2015.03.013
34. M. Van Den Haak, M. De Jong, and P. Jan Schellens, *Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue*, Behav. Inf. Technol. **22** (2003), 339–351.
35. C. Vieira, A. J. Magana, and M. Boutin, *Using pattern recognition techniques to analyze educational data*, 2017 IEEE Front. Educ. Conf. (FIE), 2017, pp. 1–3. https://doi.org/10.1109/FIE.2017.8190592
36. C. Vieira, A. J. Magana, and M. Boutin, *Add-on preferential groups (APG): analyzing student preferences of teaching methods*, Comput. Appl. Eng. Educ. **26** (2018), 1020–1032.
37. C. Vieira, A. J. Magana, and M. Boutin, *Using computational methods to analyze educational data*, 2019 IEEE Front. Educ. Conf. (FIE), 2019, pp. 1–4. https://doi.org/10.1109/FIE43999.2019.9028599
38. C. Vieira, P. Parsons, and V. Byrd, *Visual learning analytics of educational data: a systematic literature review and research agenda*, Comput. Educ. **122** (2018), 119–135.
39. C. Vieira, Y. Y. Seah, and A. J. Magana, *Students' experimentation strategies in design: is process data enough?* Comput. Appl. Eng. Educ. **26** (2018), 1903–1914.
40. C. Vieira, A. J. Magana, M. L. Falk, and R. E. Garcia, *Writing in-code comments to self-explain in computational science and engineering education*, ACM Trans. Comput. Educ. **17** (2017), 1–21.
41. C. Vieira, A. J. Magana, A. Roy, and M. L. Falk, *Student explanations in the context of computational science and engineering education*, Cogn. Instr. **37** (2019), 201–231.
42. C. Vieira, R. L. Gómez, M. Gómez, M. Canu, and M. Duque, *Implementing unplugged CS and use-modify-create to develop student computational thinking skills: a nationwide implementation in Colombia*, Educ. Technol. Soc. **26** (2023), 155–175. https://doi.org/10.30191/ETS.202307_26(3).0012
43. J. M. Wing and D. Stanzione, *Progress in computational thinking, and expanding the HPC community*, Commun. ACM **59** (2016), 10–11. https://doi.org/10.1145/2933410
44. X. Zhai, Y. Yin, J. W. Pellegrino, K. C. Haudek, and L. Shi, *Applying machine learning in science assessment: a systematic review*, Stud. Sci. Educ. **56** (2020), 111–151.

AUTHOR BIOGRAPHIES



Camilo Vieira, PhD, is an assistant professor in the Department of Education at Universidad del Norte (UniNorte; Barranquilla-Colombia). He holds a BSc degree in Systems Engineering and a Master's Degree in Engineering from Universidad Eafit (Colombia) and a PhD in Computational Sciences and Engineering from Purdue University (Indiana-US). Dr. Vieira completed a postdoctoral experience at Purdue University in information visualization, and he has been working at UniNorte since 2019, where he coordinates the research group Informática Educativa. In 2022, he was a Fulbright Visiting Scholar in the Department of Curriculum, Instruction and Special Education at the University of Virginia. He investigates how to support student complex learning and how instructors teach complex topics, particularly in computing education. He also explores how to use computational methods to understand educational phenomena.

Juan D. Ortega-Alvarez is a collegiate assistant professor in the Department of Engineering Education at Virginia Tech and a Visiting Professor of Engineering at Universidad EAFIT (Medellín, Colombia). He earned his Ph.D. in Engineering Education from Purdue University and holds an MS in Process Engineering and Energy Technology from Hochschule Bremerhaven (Germany). With over a decade of experience in academia, Juan has been actively involved in teaching undergraduate and graduate courses. Additionally, he brings more than six years of practical engineering experience, primarily focusing on the design and enhancement of chemical processing plants.

Dr. Alejandra J. Magana is the W.C. Furnas Professor in Enterprise Excellence in the Department of Computer and Information Technology and Professor in the School of Engineering Education at Purdue University. Dr. Magana holds a BE in Information Systems and an MS in Technology, both from Tec de Monterrey, and an MS in Educational Technology and a PhD in

Engineering Education, both from Purdue University. Her research program investigates how model-based cognition in Science, Technology, Engineering, and Mathematics (STEM) can be better supported by means of expert tools and disciplinary practices such as data science computation, modeling, and simulation. In 2015, Dr. Magana received the National Science Foundation's Faculty Early Career Development (CAREER) Award for investigating modeling and simulation practices in undergraduate engineering education. In 2016, she was conferred the status of Purdue Faculty Scholar for being on an accelerated path toward academic distinction, and in 2022, she was inducted into the Purdue University Teaching Academy, recognizing her excellence in teaching.

Mireille Boutin is a full professor in the Department of Mathematics and Computer Science at the Eindhoven University of Technology, Netherlands, where she leads the Discrete Algebra and Geometry (DAG) group. She is also on the faculty of the Department of Mathematics and the Elmore Family School of Electrical and Computer Engineering at Purdue University, United States. Her research is in the area of signal processing, machine learning, and applied mathematics. She is a three-time recipient of Purdue's Seed for Success Award. She is also a recipient of the Eta Kappa Nu Outstanding Faculty Award, the Eta Kappa Nu Outstanding Teaching Award, and the Wilfred "Duke" Hesselberth Award for Teaching Excellence. She previously served as Associate Editor for the journals IEEE Signal Processing Letters, IEEE Transactions on Image Processing, and Applicable Algebra in Engineering, Communication and Computing. She is currently an associate editor for La Matematica and the SIAM Journal on Applied Algebra and Geometry.

How to cite this article: C. Vieira, J. D. Ortega-Alvarez, A. J. Magana, and M. Boutin, *Beyond analytics: using computer-aided methods in educational research to extend qualitative data analysis*, Comput. Appl. Eng. Educ. (2024); e22749. https://doi.org/10.1002/cae.22749