



# A macro perspective of the perceptions of the education system via topic modelling analysis

Jenny Cifuentes<sup>1</sup> · Fredy Olarte<sup>2</sup>

Received: 6 September 2021 / Revised: 5 January 2022 / Accepted: 11 May 2022 /  
Published online: 10 June 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

Education quality has become an important issue and has received considerable attention around the world, especially due to its relevant repercussions on the socio-economical development of society. In recent years, many nations have realized the need for a highly skilled workforce to thrive in the emerging knowledge-based economy. They have consequently adopted strategies to identify the lines of action to improve the education quality. In response to the government's efforts to improve the education quality in Colombia, this study examines the current perceptions of the education system from the perspective of key local stakeholders. Therefore, we used a survey that contained open-ended questions to collect information about the limitations and difficulties of the education process for several groups of participants. The collected answers were categorized into a variety of topics using a Latent Dirichlet Allocation based model. Consequently, the students', teachers' and parents' answers were analyzed separately to obtain a general landscape of the perceptions of the education system. Evaluation metrics, such as topic coherence, were quantitatively analyzed to assess the modelling performance. In addition, a methodology for the hyper-parameters setting and the final topic labelling was presented. The results suggest that topic modelling strategies are a viable alternative to identify strategic lines of action and to obtain a macro-perspective of the perceptions of the education system.

**Keywords** Textual content analysis · Stakeholder perception · Text mining · Topic modelling · Educational system

---

Fredy Olarte contributed equally to this work.

✉ Jenny Cifuentes  
jacfuentesq@gmail.com

Fredy Olarte  
faolarted@unal.edu.co

<sup>1</sup> ICADE, Faculty of Economics and Business Administration; Department of Quantitative Methods, Universidad Pontificia Comillas, Calle Alberto Aguilera 23, Madrid, 28015, Spain

<sup>2</sup> Electrical and Electronics Engineering Department, Universidad Nacional de Colombia, Carrera 45 26-85, Bogotá, Colombia

## 1 Introduction

Surveys are a significant research tool that can help to gain insight into a study subject. Specifically, open-ended questions have been considered to be a critical element of surveys because they provide information to clarify ambiguities, to examine attitudes, and to detect spontaneous perceptions, which had not been considered during the survey planning [18]. Consequently, these questions allow the researcher to elicit a topic, even if there is a lack of knowledge in the survey that prevents the adequate formulation of closed questions. Common use cases of open-ended questions to study and analyze citizens' perceptions about social indicators include surveys on education [13, 19, 24], health care [1, 17, 20], and social service systems [7]. The results of these studies allow the identification of relevant topics that matter to stakeholders, the detection of obstacles to change performance, and they can help us to explain and understand the impact of social reforms and their possible lack of improvement.

Despite the great benefits of using open-ended questions to acquire and analyze information about stakeholders' perceptions and expectations, their processing is generally associated with a high work-load. The main reason for this is that the traditional approach associated with this task involves the work of analysts who read and manually categorize the whole dataset [18]. This process tends to be tedious and time-consuming. In addition, it can be susceptible to errors when different analysts individually process the data [22].

Several researchers have proposed strategies to explore and analyze text collections. At present, these techniques range from simple methodologies such as frequency counts [21] to more complex Machine Learning (ML) based algorithms [16, 25, 26]. In particular, Topic Modelling (TM) based strategies have emerged as an impressive paradigm to automatically process the semantic characteristics of large textual databases.

TM is oriented to group text instances, considering that each sample can be modeled as a function of latent variables called topics. In this context, a topic is defined by a set of words, which are selected by statistical methods [2]. This approach is generally considered to be an unsupervised algorithm because of the inference processed involved to represent the content of each modeled topic. Applications of this methodology include software engineering, linguistic sciences, social networking, and so on [8, 23].

Latent Dirichlet Allocation (LDA) is a text analysis method that is used to represent the topic structure present in a collection of text documents [2]. Using this approach, recent interesting results have included the identification of relevant topics for each coronavirus disease and the exploration of their corresponding research trends from academic papers and news [4], the modelling of key research topics in big data literature [14], or the identification of evolving trends and underlying topics in humanoid robots research by analyzing scientific articles and patents [9]. In education, the use of this approach has not been fully exploited. One of the studies developed in the field is focused on the analysis and visualization of cognitive information that can improve the collaborative learning in classrooms. To this end, the work in [6] implements a Vector Space Model to develop the methodology, which was consequently validated in an experimental case study. The results of this study provide significant elements in the discussion about the student learning process. Recently, the LDA-based approach has also been used to analyze the responses of a teacher self-assessment survey in an Ecuadorian university. As a result of this case study, a set of main strategies that teachers can carry out in their classes with the aim of improving student retention were identified and discussed [3]. An alternative analysis was developed in [15], where the massive open online courses (MOOC) reviews were analyzed using LDA. In the results, the

most important characteristics of courses for learners were identified and exposed as a way to improve the overall MOOC learning experience.

This paper presents a complete methodology of collection, pre-processing, topic modelling, and results analysis, based on LDA to represent the categories from several groups of stakeholders of a set of answers to open-ended questions about educational system limitations. As relevant keypoints in the analysis, this approach describes the data collection, the initial exploratory analysis based on a relevant word frequency metric, the topic modelling method, the hyper-parameters setting and the final labelling stages. The survey evaluated in this study is oriented to acquire information about the main expectations and difficulties of the current educational system in Bogota, Colombia. Considering the possible diversity in the ideas from different stakeholders (students, parents and teachers), each group is analyzed separately. Based on this process and the analysis provided by a team of experts in qualitative analysis, the results show the main similarities and differences between the considered groups.

This paper is organized as follows. Section 2 describes the methodology of pre-processing and analysis that is used to process the textual data from the case study under consideration. In addition, the algorithm used for the topic modelling and analysis, and its corresponding approach to set and interpret the hyper-parameters is presented in this section. In Section 3, the results of this case study are detailed. Finally, Section 4 outlines the conclusions of the developed work.

## 2 Methodology

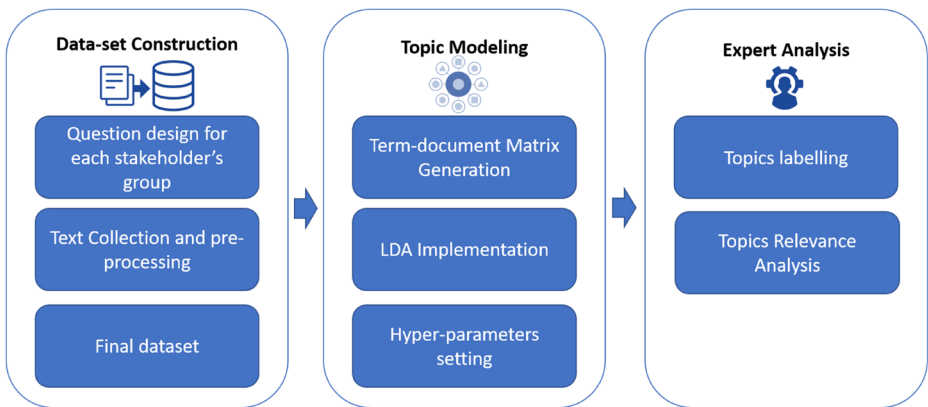
In this section, the methodology to model the topics in a set of unstructured textual data is presented. The case study analyzes the answers to open-ended questions that are designed to identify the current expectations and limitations in the educational system of Bogota from different points of view. Figure 1 details every stage of the proposed analysis methodology. During the first stage, the textual data is collected and a pre-processing process is carried out. Once the data is processed, the topic modelling analysis is performed through the implementation and tuning of an LDA-based model. In the final stage, a group of experts in this study area carry out the label identification of each topic. This task is developed using the information of keywords and bigrams from each of the stakeholder groups. It is important to highlight the relevance analysis of the topics identified, with reference to the problem under consideration. The case study studies the topics that are automatically generated to identify the main limitations that stakeholders find in the current educational system.

### 2.1 Constructing the dataset

The open-ended questions for each stakeholder are designed and the data collection is carried out in the first stage. In addition, the generated dataset is pre-processed to extract the main information to be used in the following stage.

#### 2.1.1 Question design

To identify the most significant pedagogical and technical aspects to be improved in the educational process from different point of views, each of the stakeholder groups has been asked a slightly different question formulation, as follows:



**Fig. 1** Topic modelling methodology

- (i) **Students:** According to your experience as a school/higher education student, describe what characteristics you expect will be changed in your education environment to face the challenges that arise in your life after finishing school/university?
- (ii) **Teacher:** What characteristics of the pedagogical processes of the classroom, the institution and the educational system would you change to promote integral development during secondary/high education?
- (iii) **Parents:** What elements in the educational process would you change to impact the student's lives in a significant way and allow them to face challenges on a personal, family and social level?

These questions have been designed in conjunction with a group of experts in education to focus the formulation on the points of interest for each stakeholder. In addition, a minimum number of words (250) was set to ensure that a collection of topics were addressed in each answer. See the complete set of open-ended questions as well as a sample of the multiple-choice questions included in this study, for each stakeholder, in the Appendix A.

### 2.1.2 Text collection and pre-processing

The data that is utilized in this research were obtained from the mission of educators and citizen wisdom, which is a Bogota Secretary of Education initiative that is intended to define educational public policies for the city upto 2038. The mission's main purpose is to listen to diverse citizen opinions about education. Therefore, several virtual and face-to-face spaces were carried out to collect perceptions and expectations of around one million people. Students, teachers, and parents contributed to create an educational landscape of the entire city.

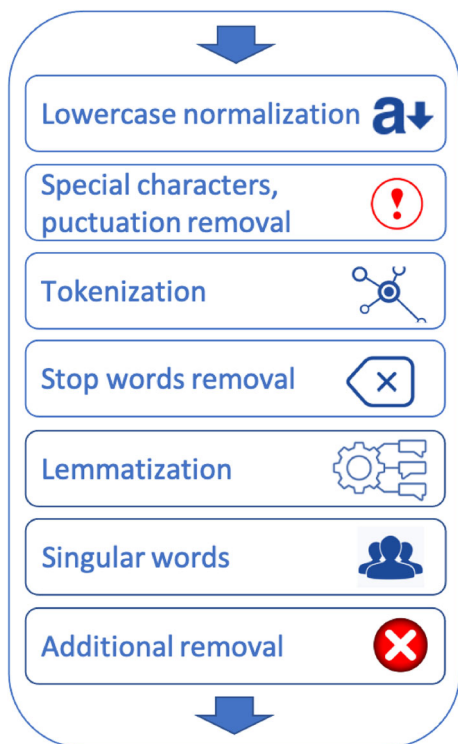
A set of open-ended questions were designed for each role, and were validated by subject matter experts and psychometrics. Responses were acquired using several mechanisms: a web platform was widely announced, paper-based forms were applied in streets and bus stations, and during new students inscriptions. In addition, a complete educational event, which was promoted by the Secretary of Education, allowed us to collect answers from more than 500000 parents.

To summarize, the data collection stage has allowed the analysis of 669456 answers from parents, 41390 answers from students and 7814 answers from teachers, the information was obtained from different sources. Then, the data has been digitized (if necessary) and subsequently a pre-processing stage was carried out to guarantee the data quality. This phase is particularly important for the analysis of unstructured textual information [10]. Figure 2 summarizes the sequential steps conducted in this process.

First, this stage involves a lowercase normalization, followed by the removal of special characters, punctuation and extra white spaces. The next step involved in the pre-processing was *tokenization*. The main objective of this task is to break down the text into smaller units, called tokens. The text can be divided by either words, characters or subwords (n-gram characters). As such, the data is tokenized by words, splitting the string elements into sub-strings. Based on this result, those common words in the language that might not add much value to the meaning of the document (stop words) are also removed. Subsequently, a lemmatization process was developed to group the different flexed forms of a word into a basic root word called lemma. In addition, the singular form of the words is obtained.

The final stage is to discard sparse terms that appear less than two times in the whole corpus, as well as those which appear in more than 95% of the documents, without losing relevant relationships inherent in the text instances. This task allows us to reduce the computation time involved in the next phases of the analysis. Likewise, duplicated answers are also removed. Consequently, the final dataset, which will be the input for the topic modelling analysis, is structured with the results of the previous described pre-processing. It is important to note that the textual information that is analyzed in this survey was acquired in Spanish. The pre-processing steps were adequately adapted to the particularities of this

**Fig. 2** Topic modelling methodology



language, considering that the implementation of natural language processing strategies (stopwords removal/lemmatization) in Spanish are still under development and some exceptions have not yet been included. Finally, the results presented in this work were translated to English by native speakers.

## 2.2 Topic modelling

After the data is processed, the topic modelling analysis is carried out, based on the following three main steps: the term-document matrix generation, where an initial exploratory analysis is performed; the implementation of the unsupervised algorithm (LDA); and the final setting of the related hyper-parameters.

### 2.2.1 Term-document matrix generation and exploratory analysis

During the processing and analysis of natural language, the textual instances are characterized by a bag of words, which is computationally represented by a term-document matrix. In this context, the word-document matrix can be considered as a simplified version of the textual corpus, and it is the input of the algorithms that are used to model the corpus topics [11]. It is important to note that the order of the textual instances does not suggest any implicit relation. In fact, during the computation of the word-document matrix, all of the textual elements are randomly mixed to carry out the required statistical processing and analysis. As such, strategies such as the Latent Semantic Analysis (PLSA) and LDA are based on the assumption of the exchangeability of words and textual instances [12].

Once the word-document matrix is generated, the words and words sequences that are the most frequently used, known as *n-grams*, are analyzed. Specifically, a uni-gram will be defined with one word and its frequency, while a bi-gram will be a set of two consecutive words and its frequency, and so on. The frequency of these sets of word or words helps to explore the most common concepts in the corpus. This analysis is carried out as a preliminary step to understand the recurrent ideas in the dataset, which will later will support the identification of the topics in the dataset. To consider the importance of each word in relation to other instances from the same corpus, the *Term Frequency - Inverse Document Frequency* (TF-IDF) is computed. To calculate the TF-IDF, it is required to compute the word frequency in a document (in this case, in an answer), and the word frequency in the other documents in the corpus. In other words, the following elements are calculated:

- Term Frequency (TF): Frequency of each token or word  $t$ , which appears in the document  $d$ ,  $\text{tf}(t, d) = f(t, d)$ .
- Inverse Document Frequency (IDF): the log of number of documents  $N$  divided by the number of documents that contain the token  $df_t$  (See (1)).

$$\text{idf}(t, N) = \log \frac{N}{df_t} \quad (1)$$

Lastly, the TF-IDF is calculated by multiplying the TF by the IDF:

$$\text{tfidf} = \text{tf}(t, d) \cdot \text{idf}(t, N) \quad (2)$$

This metric allows us to provide more relevance to those words that are repeated in more answers instead of words that are repeated a lot in just one answer.

### 2.2.2 LDA model

To obtain the topics of the set of answers analyzed, a topic modelling strategy using LDA is implemented. LDA is an unsupervised machine learning technique to assess data for patterns or latent topics. It is commonly used in studies that have small observations or unstructured text data, such as the answers of open-ended questions. LDA assigns every word a probabilistic score of the most probable topic it could belong to, where each topic is a mixture of words and each document is a mixture of topic probabilities.

In this context, the model considers the corpus ( $D = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_M\}$ ) as a collection of  $M$  documents with  $N_m$  words ( $\mathbf{w} = (w_1, w_2, \dots, w_N)$ ), with a set of  $W$  unique words. Then, each document is represented as a combination of  $k$  bag-of-words TOPICS, and each topic is modeled by means of a discrete probability distribution that establishes the probability that each word is present in a specific topic. Figure 3 shows the generation process of the LDA. In this model,  $\alpha$  and  $\eta$  are the hyper-parameters for Dirichlet distributions,  $\theta$  is the distribution of topics for each instance  $i$ , and  $\beta$  is the distribution of words for each topic  $k$ . In addition,  $z$  describes that a word is sampled in a particular topic, and  $w$  represents a simple word.

In this context, the probability distribution over words within a given answer is:

$$P(w_i) = \sum_{j=1}^T P(w_i | z_i = j) P(z_i = j) \tag{3}$$

where  $P(z_i = j)$  is the probability that the  $j$ -th topic was sampled for the  $i$ -th word, and  $P(w_i | z_i = j)$  is the probability of word  $w_i$  of topic  $j$ .

### 2.2.3 Hyperparameter tuning

LDA considers  $\alpha$ ,  $\eta$ , and  $k$  as parameters and randomizes all other values (excluding  $w$ ). Based on this consideration, the goal is to determine which  $\alpha$  and  $\eta$  maximizes the probability of generating the actual corpus by determining the best instance/topic ( $\theta$ ) and topic/word ( $\beta$ ).

For the LDA implementation, a hyperparameter tuning is applied to set the the number of topics ( $k$ ), the parameter of document-topic density ( $\alpha$ ), the parameter for word-topic

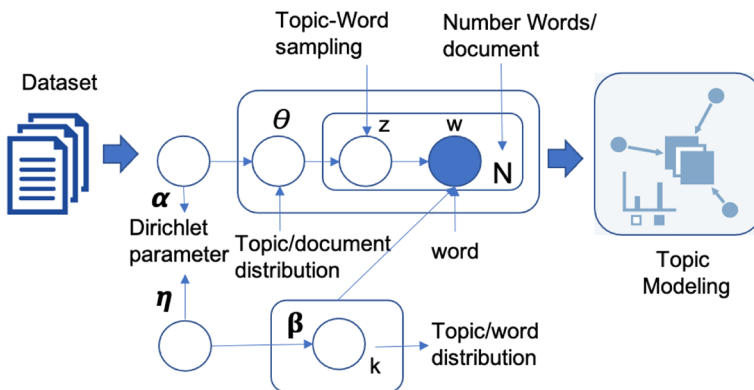


Fig. 3 LDA model representation

density ( $\eta$ ), and the number of iterations. To measure the model performance and compare, the *coherence score*  $c_v$  will be calculated. This probabilistic measure estimates if the words in the same topic go well together. This means that when the coherence score is high, the words are more closely related, while if it is very low, it contains words that do not occur in the same documents together or are not closely related.

Taking into account the corpus (bag of words associated to the complete answers) of each stakeholder’s group, a series of sensitivity tests is carried out to determine the best hyperparameters for the model. As previously stated, four parameters for the LDA modelling are considered:  $k$ ,  $\alpha$ ,  $\eta$ , and the number of iterations. Consequently, the hyperparameter tuning consists of three tests:

- (i) Finding the number of  $k$  topics.
- (ii) Finding the best Dirichlet hyperparameter  $\alpha$  and  $\eta$ . To calculate  $\alpha$ , the following approaches are considered:
  - Fixed normalized asymmetric prior of

$$\alpha = \left( \frac{1}{(1 + \sqrt{k})}, \frac{1}{(2 + \sqrt{k})} \dots \right) \rightarrow$$

$$\alpha_i = \frac{1}{(i + \sqrt{k})}, i = 1, 2, \dots, k \tag{4}$$

where  $i$  is the topic index and  $k$  is the number of topics.

- Fixed normalized asymmetric prior of 1/number of topics.
- It learns an asymmetric prior from the corpus.
- An array of uniformly distributed symmetric values for all  $k$  topics, where values from 0.01 to 1, with a step of 0.3, are considered [5].

For  $\eta$  calculation, three different approaches are involved:

- Scalar for a symmetric prior over topic/word probability,
- It learns an asymmetric prior from the corpus.
- An array of symmetric values for all  $w$  words, where values from 0.01 to 1, with a step of 0.3, are included.

By exploring these different alternatives,  $\alpha$  and  $\eta$  values with a higher coherence score are selected. In short, from the previous considerations,  $\alpha$  defines a Dirichlet distribution hyperparameter that creates the  $k$ -dimensional document-topic ( $\theta$ ) vectors, while  $\eta$  produces the  $W$ -dimensional topic-word ( $\beta$ ) vector. In turn,  $\theta$  and  $\beta$  act as parameters for categorical distributions, where topics and words are sampled, respectively.

- (iii) Obtaining the optimal number of iterations of the model: Now that the  $k$  value is set, and the best value for  $\alpha$  and  $\eta$  is calculated, the best amount of iterations is finally selected. The number of iterations controls the repetitions of a particular loop over each document. It is important to set this value high, so we select a range from 50 to 150 iterations. The chosen value provides the best coherence score.

With these steps, the best parameters ( $k$ ,  $\alpha$ ,  $\eta$ , and number of iterations) are selected for the modelling to obtain the highest  $c_v$ . This in turn generates more meaningful and interpretable topics. Hence, the final step for the topic modelling is to analyze the topics



that the model generated, draw conclusions about the theme of each topic and analyze them in terms of its distribution in the dataset.

In addition to this analysis, the *intertopic distance* is computed to analyze the closeness among the modeled topics. To visualize, first the *Jensen-Shannon divergence (JDS)* between topics is calculated. Specifically, this metric is a symmetrized and smoothed version of the Kullback-Leibler divergence, which is used to calculate similarities between two distributions. Therefore, the Jensen-Shannon divergence of  $P$  and  $Q$  is defined as:

$$\text{JSD}(P \parallel Q) = \frac{1}{2}D(P \parallel M) + \frac{1}{2}D(Q \parallel M) \quad (5)$$

Let  $M = \frac{1}{2}(P + Q)$ . The Jensen-Shannon distance is obtained by taking the square root of this divergence. Consequently, taking into account this definition, the probability distributions for each of the topics ( $\beta$ ) extracted by the LDA algorithm are analyzed and the distance between each topic is computed. Then, considering these results, a multidimensional scaling is used to project the *intertopic distances* onto a 2D plane. In this representation, the area of the circle or blob represents the importance of each topic over the entire corpus and the distances between these blobs indicate the closeness or similarity between each topic. The respective centers are defined by the calculated distance between topics, while the circle's area defines the prevalence of each topic. Hence, during the analysis, the preferred model will be the one that has the least or preferably no overlapping circles, and is spread throughout the graph.

### 2.3 Expert analysis

At this stage, based on the results of the topic modelling algorithm, the analysis of the labels that identify each of the obtained categories was carried out for each stakeholder group. Therefore, an expert team in qualitative analysis has evaluated the results of the keywords and the bigrams per each topic returned in the proposed methodology.

Before analyzing the information of each model, a manual corpus-labeling process was performed. In this task, 5% of answers from parents, and 10% of answers from students and teachers, were randomly analyzed. This approach is focused on a general reading of the chosen answers and the identification of macro-descriptors to which each stakeholder refers in the corresponding answers. This manual labeling provided relevant information to establish criteria for the final process of categories tagging.

Based on these results and the keywords/bigrams information of each model, we have titled each category in case of finding a pattern, which would allow the satisfactorily labeling. As a result of this stage, a logical association of the descriptive keywords to the related category is obtained. It is important to note that no descriptor was assigned to the words groups in case the topics seemed to be incomprehensible.

## 3 Experimental results and discussion

### 3.1 Preliminary analysis

After the dataset construction stage was finished and the term-document matrix was generated, a TF-IDF analysis was performed and the relevant terms in each corpus were identified. In the unigrams case (words), the terms that were present in more than 95%

of the answers were skipped. The most important words, bigrams and trigrams for each stakeholder's group are listed in Tables 1, 2, and 3.

Considering these results, the answers associated to each bigram and trigram were extracted and an initial qualitative analysis were performed. This stage allows us to identify the main/recurrent idea behind each bigram/trigram obtained from the preliminary analysis, for each stakeholder's group. Consequently, from the results, it is possible to see that the students answers involve ideas about having their school as a space with large green areas, with special care for the environment, where they could again have face-to-face classes and outdoor classes. Likewise, with respect to the classes, the need for dynamic and didactic classes, and ludic activities, where the teacher understands the students needs and focuses on the development of their skills, is identified. An additional relevant concept in their answers is focused on the skills to complete a resume. Finally, these participants see their educational process as an opportunity to enter a university, in which the knowledge addressed in the school could contribute for a better future and improve, in some way, their quality of life.

When analyzing the complete answers for the teacher stakeholder group, it is possible to observe a great concern about having an adequate number of students in the classroom, as well as the promotion of comprehensive development and meaningful learning during the educational process. Specifically, this actor gives particular importance to the pedagogical and learning processes in the classroom, including reading-writing processes and ludic activities, and the incorporation of technological tools and didactic material for the development of competencies. Finally, the importance of social-emotional skills and the participation of parents in the teaching process of their children is also mentioned.

The parents answers reveal that they are focused on awakening the students interest, highlighting the importance of learning in a didactic and amusing way, allowing them to develop skills that prepare them for the future and impact on their daily life. The relevance of the development of a life project is also pointed out, caring for the environment and values such as respect.

### 3.2 Topic modelling results

After carrying out the exploratory topic modelling analysis, an exhaustive search for the hyperparameters:  $k$  (number of topics),  $\alpha$  (Dirichlet distribution document-topic),  $\eta$  (Dirichlet distribution topic-word) and the number of number of iterations was performed to

**Table 1** Words, bigrams and trigrams with higher TF-IDF in the student group

Words	Bigrams	Trigrams
School	Green zone	Green zone school
Class	Big school	Imagine big school
Learn	Dynamic class	Take care environment
Lifetime	Opportunity study	Teacher understand student
Technology	Continue study	Train study need
Future	Enter university	Outdoor class
University	Environment	Complete resume
Train	Virtual class	Opportunity enter university
Opportunity	Didactic class	Teach beneficial future
Degree	Life quality	Return facetoface class

**Table 2** Words, bigrams and trigrams with higher TF-IDF in the teacher group

Words	Bigrams	Trigrams
Process	Pedagogical process	Classroom number students
Develop	Integral develop	Promote integral develop
Learn	Education system	Teach learn process
Classroom	Number student	Classroom pedagogical process
Pedagogical	Develop skill	Read write process
Institution	Pedagogical process	Decrease number student
Family	Learn process	Specialize process area
Technology	Significant learn	Few number student
Evaluation	Technology tool	Socio emotional skill
Integral	Didactic material	Reinforce learning process

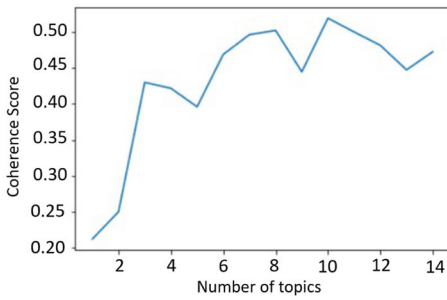
optimize the obtained results. As such, considering the methodology described in the previous section, the first step was the evaluation of models from 1 to 14 topics, and later compare and select the one that had a higher coherence value. The other parameters were set to their default in the LDA model, where parameters  $\alpha$  and  $\eta$  were both equal to a symmetric one over the number of topics ( $1/k$ ). The coherence values based on the number of topics are shown in Fig. 4.

Based on these results, the selected value will be the number of topics that marks the end of a rapid growth of the coherence values, where a suitable amount of topics is obtained and the topics can be interpreted without having many keywords being repeated in each category. It is possible to identify, from Fig. 4, that the  $k$  values with a higher  $C_v$  are obtained with 8 and 10 topics in the student group; 5, 7 and 11 in teacher group; and 9, 10 and 13 in parent group. Accordingly, the models with these number of topics were evaluated, calculating the intertopic distance and evaluating the value  $k$  that gives the most meaningful and interpretable topics.

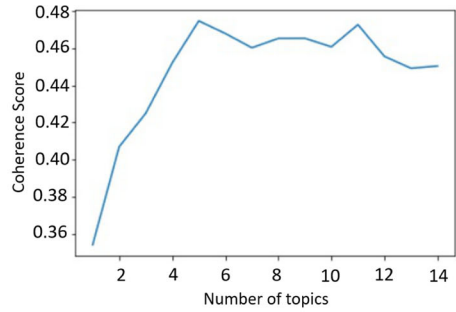
As discussed in the previous section, the best model has the least overlapping circles, and the topics are spread all over the graph representing the intertopic distance (Fig. 5). During the evaluation process, it was observed that when the amount of topics increased, there were more smaller circles (which may possibly be subtopics). In addition, more blobs that are

**Table 3** Words, bigrams and trigrams with higher TF-IDF in the parent group

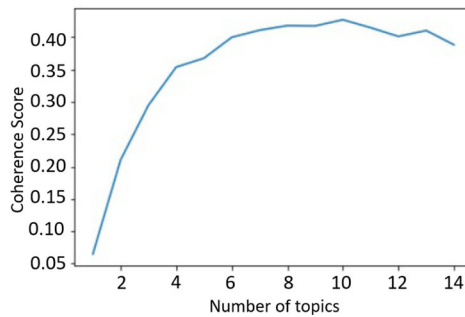
Words	Bigrams	Trigrams
Life	Life project	Didactic student interest
Child	Develop skill	Student interest tedium
Develop	Didactic student	Tedium enjoy learn
Value	Student interest	Personal family social
Learn	Enjoy learn	Develop life project
Future	Tedium enjoy	Impact life student
Skill	Daily life	Level personal life
Respect	Value respect	Take care environment
Family	Personal level	Challenge personal level
Personal	Family social	Value ethical moral



(a) Coherence score for models with 1 to 14 topics in students



(b) Coherence score for models with 1 to 14 topics in teachers

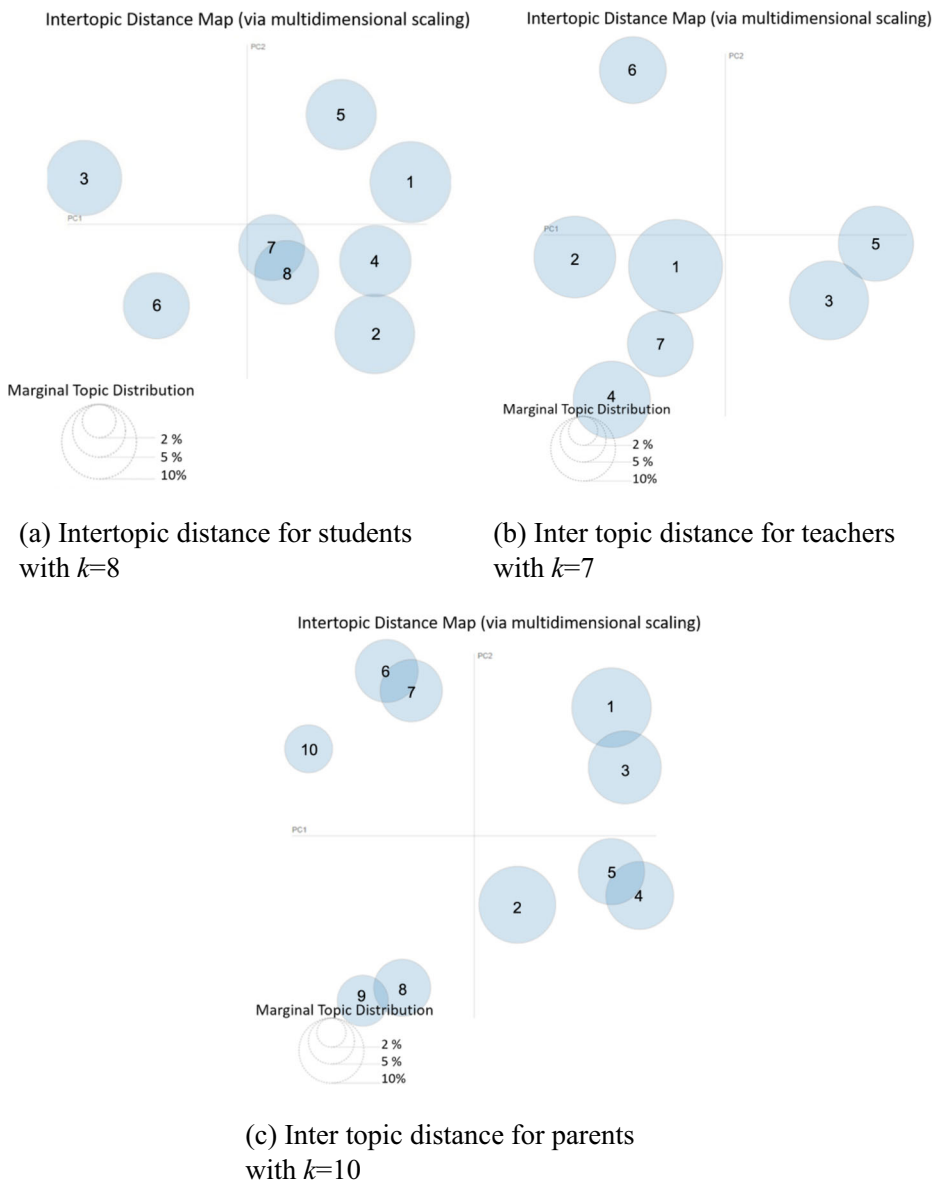


(c) Coherence score for models with 1 to 14 topics in parents

**Fig. 4** Coherence score for each stakeholders group, varying the number of topics  $k$

overlapping were present in the analysis. It is important to consider that when a greater number of topics are involved, they are less comprehensible. Therefore, largest circles and the least overlapping were obtained with  $k$  values equal to 8, 7 and 10, for the student, teacher and parent group, respectively (see Fig. 5).

With the optimal amount of topics, now  $\alpha$  and  $\eta$  parameters are tuned to obtain the highest coherence score. For the parameter  $\alpha$ , either symmetric or asymmetric values were considered; while for the  $\eta$  parameter, symmetric values were considered. For the symmetric values, different uniformly distributed values (e.g., 0.01, 0.31, 0.61, 0.91 and 1) were evaluated [5]. It is important to highlight that a low  $\alpha$  in a symmetric distribution means that it is more likely that each document may contain mixture of just a few topics. In contrast, a high  $\alpha$  means that the document is likely to contain a mixture of most of the topics and not a single topic. Likewise, for  $\eta$ , a high value represents that each topic is likely to contain a mixture of words and has a smoother distribution weight across all words. In addition to  $\alpha$  and  $\beta$ , the number of iterations was also tuned. In this case, the range of evaluation was from 50 to 150, by 10 steps of difference. The results with the highest coherence values, for each stakeholders group, can be seen in Table 4. As it can be seen in the results, while  $\alpha$  is small for the parents group—meaning that in proportion to the number of answers, the set of textual instances are modeled with a few amount of topics—, it is bigger for students and



**Fig. 5** Intertopic distance for the selected model of students, teachers and parents groups

teachers cases. Likewise, students and parents have a higher  $\eta$ , which means that the model of each topic is representative of a mixture of a considerable amount of words.

The number of iterations is similar in every group, with a smaller value for the parent group. This is an expected result, given the amount of answers analyzed for this stakeholder. With the generated model for each group, the top 10 more likely keywords, and the bigrams with larger frequency are found for each topic (See Tables 5, 6 and 7). One sign of a good topic model can be seen in the possibility of labeling the topic considering the top

**Table 4** Best parameters found for the model of the student, teacher and parent groups

Stakeholder	$\alpha$	$\eta$	Number iterations
Students	0.61	0.91	90
Teachers	0.31	0.01	90
Parents	0.01	0.91	70

words/bigrams of each group. As such, an initial category has been assigned to each cluster, based on a qualitative initial assessment, for each stakeholders group. This initial category seeks only to provide an initial label to the different groups, and they are not provided to the expert group nor informed to the LDA-model. Specifically, the labels were chosen by analyzing at the words/bigrams per topic with their probabilities and frequencies, respectively, and evaluating the answers that were most likely per topic. Although the models, presented for each stakeholder, seem to have consistent and interpretable topics, it is important to highlight that no one topic of each model was able to describe the analyzed dataset. LDA parameters are the important elements to characterize the models. This is due to the fact that LDA begins with a degree of randomness and, based on this particularity, it generates a slightly different topic model every time. However, in this case, the topics produced for each stakeholder, in each iteration, were similar.

The relevant topics that could be labeled for the students group include the need of language skills, the need of a preparation for a real world, the use of didactic strategies, the need to access to a higher education, the importance of the social relation at school, the improvement of the facilities of the educational institutions, the limitations during the virtual classes, and the importance of the use of technology. Meanwhile, teachers highlight the reduction of the number of students in the classroom, the importance of the family involvement, the integral development and the emotional intelligence in the students, and, similar to the students, an important number of answers are focused on the pedagogical strategies, the learning process and the use of the technology. Finally, the parents attach particular importance to the need of the theory and the practice in the learning process, the use of strategies to awaken the students interest, the development of talent and skills in their children, and the importance of the social interaction and the instruction in values. Similar to the students, they highlight the preparation for a real world, the access to a higher education, the use of the technology, and the limitations during the virtual classes. Finally, in accordance with the teachers, they give prominence to the family involvement during the learning process.

Based on these results, the final classification of answers in the different topics for each stakeholder group can be seen in Fig. 6. These distributions show that the preparation for a real world and the social relation at school are the most recurrent topics addressed by students. Meanwhile, teachers are more focused on the pedagogical strategies in the classroom and parents are more interested on awakening the students interest in the classes and the development of talent and skills.

### 3.3 Expert analysis results

To complete the analysis, an expert team in qualitative analysis has assessed the keywords and bigrams obtained from each topic. Based on a preliminary manual categorization, they have defined a more descriptive title for each topic. During the process, they have followed these steps:

**Table 5** Keywords, and bigrams for each topic modeled in the student group

Number	Topic	Keywords	Bigrams with higher frequency
1	Need language skills	Future, study, language, teach, degree, english, serve, need, course, practice	Opportunity study, serve future, after school, study future, teach language, study degree, buy house, learn language, language learn, english language
2	Preparation for a real world	Life, student, teach, prepare, work, focus, learn, help, confront, real	Daily life, after school, real life, confront life, work life, life teach, school teach, project life, life school, real world
3	Didactic strategies	Activity, child, learn, ludic, train, understand, student, teacher, didactic, school	Ludic activity, recreation activity, didactic activity, after school, teacher explain, student activity, school activity, train learn, understand student, outofclass activity
4	Access higher education	Student, high, knowledge, access, opportunity, quality, important, social, institution, university	Enter university, access high, after school, enter high, high free, vocational guidance, opportunity access, access university, improve quality, young child
5	Social relation at school	School, teacher, learn, change, classmate, happy, big, classroom, friend, share	Big school, imagine school, change school, share classmate, school learn, happy school, classmate teacher, big classroom, share friends, share student
6	Improvement of facilities	Improve, school, big, space, green, area, learn, play, help, zone	Green area, big school, school improve, imagine school, school space, improve facility, college area, big area, big space, green zone
7	Virtual classes	Class, dynamic, subject, virtual, didactic, life, fear, facetoiface, homework, internet	Dynamic class, didactic class, virtual class, facetoiface class, homework assign, fear class, school class, internet class, access internet, online class
8	Use of technology	Technology, use, education, implement, tool, teach, content, system, digital, screen	Technology tool, education system, use technology, implement technology, digital screen, advance technology, school technology, implement technology, digital tool, technology equity

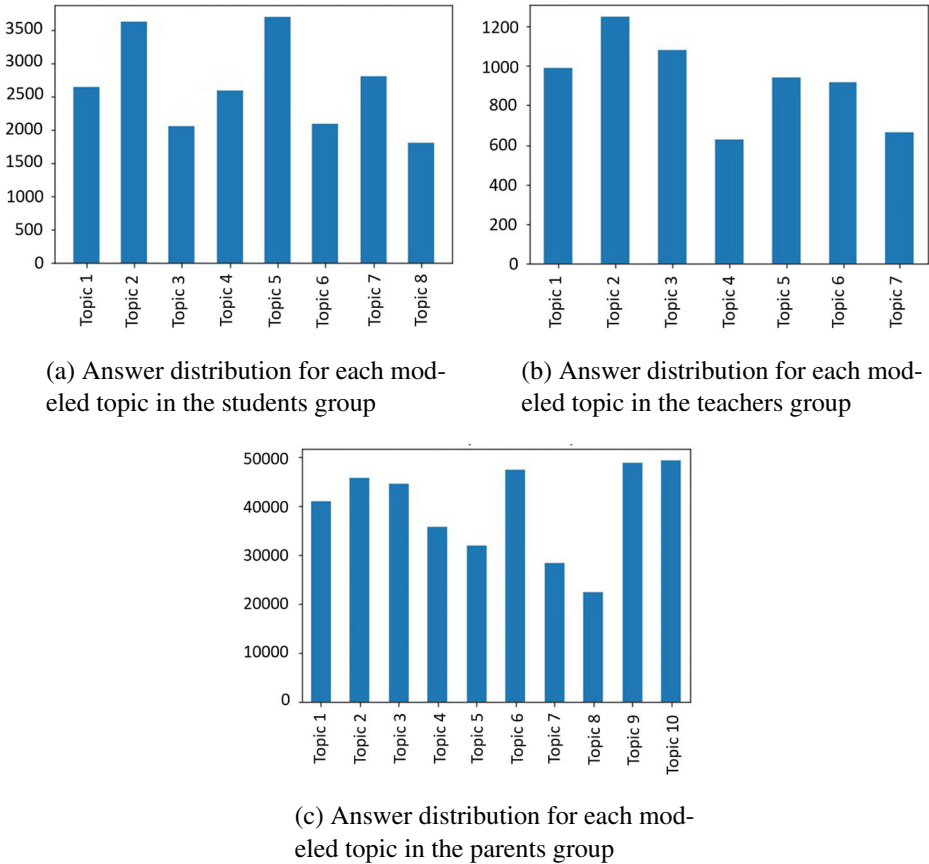
**Table 6** Keywords, and bigrams for each topic modeled in the teacher group

Number	Topic	Keywords	Bigrams with higher frequency
1	Learning process	Learn, child, improve, develop, significant, process, knowledge, train, teach, think	Significant learn, learn process, comprehensive teach, critical think, learn environment, teach process, improve process, ludic activity, develop skill, learn experience
2	Pedagogical strategies at classroom	Process, student, teacher, pedagogical, develop, child, train, initial, education, classroom	Pedagogical process, education system, develop pedagogical, pedagogical classroom, develop process, education process, develop process, teacher train, pedagogical skill, pedagogical strategies
3	Reduction of the number of students	Student, classroom, teacher, child, initial, quantity, material, number, process, learn	Student classroom, student quantity, student number, decrease quantity, teacher support, reduce number, maximum student, classroom class, child classroom, teacher specialize
4	Use of technology	Technology, student, process, material, knowledge, reality, teacher, classroom, tool, system	Technology tool, study plan, educational system, develop ability, technology resource, implementation tool, student technology, develop knowledge, technology material, integrate technology
5	Family involvement	Process, student, family, parent, class, improve, education, train, evaluation, relation	Parent family, family process, quantitative evaluation, learn process, involve parent, train process, improve process, education community, evaluation process, parent relation
6	Integral development	Develop, ability, child, play, process, read, institution, integral, write, skill	Develop skill, develop comprehensive, read write, comprehensive child, develop competence, read process, learn play, integral develop, integral skill, play art
7	emotional intelligence	Teacher, evaluation, student, develop, social, regulate, manage, emotion, child, intelligence	Emotion intelligence, hourly intensity, manage emotion, multiple intelligence, train evaluation, xxi century, emotion partner, social emotion, regulate emotion, social relation



**Table 7** Keywords, and bigrams for each topic modeled in the parent group

Number	Topic	Keywords	Bigrams with higher frequency
1	Preparation for a real world	Student, teach, life, project, problem, real, make, decision, learn, future	Life project, make decision, student future, real life, serve life, decision life, teach life, prepare real, real problem, problem solve
2	Access high education	Opportunity, school, study, university, after, life, future, career, child, degree	After school, opportunity study, work life, get degree, professional career, improve opportunity, access university, want study, give opportunity
3	Instruction in values	Value, respect, teach, life, child, responsibility, base, principle, love, student	Value respect, value principle, teach value, principle value, base value, love respect, respect responsibility, value respect, respect tolerance, right duty
4	Use of technology	Student, learn, child, improve, educational, tool, teach, develop, technology, need	Need tool, technology tool, student learn, provide tool, child learn, education system, technology interest
5	Social interaction	Personal, social, level, student, family, relation, life, develop, individual, child	Personal level, social family, social interaction, social personal, personal development, achieve goal, personal life, social individual, social relation, social skill
6	Family involvement	Student, parent, teacher, support, school, child, family, scholar, relative, relation	Family social, parent teacher, student teacher, parent relation, parent scholar, support parent, support family, relation father, relative relation, family school
7	Virtual classes	Classroom, virtual, child, facetoface, learn, teach, online, train, technology, access	Virtual class, facetoface class, environment, virtual facetoface, child learn, internet access, return class, technology internet, online courses, access technology
8	Theory and practice at the classroom	Student, practice, dynamic, learn, class, theory, relation, child, application, teach	Free expression, dynamic class, theory practice, relation practice, want study, dynamic learn, application theory, teach practice, class practice, active learn
9	Awakening students interest	Life, student, learn, didactic, confront, tedium, interest, enjoy, amuse, emphasis	Enjoy learn, student interest, place emphasis, educate entertain, amuse learn, serious didactic, tedium enjoy, daily life, didactic class, awaken interest
10	Development of talents and skills	Develop, student, skill, child, activity, talent, life, ability, ludic, improve	Develop skill, allow develop, skill student, improve skill, develop ability, practice talent, develop talent, ludic activity, improve talent, talent creativity



**Fig. 6** Answer distribution for the selected model of students, teachers and parents groups

1. The kinds of answers for each of the questions were determined and ordered according to each of the stakeholders.
2. A total sample of 5/10% of the answers was selected for the manual categorization.
3. A list of answers was drawn up for the questions of the different stakeholders and the first categories were drawn.
4. A logical grouping of descriptive categories was made and descriptors were established (Appendix B).
5. Based on the previous results, a categorization and coding manual was built for the responses of different stakeholders.
6. The answers were assigned to each category and the frequency of the categories was calculated.
7. Triangulation-analysis of qualitative results was performed with the results of the automatic classification (LDA).
8. The categories were adjusted for each analyzed group

**Table 8** Final labels for each topic in the student group

Number	Initial label	Final label
1	Need language skills	Need for a foreign language proficiency
2	Preparation for a real world	Development of life skills and competencies
3	Didactic strategies	Change of pedagogical methodologies for a significant learning
4	Access to higher education	Wider coverage and access to higher education
5	Social relation at school	Strategies to improve school coexistence
6	Improvement of facilities	Investment and modification in the infrastructure and equipping of the school
7	Virtual classes	Online classes and access limitations
8	Use of technology	Access and use of new technologies in education

Specifically, during the triangulation-analysis step, the categories established by the experts are matched with the groups obtained with the LDA model. This matching process has been developed by following the next steps:

1. Reading of all topics, bigrams and trigrams by stakeholder.
2. Perform the qualitative analysis between categories and descriptors, and bigrams and trigrams. The methodology of this process has the following characteristics:
  - (a) Each category, obtained from the manual analysis, was scored according to each group found by the LDA model, assigning a qualitative coherence value.
  - (b) The score defined from 0 to 1 took into account bigrams and trigrams, for each group. The score was made by dividing the number of bigrams or trigrams, that are consistent with the suggested category, over the total of bigrams and trigrams for each topic. The bigrams and trigrams selected by category were those present in more than 10 % of the observations.
  - (c) The category with the highest score and higher than 0.7 was the final category assigned to each LDA group.

**Table 9** Final labels for each topic in the teacher group

Number	Initial Label	Final Label
1	Learning process	More flexible curricular changes appropriate to the different learning processes of each student
2	Pedagogical strategies at classroom	Change of pedagogical methodologies for a significant learning
3	Reduction of the number of students	Changes in group sizes to ensure quality education
4	Use of technology	Access and use of new technologies in education
5	Family involvement	Greater interaction of the family in the educational process
6	Integral development	Development of skills and competencies for the comprehensive and integral training of students
7	Emotional intelligence	Education that integrates multiple intelligences

**Table 10** Final labels for each topic in the parent group

Number	Initial label	Final label
1	Preparation for a real world	Development of life skills and competences
2	Access high education	Wider coverage and access to higher education
3	Instruction in values	Instruction in values
4	Use of technology	Access and use of new technologies in education
5	Social interaction	Teaching interpersonal skills
6	Family involvement	Greater interaction of the family in the educational process
7	Virtual classes	Online classes and access limitations
8	Theory and practice at the classroom	Education focused on the practice of theoretical knowledge
9	Awaking students interest	Changes in the traditional education
10	Development of talents and skills	Vocational training focused on talents-hobbies oriented to their life projects

3. In each case, the consistency in the proportions of the manual categorization of the proposed category and the LDA group is finally validated. In all cases, results were congruous.”

Based on this analysis, the final descriptive labels selected for each topic can be seen in Tables 8, 9 and 10. From the results, it can be seen that new labels involve more details about the focus of the answers, which was the main objective of the preliminary manual labeling of the selected sample of data.

## 4 Conclusions

To assess the degree of public satisfaction in public politics (addressed in such important sectors of mutual interest as education), surveys are commonly used to understand the point of view of stakeholders (e.g., students, teachers, parents, etc). These surveys allow us to collect valuable information about possible lines of improvement during the education process. Usually, these tools include open-ended questions, focused on identifying spontaneous thoughts and discovering new lines of action. Although open-ended questions allow the acquisition of new information, they also require a large workload and manual processing time. This has been considered to be a significant disadvantage, discouraging the use of this kind of questions and avoiding the possibility of collecting information of great importance.

This study presents a complete methodology for the collection, pre-processing and automatic analysis of open-ended questions, using an unsupervised approach based on the identification of latent topics. Additional insights are provided to the topic labels obtained from the automatic results by an initial exploratory analysis using the tfidf metric and a fine labeling provided by an expert team in qualitative analysis. This approach allows us to model the topics discussed in the collected answers and obtain a macro-perspective of the education system perception from different points of view. This study will help to reduce the workload and the processing time that are required to complete the analysis of unstructured textual data from different sources, such as the answers acquired through open-ended questions.

During the analysis, three groups of stakeholders were interviewed: students, teachers and parents. Consequently, the questions were structured for each stakeholder to obtain information about the limitations identified, and the aspects to be changed in the educational system, to achieve goals consistent with the respective participant role. This application provides important information about the potential lines of action to improve the perception and satisfaction of the population in the education sector. As a result of this application, the categories generated by the models and expert feedback allowed us to clearly identify the relevant topics for each stakeholder. These results suggest that this methodology can be used to extract different kind of information in this field.

The results obtained from the methodology presented in this work show that some topics are addressed by only one group of participants. Only the students highlight the importance of a foreign language proficiency, the investment in the infrastructure and the strategies to improve school coexistence. In turn, teachers emphasize the pedagogical methodologies and curricular change, the reduction of the number of students per class and development of skills and competences focused on an integral development that integrates multiple intelligences. Finally, parents were interested in the instruction in values, the importance of teaching interpersonal skills and the changes of the traditional education to awaken the interest of the students. As a complement, both the students and parents underline the relevance of a wider coverage and access to higher education, the development of life skills and competences, and the online classes and access limitations. Teachers and parents highlight the importance of a greater interaction of the family in the educational process. As a common topic for all the groups, the access and the use of new technologies in education was reported to be an important element to consider in the change of the education system.

It is important to highlight that the proposed methodology has a practical applicability to identify prominent underlying, in a large collection of responses to open-ended questions, oriented to multiple stakeholders. The questionnaire design, acquisition, pre-processing, automatic categorization and expert feedback stages could be applied (without loss of generality) to study and analyze a macro perspective of multiple stakeholders' perceptions in any application. However, some considerations must be particularly analyzed such as the number of responses, which is required to be large to obtain a model with an acceptable performance and to take advantage of the time reduction during the categorization analysis, and the changes in the questionnaire between the different stakeholders, which will allow to extract different information for the same topic, based on multiple points of views. Remaining stages can be replicable for similar tasks such as analyzing open-ended feedback or discussion forums.

In our further work, the analysis will be focused on deepening the stakeholders' perception on the educational system, but obtaining a subdivision based on the grades and level of education. In this way, students, teachers and parents will be divided in sub-stakeholders and the questionnaire will be focused on delving deeper into the topics of interest, reported as a result of the present study, of each stakeholder. Considering that LDA-based models do not properly estimate correlations between topics, because of the nature of the Dirichlet distribution, an additional line of action is oriented to the automatic analysis of the relationships between topics through the modelling of spatial distributions. This approach will aim at avoiding the overlapping of concepts among different categories. Complementary studies could involve the acquisition of new variables such as the age, gender or residence location as well as the information from other areas (e.g., the corporate sector, administrative employees of educational institutions, etc.). This new data could help to expand the scope of our results.

## Appendix A: Full questionnaire

### A.1 Spanish version

#### A.1.1 Students

##### Multiple-choice questions

1. ¿Qué te gustaría que los estudiantes aprendieran en el colegio para que se desarrollen como seres humanos integrales?
  - (a) Acciones que le permitan sentirse seguros de sí.
  - (b) Conocer, manejar y expresar lo que sienten y piensan.
  - (c) Convivir en armonía y unidad con la familia, compañeros y comunidad.
  - (d) Cuidar del cuerpo y del ser interior como fuente de amor propio.
  - (e) Resolver problemas de la vida cotidiana.
  - (f) Seguir las responsabilidades, derechos y deberes ciudadanos.
  
2. ¿Qué estrategia te gustaría que se implementara en la educación superior para mejorar la enseñanza?
  - (a) Acercar la educación a los entornos locales mediante proyectos.
  - (b) Cooperar pedagógicamente con otros entes educativos a nivel nacional e internacional.
  - (c) Proponer espacios de diálogo con el mundo laboral.
  - (d) Reorientar las clases y trabajos de grado hacia las necesidades de la comunidad.
  - (e) Transformar la realidad desde la investigación
  - (f) Transformar los cursos en torno a las competencias profesionales que se requieren.
  - (g) Usar más las herramientas digitales complementando los recursos físicos y presenciales.
  
3. ¿Qué oportunidad te gustaría que la ciudad le ofreciera a los y las estudiantes al terminar la educación media?
  - (a) Acceder a cursos de formación para el trabajo y el desarrollo humano.
  - (b) Articulación entre la educación media e institutos de educación técnica, tecnológica u oficios.
  - (c) Contar con opciones de financiamiento por parte del Distrito de tal manera que pueda asegurar un cupo en una institución universitaria.
  - (d) Disponer de acompañamiento y apoyo en la inserción al mercado laboral.
  - (e) Posibilidades de prácticas en fundaciones artísticas, culturales o ambientales.
  - (f) Realizar intercambios que promuevan actividades ambientales o culturales.
  - (g) Realizar un intercambio a nivel nacional o internacional con grupos que promuevan la investigación.
  
4. ¿Qué necesidad de tu entorno te gustaría que se incluyera en los temas de estudio en el futuro?
  - (a) Acceso a tecnología para mejorar la producción.
  - (b) Atención en salud.
  - (c) Desarrollo de aplicaciones para celulares y tabletas.
  - (d) Desarrollo de competencias para la construcción de paz.

- (e) Formación de líderes en emprendimiento.
  - (f) Formación en valores éticos y democráticos.
  - (g) Fortalecimiento de las artes y medios audiovisuales.
  - (h) Mejora de vías terrestres y edificaciones.
5. ¿Cuál de los siguientes apoyos educativos sería más importante para los estudiantes en el futuro?
- (a) Acceder a educación pública y gratuita.
  - (b) Avanzar a un siguiente nivel de estudios con becas.
  - (c) Buscar apoyos económicos del Estado.
  - (d) Buscar financiar sus estudios.
  - (e) Estudiar algo que les permita trabajar.
  - (f) Poder estudiar fuera de la ciudad.
  - (g) Recibir apoyo de los padres para seguir estudiando.
  - (h) Terminar los estudios sin endeudarse.

### Open-ended questions

1. Describe de acuerdo a tu experiencia como estudiante de educación media/superior, ¿cuáles características esperas que se cambien en tu ambiente educativo para afrontar los desafíos que se presenten en tu vida después de terminar el colegio/universidad?
2. Si tuvieras que resumir en una frase el principal obstáculo que tienes o has tenido en el sector educativo para desarrollarte plenamente, ¿cuál sería?

### A.1.2. Teachers

#### Multiple-choice questions

1. ¿En qué le gustaría que se centrara la gestión de la institución educativa para fortalecer la formación integral de los estudiantes?
  - (a) Cubrir las necesidades de las y los estudiantes para poder estudiar.
  - (b) Formar a las familias en el acompañamiento familiar y educativo.
  - (c) Formar y capacitar al cuerpo docente.
  - (d) Fortalecer procesos de innovación educativa.
  - (e) Incorporar a la familia en los procesos de formación de la institución.
  - (f) Incorporar a otros sectores, como la industria, en los procesos de formación.
  - (g) Involucrar a la comunidad cercana de la institución en los procesos formativos.
  - (h) Mejorar la infraestructura y dotación de las instituciones.
  - (i) Fortalecer los procesos y necesidades para la formación de la infancia.
  - (j) Promover el bienestar de los integrantes de la comunidad educativa. deberes ciudadanos.
2. En el marco de la formación para docentes, orientadores y orientadoras, ¿Cuál estrategia le gustaría que se implementara para promover la transformación pedagógica?
  - (a) Consolidación de apoyos interinstitucionales sistemáticos.
  - (b) Coordinación, movilidad, flexibilidad y acuerdo entre los programas e instituciones que forman educadores.
  - (c) Encuentros entre actores educativos para organizar procesos de la formación del educador.

- (d) Espacios de formación entre pares: “Docentes que aprenden de docentes”.
  - (e) Experiencias formativas de encuentro con diversos contextos, poblaciones y propuestas educativas.
  - (f) Implementación de alianzas académicas para la innovación.
  - (g) Programas que formen desde la práctica pedagógica y las pasantías.
  - (h) Promoción de espacios alternativos de formación docente.
  - (i) Proyectos transversales que involucren el aprendizaje de otras disciplinas.
  - (j) Redes y estrategias tecnológicas para el trabajo y el intercambio.
  - (k) Aumentar la movilidad y flexibilidad entre los programas e instituciones que forman educadores.
  - (l) Diálogos que promuevan el desarrollo en diversas localidades y sectores.
3. ¿Cuál proceso le gustaría que se incluyera en la formación docente para preparar a los estudiantes en su tránsito de la educación media a la educación superior y vida laboral?
- (a) Acompañamiento personalizado a las y los estudiantes.
  - (b) Apoyo a las y los estudiantes con potencial que no estudian ni trabajan.
  - (c) Aprendizaje de temas asociados a la convivencia, reconciliación y paz.
  - (d) Capacitación en formas de enseñar que puedan atender intereses, necesidades y habilidades de las y los estudiantes.
  - (e) Construcción de acuerdos con las y los estudiantes sobre planes de estudios.
  - (f) Enseñanza-aprendizaje para resolver conflictos socioambientales.
  - (g) Generar diálogo entre estudiantes, directivos y familias sobre proyectos para las y los estudiantes
  - (h) Inclusión de la educación sexual en planes de trabajo de las y los docentes.
  - (i) Inclusión de la salud física y mental en los procesos de enseñanza.
  - (j) Integración de contenidos en educación técnica y tecnológica en los planes de estudio.
  - (k) Participación en espacios de diálogo con diversas instancias educativas.
4. ¿Cuál acción le gustaría que se realizara en el futuro para facilitar la permanencia y la terminación de los estudios de todos los estudiantes, en todos los niveles de formación?
- (a) Adaptar la enseñanza a poblaciones con diversas capacidades y estilos de aprendizajes.
  - (b) Adecuar el colegio a las necesidades de estudiantes en zonas apartadas o rurales.
  - (c) Brindar igualdad de oportunidades para toda la población.
  - (d) Contar con becas y estímulos para estudiantes que se destaquen por sus talentos.
  - (e) Facilitar la adquisición y acceso a computadores, internet, libros, revistas.
  - (f) Ofrecer programas de estudio flexibles tanto en jornadas como en contenidos.
  - (g) Ofrecer programas educativos virtuales de calidad para estudiantes.
  - (h) Propiciar un entorno escolar saludable y seguro.
5. En el futuro, ¿qué cambio facilitaría su actividad como docente?
- (a) Actividades que impacten de forma favorable en los barrios.
  - (b) Ajuste de las características de la institución educativa a las necesidades de estudiantes y docentes.
  - (c) Educación gratuita y equitativa en las aulas.
  - (d) Mayor inversión en formación y capacitación.
  - (e) Mayor pertinencia de la tecnología para educar.



- (f) Mejores condiciones socioeconómicas de mis estudiantes.
- (g) Menor cantidad de estudiantes por salón de clase.
- (h) Que los contenidos que se comparten sean más acordes a las y los estudiantes.
- (i) Salarios docentes acordes con los esfuerzos de enseñar.
- (j) Trabajar cerca del lugar de residencia.

### Open-ended questions

1. ¿Qué características de los procesos pedagógicos del aula, de la institución y del sistema educativo cambiarías para promover el desarrollo integral durante la educación secundaria/alta?
2. Si tuvieras que resumir en una frase el principal obstáculo que tienes o has tenido en el sector educativo para desarrollarte plenamente, ¿cuál sería?

### A.1.3 Parents

#### Multiple-choice questions

1. Para promover la formación integral de sus hijas e hijos, ¿Qué acción espera que haga la administración escolar?
  - (a) Cubrir las necesidades de las y los estudiantes para que puedan ir al colegio.
  - (b) Formar y capacitar el cuerpo docente.
  - (c) Fortalecer los procesos de innovación educativa.
  - (d) Incorporar a la familia en los procesos de formación del colegio.
  - (e) Involucrar a la comunidad cercana al colegio en los procesos educativos.
  - (f) Involucrar otros sectores como la industria o instituciones de educación superior en la formación.
  - (g) Mejorar la infraestructura y dotación de las instituciones.
2. ¿Cuál es el principal iniciativa pedagógico que motivarían a sus hijas e hijos a aprender?
  - (a) Entrenamiento en la comprensión de las emociones
  - (b) Experimentos y proyectos comunicativos en las aulas de clase.
  - (c) Posibilidad de elegir problemas para desarrollarlos libremente según sus intereses de aprendizaje.
  - (d) Proyectos ecológicos que fomenten la conservación del medio ambiente.
  - (e) Talleres de teatro que privilegien el aprendizaje natural.
  - (f) Uso de nuevas tecnologías como material didáctico.
3. ¿Qué le gustaría que aprendieran las y los estudiantes en su formación como bachilleres y que les ayude para su vida después de graduarse?
  - (a) Derechos y deberes como ciudadanos.
  - (b) Estrategias que permitan la disminución de la violencia intrafamiliar y sexual.
  - (c) Formas de comunicarse con el mundo y comprender dicha comunicación.
  - (d) Formas sanas de relacionarse consigo mismos(as) y con los demás.
  - (e) Habilidades que permitan el desarrollo empresarial.
  - (f) Herramientas necesarias en el desarrollo del trabajo y la educación.
  - (g) La propia historia y reconocer el mundo de otras maneras.
  - (h) Prevención del embarazo adolescente.

4. ¿Qué condición le gustaría que existiera en el futuro para facilitar la permanencia y la terminación de los estudios de las niñas, niños, jóvenes o adultos en todos los niveles de formación?
- Adaptar la enseñanza a poblaciones con diversas capacidades y estilos de aprendizajes.
  - Adecuar el colegio a las necesidades de las y los estudiantes en zonas apartadas o rurales.
  - Brindar igualdad de oportunidades para toda la población.
  - Contar con becas y estímulos para las y los estudiantes destacados.
  - Facilitar la adquisición y acceso a computadores, internet, libros y revistas.
  - Ofrecer programas de estudio flexibles tanto en jornadas como en contenidos.
  - Ofrecer programas educativos virtuales de calidad para las y los estudiantes.
  - Propiciar un entorno escolar saludable y seguro.
5. Preferiría en un futuro que la inversión en educación se centrará en:
- Ampliar la jornada escolar.
  - Aumentar el acceso a la educación superior.
  - Aumentar la atención de las niñas y niños en la primera infancia.
  - La infraestructura y el material educativo de las instituciones.
  - Los espacios de participación y recreación de las y los estudiantes.
  - Mejorar las condiciones salariales y laborales de las y los docentes.
  - Programas para reducir la reprobación de las y los estudiantes.
  - Aumentar el presupuesto desde los gobiernos y hacer vigilancia de los mismos.
  - Diversificar la jornada escolar.
  - Mejorar la formación de las y los profesores.

### Open-ended questions

- ¿Qué elementos del proceso educativo cambiaría para que tuvieran un impacto significativo en la vida de los alumnos y les permitiera afrontar los retos a nivel personal, familiar y social?
- Si tuviera que resumir en una frase el principal obstáculo que cree que tienen los estudiantes para desarrollarse plenamente en el sector educativo, ¿cuál sería?

## A.2 English Version

### A.2.1 Students

#### Multiple-choice questions

- What would you like students to learn in school in order that they develop as integral human beings?
  - Actions that allow them to feel self-confident.
  - To know, manage and express what they feel and think.
  - To live in harmony and unity with family, peers and community.
  - To take care of the body and the inner being as a source of self-love.
  - To solve daily life problems.

- (f) To follow the responsibilities, rights and duties of citizenship.
2. What strategy would you like to be implemented in higher education to improve teaching?
- (a) To bring education closer to local environments through projects.
  - (b) To cooperate pedagogically with other educational entities at national and international level.
  - (c) To propose spaces for dialogue with the labour market.
  - (d) To reorient classes and graduate work towards the needs of the community.
  - (e) To transform reality through research.
  - (f) To transform courses around the required professional skills.
  - (g) To use more digital tools complementing physical and face-to-face resources.
3. What opportunity would you like the city to offer students after high school?
- (a) Access to training courses for work and human development.
  - (b) Articulation between secondary education and technical, technological or vocational education institutes.
  - (c) To have access to financing options from the city in order to ensure a place in a university institution.
  - (d) To have accompaniment and support in the insertion to the labor market.
  - (e) Possibilities of internships in artistic, cultural or environmental foundations.
  - (f) Exchanges that promote environmental or cultural activities.
  - (g) To carry out an exchange at a national or international level with groups that promote research.
4. What needs in your environment/community would you like to be included in future study topics?
- (a) Access to technology to improve production.
  - (b) Health care.
  - (c) Development of applications for cell phones and tablets.
  - (d) Development of peace-building skills.
  - (e) Training of leaders in entrepreneurship.
  - (f) Training in ethical and democratic values.
  - (g) Strengthening of the arts and audiovisual media.
  - (h) Improvement of land roads and buildings.
5. Which of the following educational supports would be most important for students in the future?
- (a) Access to free public education.
  - (b) To advance to the next level of education with scholarships.
  - (c) To seek financial support from the State.
  - (d) To seek to finance their studies.
  - (e) To study a career with more job opportunities.
  - (f) To be able to study outside the city.
  - (g) To receive support from parents to continue studying.
  - (h) To finish their studies without getting into debt.

## Open-ended questions

1. According to your experience as a school/higher education student, describe what characteristics you expect will be changed in your education environment to face the challenges that arise in your life after finishing school/university?
2. If you had to summarize in one sentence the main obstacle you have had in the education sector to fully develop yourself, what would it be?

## A.2.2 Teachers

### Multiple-choice questions

1. What would you like the management of the educational institution to focus on in order to strengthen the comprehensive education of students?
  - (a) Meeting the needs of the students to be able to study.
  - (b) To instruct families in family and educational accompaniment.
  - (c) To educate and train the teaching staff.
  - (d) To strengthen educational innovation processes.
  - (e) To incorporate the family in the institution's educational processes.
  - (f) To incorporate other sectors, such as industry, in the training process.
  - (g) To involve the institution's community in the education process.
  - (h) To improve the infrastructure and equipment of the institutions.
  - (i) To strengthen the processes and needs for the children' education.
  - (j) To promote the well-being of members of the educational community.
2. Within the framework of training for teachers and guidance counselors, what strategy would you like to be implemented to promote pedagogical transformation?
  - (a) Consolidation of systematic inter-institutional support.
  - (b) Coordination, mobility, flexibility and agreement among programs and institutions that train educators.
  - (c) Meetings between educational actors to organize educator training processes.
  - (d) Peer-to-peer training spaces: "Teachers learning from teachers".
  - (e) Formative experiences of encounter with diverse contexts, populations and educational proposals.
  - (f) Implementation of academic alliances for innovation.
  - (g) Programs that train from pedagogical practice and internships.
  - (h) Promotion of alternative spaces for teacher training.
  - (i) Transversal projects that involve learning in other disciplines.
  - (j) Networks and technological strategies for work and exchange.
  - (k) Increasing mobility and flexibility among programs and institutions that train educators.
  - (l) Dialogues that promote development in diverse localities and sectors.
3. What process would you like to be included in teacher training to prepare students for their transition from secondary education to higher education and working life?
  - (a) Personalized accompaniment for students.
  - (b) Support for students who are neither studying nor working.
  - (c) Learning of topics associated with coexistence, reconciliation and peace.
  - (d) Teaching training to meet the students' interests, needs and abilities.

- (e) Construction of agreements with students on curricula.
  - (f) Teaching-learning to resolve socio-environmental conflicts.
  - (g) To generate dialogue between students, directors and families about projects for students.
  - (h) Inclusion of sexual education in teachers' work plans.
  - (i) Inclusion of physical and mental health in teaching processes.
  - (j) Integration of technical and technological education contents in the curricula.
  - (k) Participation in dialogue spaces with different educational instances.
4. What action would you like to be carried out in the future to facilitate the permanence and completion of studies for all students, at all levels of education?
- (a) To adapt teaching to populations with diverse learning styles and abilities.
  - (b) To adapt the school to the needs of students in remote or rural areas.
  - (c) To provide equal opportunities for the entire population.
  - (d) To have scholarships and incentives for students who stand out for their talents.
  - (e) To facilitate the acquisition and access to computers, internet, books, magazines.
  - (f) To offer flexible study programs both in terms of schedules and content.
  - (g) To offer quality virtual educational programs for students.
  - (h) To promote a healthy and safe academic environment.
5. In the future, what change would facilitate your activity as a teacher?
- (a) Activities that would favorably impact neighborhoods.
  - (b) Adjustment of the characteristics of the educational institution to the needs of the students and teachers.
  - (c) Free and equitable education in the classroom.
  - (d) Greater investment in education and training.
  - (e) Greater relevance of technology for education.
  - (f) Better socioeconomic conditions for the students.
  - (g) Less students per classroom.
  - (h) Contents shared are more in line with the students.
  - (i) Teachers' salaries according to their teaching efforts.
  - (j) Working close to the place of residence.

### Open-ended questions

1. What characteristics of the pedagogical processes of the class-room, the institution and the educational system would you change to promote integral development during secondary/high education?
2. If you had to summarize in one sentence the main obstacle you have had in the education sector to fully develop yourself, what would it be?

### A.2.3 Parents

#### Multiple-choice questions

1. In order to promote the integral education of your children, what action do you expect the school administration to take?
  - (a) To cover the needs of the students so that they can go to school.
  - (b) To educate and train the teaching staff.

- (c) To strengthen educational innovation processes.
  - (d) To incorporate the family in the school's educational processes.
  - (e) To involve the community close to the school in the educational process.
  - (f) To involve other sectors such as industry or higher education institutions in the educational process.
  - (g) To improve the infrastructure and equipment of the institutions.
2. What is the main pedagogical initiative that would motivate your children to learn?
- (a) Training in the understanding of emotions
  - (b) Experiments and communicative projects in classrooms.
  - (c) Possibility of choosing problems to develop them freely according to their learning interests.
  - (d) Ecological projects that encourage the conservation of the environment.
  - (e) Theater workshops that privilege natural learning.
  - (f) Use of new technologies as didactic material.
3. What would you like students to learn in their high school education that will help them in their lives after graduation?
- (a) Rights and duties as citizens.
  - (b) Strategies that allow the reduction of domestic and sexual violence.
  - (c) Strategies to communicate with the world and to understand such communication.
  - (d) Healthy ways of relating with themselves and with others.
  - (e) Skills that enable business development.
  - (f) Required tools in the development of the work and education.
  - (g) Their own history and to recognize the world in other ways.
  - (h) Prevention of teenage pregnancy.
4. What condition would you like to see in the future to facilitate the permanence and completion of studies for girls, boys, young people or adults at all levels of training?
- (a) To adapt teaching to populations with diverse abilities and learning styles.
  - (b) To adapt the school to the needs of students in remote or rural areas.
  - (c) To provide equal opportunities for the entire population.
  - (d) To have scholarships and incentives for outstanding students.
  - (e) To facilitate the acquisition and access to computers, internet, books and magazines.
  - (f) To offer flexible study programs both in terms of schedules and content.
  - (g) To offer quality virtual educational programs for students.
  - (h) To promote a healthy and safe academic environment.
5. In the future, I would prefer that investment in education would be focused on:
- (a) To extend the school day.
  - (b) To increase access to higher education.
  - (c) To increase the attention of girls and boys in early childhood.
  - (d) The infrastructure and educational material of the institutions.
  - (e) Participation and recreational spaces for students.
  - (f) To improve the salary and working conditions of teachers.
  - (g) Programs to reduce student failure.
  - (h) To increase government budgets and have them monitored.

- (i) To diversify the school day.
- (j) To improve teacher training.

### Open-ended questions

1. What elements in the educational process would you change to impact the student's lives in a significant way and allow them to face challenges on a personal, family and social level?
2. If you had to summarize in one sentence the main obstacle you think students have to fully develop themselves in the education sector, what would it be?

## Appendix B: Expert Analysis: Categories and Descriptors from the Manual Categorization

### B.1 Spanish Version

#### B.1.1 Students

**Table 11** Categories and descriptors used during the triangulation process and obtained from the manual categorization in the students group - Spanish version

Categoría	Descriptorios
Inversión y modificación en la infraestructura y dotación de colegios	Construcción de laboratorios. Construcción y/o mejora de bibliotecas. Mejora de ludotecas. Creación de zonas verdes y huertas. Creación, dotación y/o ampliación de zonas deportivas (patios, canchas, otros espacios para deportes). Creación y/o dotación de zonas de recreación y con fines culturales. Mejora de las condiciones de aseo de los baños. Implementación de tecnologías en los salones. Creación y/o dotación de comedores
Cambio de metodologías pedagógicas para un aprendizaje (una enseñanza) significativo	Aprendizaje - enseñanza significativa. Desarrollo de actividades pedagógicas. Clases prácticas. Educación contextualizada. Educación personalizada. Metodologías creativas para incrementar la motivación. Reducción de trabajos escolares. Desarrollo de actividades didácticas. Modificación metodologías pedagógicas
Mayor cobertura y acceso a la educación superior	Menor costo de la educación superior. Simulacros de exámenes de acceso de las universidades públicas. Difusión de la información de becas y oportunidades de acceso a la educación superior. Prioridad de acceso para estratos 1 y 2. Exámenes gratuitos de acceso a las universidades públicas. Convenios para acceder a la educación superior directamente desde el colegio
Clases remotas y limitaciones de acceso	No tener internet. No tener computador. Difícil conexión. No tener más clases virtuales. Tener clases semipresenciales
Acceso y uso de las tecnologías en la educación	Más uso de tecnologías. Mayor involucramiento de la tecnología en la educación

**Table 11** (continued)

Categoría	Descriptores
Desarrollo de habilidades y competencias para la vida	Aprendizajes que permitan resolver problemas de la vida real. Inclusión de cursos técnicos. Educación para el futuro. Educación que fomente el pensamiento crítico. Educación en habilidades para la vida laboral: presentación de entrevistas, construcción de la hoja de vida. Educación sexual. Educación emocional: salud mental (prevención de ansiedad, depresión). conocimiento para la vida cotidiana (impuestos, política y realidades locales, nacionales y globales). Impacto de las acciones diarias en el cambio climático (aprender a reciclar). Manejo de la vida financiera
Necesidad de dominio de un idioma extranjero	Formación en un segundo idioma. Aprender otros idiomas además de inglés. Colegios bilingües. Mayor exigencia en las materias de segunda lengua
Estrategias para mejorar la convivencia escolar	Espacios lúdicos que refuercen la buena convivencia. Mejora en las convivencia entre estudiantes. Mitigar acciones de violencia escolar y bullying. Mayor atención por parte de los docentes a peleas, bullying y conflictos entre los estudiantes. Reducir la indisciplina. Aumento de la seguridad y prevención de robos. Respetar el libre desarrollo de la personalidad. Disminuir la discriminación (raza, género, clase social, etc)

## B.1.2 Teachers

**Table 12** Categories and descriptors used during the triangulation process and obtained from the manual categorization in the Teachers group - Spanish version

Categoría	Descriptores
Mayor interacción de la familia en el proceso educativo	Escuela de padres (actividades) lúdicas, inducciones pedagógicas para fortalecer los procesos pedagógicos. Talleres y capacitaciones a los padres de familia. Mayor interacción entre el docente y los padres. Participación de la comunidad educativa. Mayor interacción y compromiso entre los padres y la comunidad educativa. Participación de la familia en la formación. Mayor compromiso e integración de los padres en la educación
Acceso y uso de nuevas tecnologías en la educación	Contar con recursos tecnológicos. Uso y manejo de tableros inteligentes. Tener salas de tecnología. Acceso a computadores, internet etc. Uso de herramientas interactivas para las clases. Capacitación de herramientas tecnológicas. Recursos tecnológicos para mejorar la eficiencia y la productividad en el aula
Educación que integra inteligencias múltiples	Inteligencia emocional: comunicación asertiva, autonomía, confianza. Manejo del dinero. Hábitos saludables
Desarrollo de habilidades y competencias para la formación completa e integral de los estudiantes	Pensamiento crítico (habilidades críticas, lectura crítica, actitud crítica). Formación en educación física (corporal y psicomotriz). Formación humanística. Desarrollo de habilidades artísticas y/o que fomenten la creatividad



**Table 12** (continued)

Categoría	Descriptorios
Cambios curriculares más flexibles y adecuados a los diferentes procesos de aprendizaje de cada alumno	Currículo participativo por todos los agentes, Currículo acorde a las necesidades del contexto del país. Innovación y flexibilidad curricular, transversales e integrales en el proceso de enseñanza y aprendizaje
Cambios en las metodologías pedagógicas para un aprendizaje significativo	Metodologías que incluyan la implementación de proyectos, utilizar estrategias y metodologías lúdico-didácticas que motiven al estudiante. Buscar aplicar los conocimientos del estudiante en su cotidianidad y en contextos específicos. Desarrollar autonomía en el estudiante. Metodologías más aplicadas basadas en proyectos y problemas
Cambios en el tamaño de los grupos para asegurar calidad educativa	Clases de más de 30 estudiantes. Grupos de aprendizaje más pequeños. Grupos muy grandes con aprendizaje menos efectivo. Control de las dimensiones de los grupos

### B.1.3 Parents

**Table 13** Categories and descriptors used during the triangulation process and obtained from the manual categorization in the Parents group - Spanish version

Categoría	Descriptorios
Mayor interacción de la familia en el proceso educativo	Generar responsabilidad desde los hogares. Configurar asociación de padres (capacitación y guías a los padres). Tener mucho acompañamiento de parte de los padres. Hacer más seguimiento y supervisión a los padres en el proceso educativo. Educar a los padres para que les enseñen valores a sus hijos. Generar compromiso de los padres para una mejor comprensión en lo educativo y lo familiar. Comunicación abierta padre-docente
Formación en valores	Valores como: espiritualidad, cuidado, respeto, honestidad, valoración del medio ambiente, ética, normas de comportamiento, valores en convivencia y responsabilidad
Desarrollo de habilidades y competencias para la vida	Cambios a una educación más significativa y acorde a las necesidades y desafíos de la vida real. Currículos ajustados a las necesidades actuales del país (contenidos teóricos-prácticos). Conocimiento en clases de finanzas personales para que desarrollen sus propias empresas (emprendimiento)- educación laboral (trámites administrativos necesarios en la vida adulta: sacar citas, sacar tarjetas crédito y débito, entre otros). Tener una educación enfocada a la vida cotidiana, conociendo mas su país (historia nacional y del mundo); Brindar herramientas para afrontar un cargo laboral. Tener preparación en planes económicos para hacer sus propias empresas
Acceso y uso de nuevas tecnologías en educación	Mayor uso de las tecnologías en la forma de transmitir conocimiento. Acceso equitativo a la tecnología. Más sesiones educativas en el uso de herramientas tecnológicas
Educación enfocada en la práctica del conocimiento teórico	metodologías que impliquen mayor práctica y menos contenido. Clases prácticas. Desarrollo de laboratorios. Aprender haciendo. Búsqueda y enseñanza de artes y oficios

**Table 13** (continued)

Categoría	Descriptores
Cambios en la educación tradicional	desarrollar autonomía en el estudiante. Metodologías personalizadas (se tiene en cuenta las preferencias del estudiante). Transmisión de conocimientos de forma atractiva. Educación más interactiva. Metodologías que estimulen a los estudiantes a aprender. Fomentar la lectura y desarrollo del pensamiento crítico
Enseñanza de habilidades interpersonales	resiliencia. Control de emociones. Comunicación asertiva. Resolución de conflictos. Tener clases de educación sexual (junto con estas se pueden tratar temas de planificación tanto para hombre como para mujer)
Formación vocacional enfocada en talentos-hobbies, orientada al proyecto de vida	enseñar cómo hacer un proyecto de vida. Tener un enfoque hacia temas que le sirvan a los niños para desarrollarse en un futuro. Elaborar proyectos que los lleven a nuevos desafíos. Contenidos y actividades enfocados en los intereses, aptitudes y talentos de cada estudiante. Proyectos en torno a diferentes ocupaciones. Elección de materias extracurriculares dependiendo de los intereses. Implementación en orientación vocacional. Desarrollo de los talentos y cualidades para potencializar las fortalezas de los estudiantes
Mayor cobertura y acceso a la educación superior	Acceso gratuito a la educación superior. Igualdad de oportunidades de acceder a educación superior. Tener educación superior pública y privada con igualdad de condiciones
Clases Virtuales y limitaciones de acceso	acceso a herramientas para clases virtuales. Limitaciones de internet como recurso educativo. Incidencia negativa de las clases virtuales en la socialización. Carencia de estructura pedagógica en clases virtuales. Necesidad de retomar clases presenciales

## B.2 English Version

### B.2.1 Students

**Table 14** Categories and descriptors used during the triangulation process and obtained from the manual categorization in the students group - English version

Category	Descriptors
Investment and modification in the infrastructure and equipping of the school	Construction of laboratories. Construction and/or improvement of libraries. Improvement of toy libraries. Creation of green areas and vegetable gardens. Creation, equipment and/or expansion of sports areas (playgrounds, courts, other spaces for sports). Creation and/or equipment of recreation and cultural areas. Improvement of toilet cleaning conditions. Implementation of technologies in classrooms. Creation and/or equipment of school canteens
Change of pedagogical methodologies for a significant learning	Significant learning - teaching. Development of pedagogical activities. Practical classes. Contextualized education. Personalized education. Creative methodologies to increase motivation. Reduction of home works. Development of didactic activities. Modification of pedagogical methodologies. Development of didactic activities. Modification of teaching methodologies

**Table 14** (continued)

Category	Descriptors
Wider coverage and access to higher education	Lower cost of higher education;. Entrance exams drills for public universities. Dissemination of information on scholarships and opportunities for access to higher education. Priority access for strata 1 and 2. Free entrance exams to public universities. Agreements to access higher education directly from school
Online classes and access limitations	No internet. No computer. Difficult connection. No more virtual classes. Semi face-to-face classes
Access and use of new technologies in education	Increased use of technologies. Increased involvement of technology in education
Development of life skills and competencies	Learning to solve real life problems. Inclusion of technical courses. Education for the future. Education that encourages critical thinking. Education in life skills: interviewing, completing a resume. Sex education. Emotional education: mental health (prevention of anxiety, depression). Life skills education (taxes, politics, local, national and global realities). Impact of daily actions on climate change (learning to recycle). Financial life management
Need for a foreign language proficiency	Second language training. Learning languages other than English. Bilingual schools. Higher demands in second language courses
Strategies to improve school coexistence	Playful spaces that reinforce good coexistence. Improve coexistence among students. Mitigate actions of school violence and bullying. Greater attention by teachers to fights, bullying and conflicts among students. Reduce indiscipline. Increase security and prevent theft. Respect the free development of personality. Reduce discrimination (race, gender, social class, etc.)

## B.2.2 Teachers

**Table 15** Categories and descriptors used during the triangulation process and obtained from the manual categorization in the Teachers group - English version

Category	Descriptors
Greater interaction of the family in the educational process	Parents' association (recreational activities, pedagogical inductions to strengthen the pedagogical processes). Workshops and training for parents. Greater interaction between the teacher and parents. Participation of the educational community. Greater interaction and commitment between parents and the educational community. Participation of the family in training. Greater commitment and integration of parents in education
Access to and use of new technologies in education	Having technological resources. Use and management of smart boards. Having technology rooms. Access to computers, internet, etc. Use of interactive tools for classes. Training on technological tools. Technology resources to improve efficiency and productivity in the classroom
Education that integrates multiple intelligence	Emotional intelligence: assertive communication, autonomy, confidence; money management. Healthy habits

**Table 15** (continued)

Category	Descriptors
Development of skills and competencies for the comprehensive and integral training of students	Critical thinking (critical skills, critical reading, critical attitude). Physical education training (corporal and psychomotor). Humanistic training. Development of artistic skills and/or skills that encourage creativity
More flexible curricular changes appropriate to the different learning processes of each student	Curriculum participative by all agents. Curriculum according to the needs of the country's context. Curricular innovation and flexibility. Transversal and integral curriculum in teaching and learning process
Change of pedagogical methodologies for a significant learning	Methodologies that include the implementation of projects, use recreational-didactic strategies and methodologies that motivate the student seeking to apply the students' knowledge in their daily lives and in specific contexts and developing autonomy in the student. Personalized methods. More applied methodologies based on projects and problems
Changes in class size to ensure educational quality	Class sizes of more than 30 students. Smaller learning groups. Very large groups with less effective learning. Control of class sizes

### B.2.3 Parents

**Table 16** Categories and descriptors used during the triangulation process and obtained from the manual categorization in the Parents group - English version

Category	Descriptors
Greater interaction of the family in the educational process	Generating responsibility from home. Forming a parents' association (training and guidance to parents). Having a lot of accompaniment from parents. More follow-up and supervision of parents in the educational process. Educating parents to teach values to their children. Generating commitment from parents for a better understanding of the educational and family aspects. Open communication between parents and teachers
Instruction in Values	Values such as spirituality, care, respect, honesty, valuing the environment, ethics, codes of conduct, values in coexistence and responsibility
Development of skills and competencies for life	Changes to a more meaningful education according to the needs and challenges of real life. Curricula adjusted to the current needs of the country (theoretical-practical contents). Knowledge in personal finance classes to develop their own companies (entrepreneurship) - labour education (necessary administrative procedures in adult life: make appointments, get credit and debit cards, among others). Having an education focused on daily life, knowing more about your country (national and world history). Tools to face a job position. Preparation in economic plans to start their own companies
Access and use of new technologies in education	Greater use of technologies to transmit knowledge. Equitable access to technology. More educational sessions on the use of technological tools

**Table 16** (continued)

Category	Descriptors
Education focused on the practice of theoretical knowledge	Methodologies involving more practice and less content. Practical classes. Laboratory development. Learning-by-doing. Search and teaching of arts and crafts
Changes in the traditional education	Developing student autonomy. Personalized methodologies (student preferences are taken into account). Knowledge transmission in an attractive way. More interactive education. Methodologies that encourage students to learn. Encourage reading and development of critical thinking
Teaching interpersonal skills	resilience. emotion management. assertive communication. conflict resolution. Sex education classes (along with these, family planning issues for both men and women)
Vocational training focused on talents-hobbies, oriented to the life projects	Teaching how to make a life project. Focusing on topics that will help children to develop in the future. Elaboration of projects that will lead them to new challenges. Contents and activities focused on the interests, aptitudes and talents of each student. Projects around different occupations. Choice of extracurricular subjects depending on the interests. Implementation of vocational guidance. Development of talents and qualities to enhance the strengths of students
Wider coverage and access to higher education	Free access to higher education. Equal opportunity to access higher education. Equal access to public and private higher education. Quality higher education for all
Online classes and access limitations	Access to tools for virtual classes. Limitations of the internet as an educational resource. Negative impact of virtual classes on socialization. Lack of pedagogical structure in virtual classes. Need to return to face-to-face classes

**Acknowledgements** We gratefully acknowledge the contribution of Bogota secretary of education during instrument design and data gathering.

**Funding** The authors have no financial or proprietary interests in any material discussed in this article.

## Declarations

**Competing interests** The authors have no financial or proprietary interests in any material discussed in this article.

## References

1. Bankauskaite V, Saarelna O (2003) Why are people dissatisfied with medical care services in lithuania? a qualitative study using responses to open-ended questions. *Int J Qual Health Care* 15(1):23–029
2. Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. *J Mach Learn Res* 3:993–1022
3. Buenaño-Fernandez D, González M, Gil D et al (2020) Text mining of open-ended questions in self-assessment of university teachers: an lda topic modeling approach. *IEEE Access* 8:35 330:318–35
4. Cheng X, Cao Q, Liao SS (2020) Covid19: an overview of literature on covid-19, mers and sars: using text mining and latent dirichlet allocation. *Journal of Information Science*. p 0165551520954674
5. El Akrouchi M, Benbrahim H, Kassou I (2021) End-to-end lda-based automatic weak signal detection in web news. *Knowl Based Syst* 212(106):650
6. Erkens M, Bodemer D, Hoppe HU (2016) Improving collaborative learning in the classroom: text mining based grouping and representing. *Int J Comput-Support Collab Learn* 11(4):387–415

7. Faherty VE (2009) *Wordcraft: applied qualitative data analysis (QDA): tools for public and voluntary social services*. Sage
8. Jelodar H, Wang Y, Yuan C et al (2019) Latent dirichlet allocation (lda) and topic modeling: models, applications, a survey. *Multimed Tools Appl* 78(11):15,169–15,211
9. Kumari R, Jeong JY, Lee BH et al (2019) Topic modelling and social network analysis of publications and patents in humanoid robot technology. *J Inf Sci*, pp 0165551519887878
10. Kyriakopoulou A, Kalamboukis T (2013) The impact of semi-supervised clustering on text classification. In: *Proceedings of the 17th Panhellenic Conference on Informatics*, pp 180–187
11. Liu L, Tang L, Dong W et al (2016) An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus* 5(1):1–22
12. Lu Y, Mei Q, Zhai C (2011) Investigating task performance of probabilistic topic models: an empirical study of pls and lda. *Inf Retr* 14(2):178–203
13. Mahmoud M, Dafoulas G, Abd ElAziz R et al (2020) Learning analytics stakeholders' expectations in higher education institutions: a literature review. *Int J Inf Learn Technol*
14. Mohammadi E, Karami A (2020) Exploring research trends in big data across disciplines: a text mining analysis. *J Inf Sci*, pp 0165551520932855
15. Nanda G, Douglas KA, Waller DR et al (2021) Analyzing large collections of open-ended feedback from mooc learners using lda topic modeling and qualitative analysis. *IEEE Trans Learn Technol*
16. Nguyen DQ, Billingsley R, Du L et al (2015) Improving topic models with latent feature word representations. *Transactions of the Association for Computational Linguistics* 3:299–313
17. Pope C, Van Royen P, Baker R (2002) Qualitative methods in research on healthcare quality. *BMJ Quality and Safety* 11(2):148–152
18. Roberts ME, Stewart BM, Tingley D et al (2014) Structural topic models for open-ended survey responses. *Am J Polit Sci* 58(4):1064–1082
19. Romanowski MH, Ellili-Cherif M, Al Ammari B et al (2013) Qatar's educational reform: the experiences and perceptions of principals, teachers and parents
20. Runge CE, Waller M, MacKenzie A et al (2014) Spouses of military members' experiences and insights: qualitative analysis of responses to an open-ended question in a survey of health and wellbeing. *PLoS one* 9(12):e114–755
21. Ten Kleij F, Musters PA (2003) Text analysis of open-ended survey responses: a complementary method to preference mapping. *Food Qual Prefer* 14(1):43–52
22. Tinsley HE, Weiss DJ (1975) Interrater reliability and agreement of subjective judgments. *J Couns Psychol* 22(4):358
23. Tutubalina E, Nikolenko S (2018) Exploring convolutional neural networks and topic models for user profiling from drug reviews. *Multimed Tools Appl* 77(4):4791–4809
24. Whittle S, Whelan B, Murdoch-Eaton D et al (2007) Dream and beyond; studies of the educational environment as a means for its enhancement. *Education for health* 20(1):7
25. Yan X, Guo J, Lan Y et al (2013) A biterm topic model for short texts. In: *Proceedings of the 22nd international conference on World Wide Web*, pp 1445–1456
26. Zuo Y, Zhao J, Xu K (2016) Word network topic model: a simple but general solution for short and imbalanced texts. *Knowl Inf Syst* 48(2):379–398

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.