Intelligent Tutoring System: A Bibliometric Analysis and Systematic Literature Review

Oscar-Andrés Cuéllar-Rojas
Universitat Politècnica de València: Universitat Politecnica de Valencia

Mauricio Hincapié
Institución Universitaria Pascual Bravo: Institucion Universitaria Pascual Bravo

Manuel Contero
Universitat Politècnica de València: Universitat Politecnica de Valencia

David Güemes-Castorena (guemes@tec.mx)
Tecnologico de Monterrey  https://orcid.org/0000-0002-6346-2873

Research

Keywords: tutoring system, bibliometric analysis, literature review, text mining, educational innovation

DOI: https://doi.org/10.21203/rs.3.rs-673038/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

This article is a bibliometric analysis and literature review, having a central axis the evaluation mediated by Intelligent Tutors Systems (ITS) in education, seeking to establish state of the art on the implementations executed in the last 42 years and their impact on the evaluation process. It was based on a bibliometric analysis of 1,890 abstracts, allowing to establish the main information sources in the field. The first filter was carried out with R software and bibliometric techniques with a general search equation that allowed access to all the production of ITS registered in Scopus; this analysis used keywords and summaries. Subsequently, with the help of artificial intelligence, text mining was used to identify topics of interest in the scientific community, followed by new filtering. Finally, the selected full texts were analyzed with NVIVO software extracting emerging challenges in the field, obtaining 164 complete texts for analysis. Among the main findings, the primary purpose of evaluation in ITS was summative, peer evaluation and self-evaluation did not have the same level of importance as hetero evaluation, and ITS focus was quantitative; all this allowed us to conclude that the analyzed texts did not implement a holistic perspective.

Introduction

According to Alvarez de Zayas [1], assessment is a systemic, holistic, and dialectical process; in other words, a complex process. However, this conception of evaluation does not always correspond to what those involved in educational processes put into practice. For example, in higher education, it is common that the preferred instrument for collecting information is the exam [2]. Confounding evaluating with grading, measuring, correcting, classifying, or examining and focusing attention on the quantitative aspect [3]. Although the grading process is related to evaluation and provides valuable data for decision making (see Fig. 1), they need to be complemented with multiple instruments that integrate qualitative and continuous aspects that allow transforming classroom dynamics and not only at the end of the academic periods. In other words, they must be aligned with the true meaning of evaluation, a formative, regulatory, pedagogical, communicative, and environmental sense.

In the case of basic sciences, the misinterpreted evaluation focused on results aggravates problems of performance, grade repetition, and, in some cases, dropout. For example, according to Castillo-Sanchez et al., [4], one of the leading causes of repetition in the first mathematics course is low academic performance in the first partial exam.

Introductory science courses are conventionally graded through exams, with percentage distribution, depending on the university. For example, in the Mathematics School at the National University of Colombia, there are three midterms of 25%, 30%, and 30%, respectively, and a short exam of 15% [5]; this implies that the student receives feedback on his learning process only in some specific moments and not in all classes.

However, given this approach, there is a question that is difficult to avoid. How to implement an evaluation process that overcomes these difficulties in courses with many students? This question has already been addressed, although not resolved. Digital Technologies offer the educational community a wide range of possibilities to collect information, such as interactive videos, simulations, and surveys, among others [7]. All of them configurable to be assessed automatically, without investing excessive teacher time. However, if these tools were implemented, it would continue without solving the fundamental evaluative aspect. What decisions to make with the data? Or even more complex, how to analyze these data?

One of the favorable environments for these implementations is that of Intelligent Tutoring System (ITS). There it is possible to transition from exam-centered grading to one that draws on multiple instruments. In this context, the student receives constant feedback, both cognitively and metacognitively [8].

The main task of an ITS is to evaluate students' knowledge acquisition throughout the education process. In general, Adaptive ITS provides learning environments in which all relevant information about students is kept and used to guide them [9].

ITS uses artificial intelligence principles and methods, for example, Neural Networks, to make inferences and learn autonomously. Thanks to this characteristic, ITS is adaptive since it alters its structure, functionality, or interface for the user and their needs [10].

ITS has different configurations according to the application context, but four modules stand out in educational courses: 1) pedagogical module, 2) student module (diagnosis), 3) expert module, and 4) communications module. These modules are complemented by the models created from the data they provide, represented in blue. (See Fig. 2).

This structure integrates naturally with massive courses, favoring learning environments with little teacher interaction. Student and teacher interactions with these modules produce large volumes of mixed data. Unfortunately, this type of information is difficult to analyze on a massive scale. Bearing in mind that Massive Online Course has exceeded 180 million students [12], and the number of participants per course easily exceeds 1,000 in some of them [13]. These figures justify the use of mass-grading strategies; with them, it is possible to achieved constant and automatic feedback, minimizing the interaction with the tutor, turning the student into the protagonist of the learning process. However, the amount of data generated by this constant interaction grows exponentially and quickly, exceeding the human capacity to analyze them and make decisions that are not always quantitative in nature.

This system responds to qualitative questions about each student, as specific as:

1. Which of the concepts covered in class require further study?
2. What are the levels of performance in the fundamental competencies of the course from the first class?
3. What methodological adjustments are required in the course to favor the student process?
4. What curricular adjustments are necessary to favor the development of the competencies offered by the course?
5. What feedback do teachers and students require to make decisions that favor the acquisition of the competencies offered by the course?
Furthermore, all those related to the individual process of the subjects are complex even for a conventional number of students, and since the evaluative processes of this level of personalization require an investment of time on behalf of the educational actors that do not correspond to the implementation model (maximizing the number of participants, minimizing tutors).

Thinking about these tasks for massive groups requires an intelligent data processing system that learns from the data and acts as a virtual master, performing accurate decision-making evaluation. However, the approaches to this problem are still under development. Fundamental variables have been considered [14][15]. For example, students’ self-regulation or motivation have been included in some ITS. However, aspects such as diagnostic, formative, and summative evaluation have not been considered together. Therefore, a systematic review was carried out to identify and evaluate articles that propose implementations of evaluation systems using machine learning techniques for massive volumes of data.

**Methodology**

A funnel system is proposed to access a broad spectrum of information and to have an objective view of it, with three filtering moments to select the complete papers included for analysis (see Fig 3).

The first filter was made with R software and bibliometric techniques. Then, a general search equation allowed access to all-time production on ITS registered in Scopus (only papers were selected). This analysis was carried out using keywords and summaries.

Subsequently, with the help of artificial intelligence, text mining was used to identify topics of interest in the scientific community, followed by new filtering. The selected full texts were analyzed with NVIVO software to extract emerging challenges in the field. This study aims to answer the following questions:

Q1: What is the ITS primary evaluation purpose?

Q2: What is the main evaluating agent (in evaluation processes)?

Q3: What is the main approach used in the selected ITS?

Q4: Is the ITS evaluation process implemented holistically?

These questions arise from the need to understand evaluation in the context of learning, in particular deep learning. Specifically, a holistic and complex evaluation that can account for the student’s capacity for critical analysis of new ideas and their integration with previous knowledge, thus favoring understanding and retention in the long term to later be used to solve problems in different contexts.

An evaluation that accounts for summative aspects, but also for the levels of cognitive skills such as “analysis” (comparing, contrasting) and “synthesis” (integrating knowledge in a new dimension), integrated with metacognitive aspects that promote understanding and application of lifelong learning can be considered a holistic evaluation.

**Bibliometric analysis:**

With the search equation, *intelligent tutoring* the following results presented in Table 1 were obtained. However, it is crucial to bear in mind that this general equation is only considered since it is expected to obtain new filtering criteria that will lead to a more refined equation.

**Table 1** Characteristics of the data

| Main information about data               |  |
|------------------------------------------|---|
| Timespan                                 | 1979 ↔ 2021 |
| Sources (Journals)                       | 618 |
| Documents                                | 1,890 |
| Average citations per document           | 21.12 |
| Document types                           |  |
| Article                                  | 1,890 |
| Authors                                  | 3,819 |
| Authors collaboration                    |  |
| Single-authored documents                | 322 |
| Documents per Author                     | 0.495 |
| Authors per Document                     | 2.02 |
A total of 1,890 results were found in Scopus, covering 42 years of academic production. The texts considered were articles published in specialized journals, although it is recognized that this field of knowledge has important dissemination through conferences. However, due to the objective of the study to identify structured knowledge with a high level of depth, conference papers were not included in this analysis. Thus, a total of 3,819 authors were considered in this initial search.

The academic production origin was in 1979; in 2014, it reached its maximum (105 papers,) and since 2016, such production has slightly decreased (Fig 4.)

Fig 5 shows that the largest source of texts was the International Journal of Artificial Intelligence in Education, classified in Q1. Fig 6 shows the top 5 most cited journals in relation to ITS. The journal Computers and Education stands out with a total of 4,814 citations.

The main authors by total citations in the chosen period are presented in Fig 7. For example, Kenneth R. Koedinger, professor of human-computer interaction and psychology at Carnegie Mellon University, is the founding and current director of the Pittsburgh Learning Science Center, with 2,112 citations. Garfield claimed that Keywords Plus terms could capture an article's content with greater depth and variety [16]. However, Keywords Plus is as effective as Author Keywords in bibliometric analysis investigating the knowledge structure of scientific fields, but it is less comprehensive in representing an article's content [17].

In Fig 8, computer-aided instruction stands out as the main topic, representing 17% of the frequencies examined in the text references. Finally, For the elaboration of Fig 9, it was considered that the co-occurrences could be normalized using similarity measures such as the Salton cosine, the Jaccard index, the equivalence index, and the strength of association [18].

The selected algorithm was the strength of the association since it is proportional to the relationship between the observed number of co-occurrences of objects $i$ and $j$, and the expected number of co-occurrences of objects $i$ and $j$ under the assumption that occurrences of $i$ and $j$ are statistically independent.

For the grouping strategy, "Walktrap" was selected as one of the best alongside "Louvain" [19]. The graph is interpreted considering the following characteristics:

1. Centrality / Periphery (Position)
2. Dimension of the bubble (number of citations)
3. Strength of relationships (links)
4. Clusters (and density)
5. Bridges.

The colors represent the groups to which each word belongs. In this case, there are three groups. In the first one in red, the theme of computer-aided instruction is dominant in citations. There is no central theme in the citation in the green one but in relationships, and it is Expert Systems relating topics of interest such as artificial intelligence. Finally, the third group, colored blue, seems to be a subgroup of the first one focused on educational issues.

**Text Mining**

Although the bibliometric analysis finds the authors and journals with the most impact in the specific field, the possible thematic fields based on the analysis of the Keywords Plus and a classification of these in groups it is necessary to perform additional analysis to identify more specific thematic groups, for which the Software Knime [20] was used.

Fig 10 shows the scheme under which the database downloaded from Scopus was processed. Data was previously filtered from 2003 when a production peak occurred and is of interest. Finally, in this analysis, all the abstracts of the selected papers were considered.

Search terms:

```
TITLE-ABS-KEY("intelligent tutoring System") AND (LIMIT-TO (DOCTYPE,"ar")) AND (LIMIT-TO (PUBYEAR,2021) OR LIMIT-TO (PUBYEAR,2020) OR LIMIT-TO (PUBYEAR,2019) OR LIMIT-TO (PUBYEAR,2018) OR LIMIT-TO (PUBYEAR,2017) OR LIMIT-TO (PUBYEAR,2016) OR LIMIT-TO (PUBYEAR,2015) OR LIMIT-TO (PUBYEAR,2014) OR LIMIT-TO (PUBYEAR,2013) OR LIMIT-TO (PUBYEAR,2012) OR LIMIT-TO (PUBYEAR,2011) OR LIMIT-TO (PUBYEAR,2010) OR LIMIT-TO (PUBYEAR,2009) OR LIMIT-TO (PUBYEAR,2008) OR LIMIT-TO (PUBYEAR,2007) OR LIMIT-TO (PUBYEAR,2006) OR LIMIT-TO (PUBYEAR,2005) OR LIMIT-TO (PUBYEAR,2004) OR LIMIT-TO (PUBYEAR,2003) )
```

Fig 11 shows the workflow developed in Knime, with which it was possible to analyze 1,369 abstracts and extract the hidden thematic structure, identifying the topics that best describe a set of documents.

Table 2 describes each item presented in figure 11.

**Table 2. Item description of Knime workflow**
Table 3 has the description of each item presented in Figure 12.

**Table 3. Metanode preprocessing item description**

| Image | Name            | Description                                                                 |
|-------|-----------------|-----------------------------------------------------------------------------|
| ![Image](image1.png) | Excel reader    | It allows incorporating a database obtained from Scopus in Excel format      |
| ![Image](image2.png) | Missing Value Column Filter | This node removes all columns from the input table that contain more missing values. |
| ![Image](image3.png) | Strings to Document | It converts the specified strings to documents. For each row, a document will be created and attached to that row. |
| ![Image](image4.png) | Preprocessing | This is a metanode, which groups several nodes responsible for multiple tasks, including Part of Speach tagging, lemmatization, stop word, number, filtering. Inside this metanode are the elements shown in Fig 12. |

One of the main elements of this algorithm is the Topic Extractor, with which it is possible to achieve the following:

- Automatically finds the top K topics with the most relevant N keywords discussed in a collection of unlabeled documents (considered unsupervised).
- It represents documents as random mixtures over latent topics, where a distribution over words characterizes each topic.
- Syntax or order of the words in the document is not important (bag of words model).
- Document order is not important.
- The same word can belong to different topics.
- The number of topics needs to be selected/known in advance.
- Two important hyperparameters of the Dirichlet distributions:
  - \( \alpha \) Controls the per-document topic distribution.
  - \( \beta \) Controls the per-topic word distribution.

This process is known as the Simple parallel threaded implementation of LDA [21][22] (see Figure 13).

In Figure 14, the process for dimensional reduction is presented, and in Table 4, there is the description of each item in figure 14:

**Table 4. Topic extractor items description**

| Image | Name            | Description                                                                 |
|-------|-----------------|-----------------------------------------------------------------------------|
| ![Image](image5.png) | Punctuation Erasure | Removes all punctuation characters of terms contained in the input documents |
| ![Image](image6.png) | Number Filter    | Filters all the numerical values present in the entered documents.          |
| ![Image](image7.png) | N Chars Filter   | Filters all terms contained in the input documents with less than the specified number of N characters |
| ![Image](image8.png) | Stanford Tagger  | This node assigns each term a part of speech tag.                           |
| ![Image](image9.png) | Stanford Lemmatizer | Lemmatizes terms contained in the input documents.                      |
| ![Image](image10.png) | Case converter   | Uppercase and lowercase converter                                            |
t-SNE is a manifold learning technique. It is most often used for visualization purposes and can capture nonlinear structures in the data.

Color Manager  Assign a color label to data groups

Joiner  Joins two tables in a database-like way.

Interactive visualization  This is the metanode in charge of allowing the visualization of emerging topics in an interactive way. The nodes found inside are shown in Fig 15.

Table 5 describes the items in the Metadone interactive visualization in Figure 15.

Table 5. Metadone interactive visualization items description

| Image | Name      | Description                                                                 |
|-------|-----------|-----------------------------------------------------------------------------|
|       | **t-SNE**| t-SNE is a manifold learning technique. It is most often used for visualization purposes and can capture nonlinear structures in the data. |
|       | **Color Manager** | Assign a color label to data groups                                           |
|       | **Joiner** | Joins two tables in a database-like way.                                    |
|       | **Interactive visualization** | This is the metanode in charge of allowing the visualization of emerging topics in an interactive way. The nodes found inside are shown in Fig 15. |

In Table 6 is a description of the items in the Metanode Tagging.

Table 6. Metanode Tagging items description

| Image | Name      | Description                                                                 |
|-------|-----------|-----------------------------------------------------------------------------|
|       | **GroupBy** | It is responsible for grouping the rows of a table by the unique values in the columns of the selected group. |
|       | **Table View** | Displays data in an HTML table view. The view offers several interactive features, as well as the possibility to select rows. |
|       | **Scatter Plot** | With this node, a scatter plot is obtained.                                |
|       | **Tagging** | This metanode groups the nodes presented in Fig 16. It is the last metanode in this section. It is done in labeling that will allow viewing the word cloud and the texts associated with each topic. |
After going through these nodes, the algorithm returned the following result. In Fig 17, all the selected terms are classified into five topics from the 1,369 abstracts; each topic requires interpretation. However, the focus of the analysis was to determine if some of them were related to the category of interest: evaluation.

The program interface allows the analyst to explore each of the five topics, as shown in Fig 18.

For example, topic_0 contains the terms game, instruction, intelligent, language, reading, skill, strategy, study, system. In the "document" column, the text and the weight of contribution to each of the terms were displayed.

The topic_3, represented in yellow in Fig 19, emerges naturally among the analyzed abstracts. The terms that compose it are affective, assessment, data, emotion, method, model, performance, result, student, and system, all of them with high values for this studio. Therefore, this result— with high values— was the selection criteria to link the full texts analyzed in Nvivo in the next phase.

One hundred sixty-four papers were selected from the text mining of the emerging group represented in Fig 19. It is essential to consider that the weight of the term assessment is not high compared to the other terms identified in topic_3 and even less compared to the total number of identified terms, as shown in Fig 20.

**Results**

The results are presented in this section; a year-wise representation is given in Fig 21.

These results are characterized by research questions posed earlier in this study. The variables of selected studies are presented in Table 7.

Q1: What is the ITS primary evaluation purpose?

Q2: What is the main evaluating agent (in evaluation processes)?

Q3: What is the main approach used in the selected ITS?

Q4: Is the ITS evaluation process implemented holistically?

**Table 7.** Analyzed variables
Three pillars were considered to answer these questions: the purpose, agent, and evaluation approach in each of the selected papers. With the help of the Nvivo program [23], a case has been created for each. Subsequently, the percentages of their presence in the selected complete papers have been identified with a search matrix. Finally, considering that a proper holistic evaluation uses all the pillars comprehensively, the holistic column has been completed, finding that none of the studies possess the simultaneous presence of all the sub-variables. Table 8 summarizes the results and identifies if the study was holistic or not.

**Table 8. Results**

| Variables | Diagnostic evaluation | Formative evaluation | Summative evaluation | Self-Assessment | Co-evaluation | Hetero evaluation | Qualitative evaluation | Quantitative evaluation | Holistic |
|-----------|-----------------------|----------------------|----------------------|-----------------|---------------|-------------------|-----------------------|------------------------|---------|
| Purpose   |                       |                      |                      |                 |               |                   |                       |                        |         |
| Evaluating Agent | Self-Assessment |                      |                      |                 |               |                   |                       |                        |         |
| Co-evaluation |               |                      |                      |                 |               |                   |                       |                        |         |
| Hetero evaluation |               |                      |                      |                 |               |                   |                       |                        |         |
| Approach   | Qualitative evaluation |                  |                      |                 |               |                   |                       |                        |         |
| Quantitative evaluation |                  |                      |                      |                 |               |                   |                       |                        |         |
| Holistic   | Yes                   |                      |                      |                 |               |                   |                       |                        |         |
|           | No                    |                      |                      |                 |               |                   |                       |                        |         |
| Paper | Purpose | A: Diagnostic | B: Formative | C: Summative | Evaluating Agent | Approach |
|-------|---------|---------------|--------------|--------------|-----------------|----------|
|       |         |               |              |              | A: Hetero-assessment | B: Peer assessment | C: Self-Assessment | A: Qualitative | B: Quantitative | Holistic |
|       | A:     |               |              |              |                 |          |                    |               |                |         |
| 1     | [24]   | 16.67%        | 0%           | 83.33%       | 94.74%          | 0%       | 5.26%              | 0%            | 100%          | No       |
| 2     | [25]   | 0%            | 0%           | 100%         | 91.49%          | 0%       | 8.51%              | 0%            | 100%          | No       |
| 3     | [26]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 4.35%         | 95.65%        | No       |
| 4     | [27]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 0%            | 100%          | No       |
| 5     | [28]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 4.76%         | 95.24%        | No       |
| 6     | [29]   | 0%            | 0%           | 100%         | 52.47%          | 0%       | 47.53%             | 0%            | 100%          | No       |
| 7     | [30]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 0%            | 100%          | No       |
| 8     | [31]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 0%            | 100%          | No       |
| 9     | [32]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 0%            | 100%          | No       |
| 10    | [33]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 11.11%        | 88.89%        | No       |
| 11    | [34]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 0%            | 100%          | No       |
| 12    | [35]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 1.96%         | 98.04%        | No       |
| 13    | [36]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 0%            | 100%          | No       |
| 14    | [37]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 0%            | 100%          | No       |
| 15    | [38]   | 8.33%         | 0%           | 91.67%       | 100%            | 0%       | 0%                 | 0%            | 100%          | No       |
| 16    | [39]   | 100%          | 0%           | 0%           | 100%            | 0%       | 0%                 | 0%            | 0%            | No       |
| 17    | [40]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 50%           | 46.15%        | No       |
| 18    | [41]   | 100%          | 0%           | 0%           | 100%            | 0%       | 0%                 | 0%            | 0%            | No       |
| 19    | [42]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 10%           | 90%           | No       |
| 20    | [43]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 85.06%        | 25.30%        | No       |
| 21    | [44]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 9.64%         | 1.89%         | No       |
| 22    | [45]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 100%          | 0%            | No       |
| 23    | [46]   | 17.65%        | 0%           | 82.35%       | 100%            | 0%       | 0%                 | 0%            | 100%          | No       |
| 24    | [47]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 0%            | 100%          | No       |
| 25    | [48]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 0%            | 100%          | No       |
| 26    | [49]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 0%            | 100%          | No       |
| 27    | [50]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 0%            | 100%          | No       |
| 28    | [51]   | 0%            | 0%           | 0%           | 0%              | 0%       | 0%                 | 0%            | 0%            | No       |
| 29    | [52]   | 0%            | 0%           | 83.33%       | 100%            | 0%       | 0%                 | 0%            | 100%          | No       |
| 30    | [53]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 14.29%        | 85.71%        | No       |
| 31    | [54]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 0%            | 100%          | No       |
| 32    | [55]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 0%            | 100%          | No       |
| 33    | [56]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 0%            | 100%          | No       |
| 34    | [57]   | 8%            | 4%           | 88%          | 100%            | 0%       | 0%                 | 0%            | 100%          | No       |
| 35    | [58]   | 2.94%         | 17.65%       | 79.41%       | 74.51%          | 0%       | 25.49%             | 0%            | 100%          | No       |
| 36    | [59]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 0%            | 100%          | No       |
| 37    | [60]   | 0%            | 0%           | 100%         | 54.72%          | 0%       | 45.28%             | 0%            | 100%          | No       |
| 38    | [61]   | 16.67%        | 0%           | 83.33%       | 75.68%          | 0%       | 24.32%             | 0%            | 100%          | No       |
| 39    | [62]   | 0%            | 0%           | 100%         | 100%            | 0%       | 0%                 | 0%            | 100%          | No       |
| 40    | [63]   | 0%            | 42.11%       | 57.89%       | 100%            | 0%       | 0%                 | 0%            | 100%          | No       |
| 41    | [64]   | 0%            | 35.48%       | 64.52%       | 100%            | 0%       | 0%                 | 0%            | 100%          | No       |
Next, we present each research question and its results.

Q1: What is the main purpose of the evaluation in these ITS?

According to the data found, the primary purpose of the evaluation is summative; that is, most of the evaluation sections in the ITS analyzed tried to establish reliable balances of the results obtained focusing on the collection of information and the elaboration of instruments that allow reliable measurements of the knowledge to be evaluated at the end of a teaching-learning process.

Q2: What is the main evaluating agent (in evaluation processes)?
The main evaluating agents were those external to the student or their peers; that is, hetero evaluation was prioritized. This is consistent with the purpose found in question 1. Most ITS identify gaps or “weak spots” that need to be reinforced before moving forward with the program and design redress activities aimed at the group or individuals who require it.

Q3: What is the main approach used in the selected ITS?

The main approach found was the quantitative one; this makes much sense since smart tutors use data to achieve process automation. However, qualitative approaches were evidenced to a lesser extent, and in some cases, the use of both was possible, thanks to the technological development that allows emotional interpretation and the participants’ language.

Q4: Is the evaluation process implemented in ITS holistic?

To answer this question, the criterion was the following: in each of the selected papers, diagnostic, formative, and summative evaluation elements were sought. It was also tracked whether the STI made use of heteroevaluation, peer review, and self-assessment. Furthermore, finally, it was determined whether it integrated qualitative and quantitative approaches. All this to account for a holistic assessment that favors deep learning. Texts that met all these criteria would be classified as holistic.

Under the criteria applied, it is possible to affirm that holistic designs were not found in the analyzed texts. Mainly, special attention is required to the diagnostic and formative evaluation. Furthermore, it is also necessary to encourage the participation of other agents in the evaluation processes of ITS, specifically peer evaluation and the participation of other actors, such as the family. Finally, the mixed approach can enrich the reading of the process; the qualitative evaluative aspects in ITS are a technical challenge; however, these can be included through properly trained bots.

Emerging challenges

Based on Table 8, it was possible to identify the analysis focuses and propose the following challenges.

**Demonstrate the pedagogical value of scaffolding by intelligent tutors.**

According to Arevalillo-Herráez [166], to facilitate problem-centered instructional models, it is necessary to provide scaffolding, that is, contingent support from another more capable person who helped others solve complex problems and acquire valuable skills in doing so, these include deep content learning, argumentation skills, and problem-solving skills. Traditionally providing this type of coaching requires small groups and personalized training processes.

With the help of intelligent tutoring systems, it is possible to provide this support in large groups; however, the expected learning outcomes of scaffolding respond to different variables, such as cognitive, motivational, or metacognitive aspects. In the cognitive aspects, it has been found that Intelligent Tutoring Systems favor significant progress. However, the motivational and metacognitive aspects require further research to demonstrate their pedagogical value. This can be evidenced by the priority given in the selected full texts to evaluating summative aspects.

**Link an efficient evaluation mechanism.**

Current trends indicate that online learning has become a vital learning mode; however, a holistic evaluation mechanism was not identified in the analyzed texts.

The learning performance assessment aims to assess what students learned during the learning process. It is usually summative or formative; however, both have been confused with the rating in some ITS, focusing on materializing a numerical value. This is clearly due to the learning framework in which each research is inscribed. However, to mobilize higher thinking skills such as problem-solving, critical thinking, or creativity (typical of deep learning) and according to the results found in Table 2, it is necessary to complement this approach with qualitative approaches.

**Use multiple data sources.**

The fundamental challenges to consider when thinking about an intelligent tutor are usually the data sources to feed the predictive models, which come from the summative assessment, such as the result of exercise A or the performance in unit B. However, it is crucial to determine the pedagogical value of the actions that led to these results and the implications of these data in predicting the participants’ performance [30][35].

**Need to link e-learning environments with intelligent tutoring systems.**

Assessment of students’ performance on exercises could delay the tutor’s feedback to students for days or even weeks. Then, in some cases, tutors may have to reduce the number of assignments given to their students due to time constraints. Especially in large-scale courses, accurate and meaningful evaluation is a demanding task for tutors. Moreover, accuracy is often difficult to achieve for both subjective and objective reasons.

**Possible solutions to the emerging remains.**

In the above discussion, several challenges were identified. To address them, the following research challenges are posed.

- **Understand and implement the difference between evaluating and grading.**
Intelligent Tutoring Systems require moving toward an interpretation of the numerical results, which allows for feedback as proposed by Daniel Wilson, director of the “Zero” project at Harvard University, who indicates that the feedback process consists of four ascending phases: Clarify, Value, Express concerns and Make suggestions, which allows focusing on communication with the student in the construction of meaning, towards the achievement of deep learning [187]. Currently, developments have focused on grading.

- **Designing a holistic framework.**

The theory of conscious processes, elaborated by Álvarez de Zayas, [1] is of a systemic, holistic, and dialectical nature, that is to say, complex. It presents a redefinition of the School as a space where teaching and systematization take place to essentially give way to the training process. An ITS designed under this perspective understands evaluation in a systemic, articulated, holistic, and dialectical way. The test is relational and is not the only instrument to obtain information about the teaching and learning processes. It includes aspects related to the purpose, the extension, the evaluating agents, the moments, the approaches, and comparison standards. Dialectically produced tools are used between components and between actors.

- **Focus on the process, not just the outcome.**

To provide a solution to this aspect, ITS must move toward formative evaluation, which implies collecting, analyzing, and identifying student progress (learning monitoring) and reflecting, providing feedback, reorienting, and creating support strategies for students (pedagogical use of the results). The latter is a technological challenge, which implies training the ITS not only with quantitative data.

- **Implement Learning Analytics Systems that impact the curriculum.**

When the evaluation process is done correctly, changes to the curriculum emerge naturally, enabling the student to access authentic deep learning. This line of research would imply establishing a framework that allows artificial intelligence to detect new learning goals for the student based on the analysis of mixed data.

**Conclusions**

The use of text mining was fundamental to extract knowledge from a wide field of academic production. Other researchers in different fields can use the workflow adapted in KNIME to optimize reading time and focus attention only on the aspects of interest.

Based on smart tutors' research, it was possible to identify that progress has been made in detecting concepts that require further study in the constant feedback to students and teachers in a personalized and automatic way. First, however, it is necessary to propose a framework that offers mixed feedback to students and teachers that facilitates decision-making based on implementing predictive methods, an evaluation that transcends the grading, which is possible thanks to the fusion between pedagogical and technological aspects.

Deep learning seeks to give meaning to new information; that is, it aims to incorporate a critical perspective on certain learning and, in doing so, favor its understanding to allow its long-term retention. Achieving it requires moving towards a complex evaluation that involved different forms of evaluation, actors, moments, approaches, and analysis.

The ITS requires moving towards an interpretation of the numerical results, allowing communication with the student to focus on constructing meaning towards a holistic evaluation. This holistic evaluation includes the student's participation and peers' diagnostic, formative, and summative aspects. These changes will make it possible to account for the depth of learning achieved.

Moving towards this type of evaluation involves analyzing quantitative and qualitative variables. Therefore, creating a framework that allows artificial intelligence to integrate all these variables and effectively communicate its results is necessary. In other words, and ITS is required, capable of assessing and measuring all variables related to deep learning and achieving a truly holistic assessment.

**Declarations**

**Ethics approval and consent to participate.**

Not applicable

**Consent for publication**

Not applicable

**Availability of data and materials**

All data generated or analyzed during this study are included in this published article.

**Competing interests**

The authors declare that they have no competing interests.

**Funding**
References

1. C. Álvarez de Zayas, *Los Enfoques. Las Tendencias*. Cochabamba. Editorial Edad de Oro., 2010.

2. H. Gibb et al., "Utility of recent studies to assess the National Research Council 2001 estimates of cancer risk from ingested arsenic," *Environ. Health Perspect.*, vol. 119, no. 3, pp. 284–290, 2011.

3. J. M. Álvarez, "El campo semántico de la evaluación. Más allá de las definiciones," *Evaluar para Conoc. examinar para excluir*, no. 1985, 2001, [Online]. Available: http://farq.edu.uy/estructura/unidades_de_gestion/uap/matevalaprend/Juan Manuel Alvarez Mendez.pdf.

4. M. Castillo-Sánchez, R. Gamboa-Araya, and R. Hidalgo-Mora, "Factores que influyen en la deserción y reprobación de estudiantes de un curso universitario de matemáticas," *Uniciencia*, vol. 34, no. 1, pp. 219–245, 2020.

5. O. A. Cuéllar Rojas, "Validación de una propuesta evaluativa integral para el curso de cálculo diferencial de la universidad nacional sede Medellín, basada en el uso de un LMS (Learning Management System), Moodle," *Fac. Ciencias*, 2013.

6. C. Hamodi, V. M. López-Pastor, and A. Pastor, "Medios, técnicas e instrumentos de evaluación formativa y compartida del aprendizaje en educación superior," *Perfiles Educ.*, vol. 37, pp. 146–161, Dec. 2014, doi: 10.1016/j.pe.2015.10.004.

7. D. Torres Mancera and D. Gago Saldaña, "Los Moocs Y Su Papel En La Creación De Comunidades De Aprendizaje Y Participación," *RIED. Rev. Iberoam. Educ. a Distancia*, vol. 17, no. 1, pp. 13–34, 2014, doi: 10.5944/ried.17.1.11150.

8. A. Shemshack and J. M. Spector, "A systematic literature review of personalized learning terms," *Smart Learn. Environ.*, vol. 7, no. 1, 2020, doi: 10.1186/s40561-020-00140-9.

9. C. Lemke, "Intelligent Adaptive Learning," *Vend. Supplied Whitepaper*, 2013, [Online]. Available: http://www-static.dreambox.com/wp-content/uploads/2013/03/white-paper-intelligent-adaptive-learning-21st-century-teaching-and-learning.pdf.

10. A. Anohina, "Advances in Intelligent Tutoring Systems: Problem-solving Modes and Model of Hints," *Int. J. Comput. Commun. Control*, vol. 2, no. 1, p. 48, 2007, doi: 10.15837/ijcc.2007.1.2336.

11. M. M. Al-Hanjori, M. Z. Shaath, and S. S. Abu-Naser, "Learning computer networks using intelligent tutoring system," 2017.

12. D. Shah, "Class Central," *By the numbers: MOOCs in 2020*, 2020. https://www.classcentral.com/report/mooc-stats-2020/.

13. C. Kivelä, "Informe sobre permanencia de estudiantes en MOOCs (Attrition and Retention Aspects in MOOC Environments)," *Austria.*, 2016.

14. R. Rajendran, S. Iyer, and S. Murthy, "Personalized Affective Feedback to Address Students’ Frustration in ITS," *IEEE Trans. Learn. Technol.*, vol. 12, no. 1, pp. 87–97, 2019, doi: 10.1109/TLT.2018.2807447.

15. E. M. Torres-Madroño, M. C. Torres-Madroño, and L. D. R. Botero, "Challenges and possibilities of ICT-mediated assessment in virtual teaching and learning processes," *Futur. Internet*, vol. 12, no. 12, pp. 1–20, 2020, doi: 10.3390/fi12120232.

16. E. Garfield and J. H. Sher, "Key words plus [TM]-algorithmic derivative indexing," *Journal-American Soc. Inf. Sci.*, vol. 44, p. 298, 1993.

17. J. Zhang, Q. Yu, F. Zheng, C. Long, Z. Lu, and Z. Duan, "Comparing keywords plus of WOS and author keywords: A case study of patient adherence research," *J. Assoc. Inf. Sci. Technol.*, vol. 67, no. 4, pp. 967–972, 2016.

18. N. J. van Eck and L. Waltman, "How to normalize co-occurrence data? An analysis of some well-known similarity measures," *J. Am. Soc. Inf. Sci. Technol.*, vol. 60, no. 8, pp. 1635–1651, 2009.

19. A. Lancerinetti, M. Kivelä, J. Saramäki, and S. Fortunato, "Characterizing the community structure of complex networks," *PLoS One*, vol. 5, no. 8, p. e11976, 2010.

20. M. R. Berthold et al., "KNIME-the Konstanz information miner: version 2.0 and beyond," *ACM SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 26–31, 2009.

21. D. Gaston, C. Newman, G. Hansen, and D. Lebrun-Grandie, "MOOSE: A parallel computational framework for coupled systems of nonlinear equations," *Nucl. Eng. Des.*, vol. 239, no. 10, pp. 1768–1778, 2009.

22. L. Yao, D. Mimno, and A. McCallum, *Efficient methods for topic model inference on streaming document collections*. 2009.

23. Q. S. R. NVivo, "NVivo qualitative data analysis software," 2020.

24. K. Muldner and W. Burleson, "Utilizing sensor data to model students’ creativity in a digital environment," *Comput. Human Behav.*, vol. 42, pp. 127–137, 2015, doi: 10.1016/j.chb.2013.10.060.

25. F. Alqahtani, S. Katsigiannis, and N. Ramzan, "Using Wearable Physiological Sensors for Affect-Aware Intelligent Tutoring Systems," *IEEE Sens. J.*, vol. 21, no. 3, pp. 3366–3378, 2021, doi: 10.1109/JSEN.2020.3023886.

26. M. T. Sanz García, J. A. González-Calero, D. Amau, and M. Arenvallo-Herráez, "Using reading comprehension to build a predictive model for the fourth-grade grade students’ achievement when solving word problems in an intelligent tutoring system," *Rev. Educ.*, vol. 2019, no. 384, pp. 41–69, 2019, doi: 10.4438/1988-592X-RE-2019-384-409.

Page 15/33
27. M. Van Amelsvoort, B. Joosten, E. Krahmer, and E. Postma, "Using non-verbal cues to (automatically) assess children's performance difficulties with arithmetic problems," Comput. Human Behav., vol. 29, no. 3, pp. 654–664, 2013, doi: 10.1016/j.chb.2012.10.016.

28. G. Zheng, S. E. Fancsali, S. Ritter, and S. R. Berman, "Using instruction-embedded formative assessment to predict student summative test scores and achievement levels in mathematics," J. Learn. Anal., vol. 6, no. 2, pp. 153–174, 2019, doi: 10.18608/jla.2019.62.11.

29. J. D. Gobert, Y. J. Kim, M. A. Sao Pedro, M. Kennedy, and C. G. Betts, "Using educational data mining to assess students' skills at designing and conducting experiments within a complex systems microworld,"Think. Ski. Creat., vol. 18, pp. 81–90, 2015, doi: 10.1016/j.tsc.2015.04.008.

30. J. Anderson, S. Betts, J. Ferris, J. Fincham, and J. Yang, "Using brain imaging to interpret student problem solving," IEEE Intell. Syst., vol. 26, no. 5, pp. 22–29, 2011, doi: 10.1109/MISS.2011.57.

31. J. R. Anderson, "Tracking problem solving by multivariate pattern analysis and Hidden Markov Model algorithms," Neuropsychologia, vol. 50, no. 4, pp. 487–498, 2012, doi: 10.1016/j.neuropsychologia.2011.07.025.

32. J. R. Anderson, S. Betts, J. L. Ferris, and J. M. Fincham, "Tracking children's mental states while solving algebra equations," Hum. Brain Mapp., vol. 33, no. 11, pp. 2650–2665, 2012, doi: 10.1002/hbm.21391.

33. V. Rus and D. Stefanescu, "Toward non-intrusive assessment in dialogue-based intelligent tutoring systems," Lect. Notes Educ. Technol., no. 9789128786633, pp. 231–241, 2016, doi: 10.1007/978-981-287-868-7_26.

34. B. Paaßen, C. Göpfert, and B. Hammer, "Time Series Prediction for Graphs in Kernel and Dissimilarity Spaces," Neural Process. Lett., vol. 48, no. 2, pp. 669–689, 2018, doi: 10.1007/s11063-017-9684-5.

35. P. Penumatsa et al., "The right threshold value: What is the right threshold of cosine measure when using latent semantic analysis for evaluating student answers?", Int. J. Artif. Intell. Tools, vol. 15, no. 5, pp. 767–777, 2006, doi: 10.1142/S021821960600293X.

36. J. Krivec and M. Guid, "The influence of context on information processing," Cogn. Process., vol. 21, no. 2, pp. 167–184, 2020, doi: 10.1007/s10339-020-00958-8.

37. J. Whitehill, Z. Serpell, Y. C. Lin, A. Foster, and J. R. Movellan, "The faces of engagement: Automatic recognition of student engagement from facial expressions," IEEE Trans. Affect. Comput., vol. 5, no. 1, pp. 86–98, 2014, doi: 10.1109/TAFCC.2014.2316163.

38. K. V. Kuk, I. Z. Milentijević, D. M. Randelović, B. M. Popović, and P. Čišar, "The design of the personal enemy - MIMLebot as an intelligent agent in a game-based learning environment," Acta Polytech. Hung., vol. 14, no. 4, pp. 121–139, 2017, doi: 10.12700/APH.14.4.2017.4.7.

39. S. Yang, S. Ding, and Z. Yao, "The theory about CD-CAT based on FCA and its application," Int. J. Distance Educ. Technol., vol. 7, no. 4, pp. 61–78, 2009, doi: 10.4018/jdet.2009062404.

40. J. K. Olsen, K. Sharma, N. Rummel, and V. Alevren, "Temporal analysis of multimodal data to predict collaborative learning outcomes," Br. J. Educ. Technol., vol. 51, no. 5, pp. 1527–1547, 2020, doi: 10.1111/bjet.12982.

41. F. Kabanza and K. Rousseau, "Teaching while selecting images for satellite-based forest mapping," Int. J. Knowledge-Based Intell. Eng. Syst., vol. 9, no. 3, pp. 183–189, 2005, doi: 10.3233/KES-2005-9302.

42. E. L. Snow, A. D. Likens, L. K. Allen, and D. S. McNamara, "Taking Control: Stealth Assessment of Deterministic Behaviors Within a Game-Based System," Int. J. Artif. Intell. Educ., vol. 26, no. 4, pp. 1011–1032, 2016, doi: 10.1007/s40593-015-0085-3.

43. F. Yang and F. W. B. Li, "Study on student performance estimation, student progress analysis, and student potential prediction based on data mining," Comput. Educ., vol. 123, no. October 2017, pp. 97–108, 2018, doi: 10.1016/j.compedu.2018.04.006.

44. I. Jraidi and C. Frasson, "Student's uncertainty modeling through a multimodal sensor-based approach," Educ. Technol. Soc., vol. 16, no. 1, pp. 219–230, 2013.

45. A. R. Abbasi, M. N. Dailey, N. V. Afzulpurkar, and T. Uno, "Student mental state inference from unintentional body gestures using dynamic Bayesian networks," J. Multimodal User Interfaces, vol. 3, no. 1, pp. 21–31, 2010, doi: 10.s12193-009-0023-7.

46. A. Abdi et al., "A new approach to personalization: Integrating e-learning and m-learning," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 26, no. 4, pp. 351–352, 2017, doi: 10.1007/s11423-012-9250-9.

47. I. Šarić-Grbić, A. Grubišić, L. Šerić, and T. J. Robinson, "Student clustering based on learning behavior data in the intelligent tutoring system," Int. J. Distance Educ. Technol., vol. 18, no. 2, pp. 73–89, 2020, doi: 10.4018/IJDET.2020041015.

48. J. Mostow and J. Beck, "Some useful tactics to modify, map and mine data from intelligent tutors," Nat. Lang. Eng., vol. 12, no. 2, pp. 195–208, 2006, doi: 10.1017/S1351324906004153.

49. E. Guzmán and R. Conejo, "Self-assessment in a feasible, adaptive web-based testing system," IEEE Trans. Educ., vol. 48, no. 4, pp. 688–695, 2005, doi: 10.1109/TE.2005.854571.

50. E. Alepis, M. Virvou, and K. Kabassi, "Requirements analysis and design of an affective bi-modal intelligent tutoring system: The case of keyboard and microphone," Stud. Comput. Intell., vol. 104, no. 2008, pp. 9–24, 2008, doi: 10.1007/978-3-540-77471-6_2.

51. J. Khalfallah and J. Ben Hadj Slama, "Relevant metrics for facial expression recognition in intelligent tutoring system," Lect. Notes Educ. Technol., no. 9789110241848, pp. 119–122, 2017, doi: 10.1007/978-981-10-2419-1_17.

52. W. Chen, J. Mostow, and G. Aist, "Recognizing young readers’ spoken questions," Int. J. Artif. Intell. Educ., vol. 21, no. 4, pp. 255–269, 2013, doi: 10.3233/JAI-130031.

53. D. J. Litman and K. Forbes-Riley, "Recognizing student emotions and attitudes on the basis of utterances in spoken tutoring dialogues with both human and computer tutors," Speech Commun., vol. 48, no. 5, pp. 559–590, 2006, doi: 10.1016/j.specom.2005.09.008.
54. R. D. Nielsen, W. Ward, and J. H. Martin, "Recognizing entailment in intelligent tutoring systems," Nat. Lang. Eng., vol. 15, no. 4, pp. 479–501, 2009, doi: 10.1017/S135312490999012X.

55. O. Castillo, P. Melin, W. Pedrycz, and J. Kacprzyk, "Recent Advances on Hybrid Approaches for Designing Intelligent Systems," Stud. Comput. Intell., vol. 551, 2014, doi: 10.1007/978-3-319-05170-3.

56. H. Kaya, T. Özkaptan, A. A. Salah, and F. Gürge, "Random discriminative projection based feature selection with application to conflict recognition," IEEE Signal Process. Lett., vol. 22, no. 6, pp. 671–675, 2015, doi: 10.1109/LSP.2014.2365393.

57. C. Y. Ting and S. Phon-Amnuaisuk, "Properties of Bayesian student model for INQPRO," Appl. Intel., vol. 36, no. 2, pp. 391–406, 2012, doi: 10.1007/s10489-010-0267-7.

58. H. Moradi, S. Moradi, and L. Kashani-Vahid, "Students' Performance Prediction Using Multi-Channel Decision Fusion," Stud. Comput. Intell., vol. 524, pp. 151–174, Nov. 2014, doi: 10.1007/978-3-319-02738-8-6.

59. C. N. Moridis and A. E. Economides, "Prediction of student's mood during an online test using formula-based and neural network-based method," Comput. Educ., vol. 53, no. 3, pp. 644–652, 2009, doi: 10.1016/j.compedu.2009.04.002.

60. S. M. R. Abidi, M. Hussain, Y. Xu, and W. Zhang, "Prediction of confusion attempting algebra homework in an intelligent tutoring system through machine learning techniques for educational sustainable development," Sustain., vol. 11, no. 1, 2018, doi: 10.3390/su11010105.

61. J. Smith, P. Legg, M. Matovic, and K. Kinsey, "Predicting user confidence during visual decision making," ACM Trans. Interact. Intell. Syst., vol. 8, no. 2, 2018, doi: 10.1145/3185524.

62. I. Pandarova, T. Schmidt, J. Hartig, A. Boubekki, R. D. Jones, and U. Brefeld, "Predicting the Difficulty of Exercise Items for Dynamic Difficulty Adaptation in Adaptive Language Tutoring," Int. J. Artif. Intell. Educ., vol. 29, no. 3, pp. 342–367, 2019, doi: 10.1007/s40593-019-00180-4.

63. S. Nam, G. Frishkoff, and K. Collins-Thompson, "Predicting Students' Disengaged Behaviors in an Online Meaning-Generation Task," IEEE Trans. Learn. Technol., vol. 11, no. 3, pp. 362–375, 2018, doi: 10.1109/TLT.2017.2720738.

64. D. Ríofrío-Luczando, J. Ramírez, and M. Berrocal-Lobo, "Predicting student actions in a procedural training environment," IEEE Trans. Learn. Technol., vol. 10, no. 4, pp. 463–474, 2017, doi: 10.1109/TLT.2017.2658569.

65. M. Haridas, G. Gutjahr, R. Ramam, R. Ramaraju, and P. Nedungadi, "Predicting school performance and early risk of failure from an intelligent tutoring system," Educ. Inf. Technol., 2020, doi: 10.1007/s10639-020-10144-0.

66. S. K. D’Mello, S. D. Craig, J. Sullins, and A. C. Graesser, "Predicting affective states expressed through an emote-aloud procedure from autotutor’s mixed-initiative dialogue," Int. J. Artif. Intell. Educ., vol. 16, no. 1, pp. 3–28, 2006.

67. G. M. Baschera and M. Gross, "Poisson-based inference for perturbation models in adaptive spelling training," Int. J. Artif. Intell. Educ., vol. 20, no. 4, pp. 333–360, 2010, doi: 10.3233/JAI-2010-011.

68. G. Paravati, F. Lamberti, V. Gatteschi, C. Demartini, and P. Montuschi, "Point cloud-based automatic assessment of 3D computer animation coursework," IEEE Trans. Learn. Technol., vol. 10, no. 4, pp. 532–543, 2017, doi: 10.1109/TLT.2016.2638811.

69. V. G. Kaburlasos, C. C. Marinagi, and V. T. Tsoukalas, "Personalized multi-student improvement based on Bayesian cybernetics," Comput. Educ., vol. 51, no. 4, pp. 1430–1449, 2008, doi: 10.1016/j.compedu.2008.01.004.

70. R. Janning, C. Schatten, and L. Schmidt-Thieme, "Perceived Task-Difficulty Recognition from Log-file Information for the Use in Adaptive Intelligent Tutoring Systems," Int. J. Artif. Intell. Educ., vol. 26, no. 3, pp. 855–876, 2016, doi: 10.1007/s40593-016-0097-9.

71. R. A. Sottilare and M. Proctor, "Passively classifying student mood and performance within intelligent tutors," Educ. Technol. Soc., vol. 15, no. 2, pp. 101–114, 2012.

72. Z. Ul-Qayyum and W. Altarf, "Paraphrase identification using semantic heuristic features," Res. J. Appl. Sci. Eng. Technol., vol. 4, no. 22, pp. 4894–4904, 2012.

73. S. Ganzfried and F. Yusuf, "Optimal weighting for exam composition," Educ. Sci., vol. 8, no. 1, pp. 1–14, 2018, doi: 10.3390/educsci8010036.

74. C. Y. Ting and S. Phon-Amnuaisuk, "Optimal dynamic decision network model for scientific inquiry learning environment," Appl. Intel., vol. 33, no. 3, pp. 387–406, 2010, doi: 10.1007/s10489-009-0174-y.

75. J. D. Gobert, R. S. Baker, and M. B. Wixon, "Operationalizing and Detecting Disengagement Within Online Science Microworlds," Educ. Psychol., vol. 50, no. 1, pp. 43–57, 2015, doi: 10.1080/00461520.2014.999919.

76. A. Cully and Y. Demiris, "Online Knowledge Level Tracking with Data-Driven Student Models and Collaborative Filtering," IEEE Trans. Knowl. Data Eng., vol. 32, no. 10, pp. 2000–2012, 2020, doi: 10.1109/TKDE.2019.2912367.

77. C. Cunha-Perey, M. Arevalillo-Herraez, L. Marco-Gimenez, and D. Arnau, "On Incorporating Affective Support to an Intelligent Tutoring System: An Empirical Study," Rev. Iberoam. Tecnol. del Aprendiz., vol. 13, no. 2, pp. 63–69, 2018, doi: 10.1109/RITA.2018.2831760.

78. N. Sharada, M. Shashi, and X. Xiong, "Modeling student knowledge retention using deep learning and random forests," J. Eng. Appl. Sci., vol. 13, pp. 1347–1353, Jan. 2018, doi: 10.3923/jeas.2018.1347.1353.

79. J. R. Anderson, S. Betts, J. L. Ferris, and J. M. Fincham, "Neural imaging to track mental states while using an intelligent tutoring system," Proc. Natl. Acad. Sci. U. S. A., vol. 107, no. 15, pp. 7018–7023, 2010, doi: 10.1073/pnas.1000942107.

80. M. Holmes, A. Latham, K. Crockett, and J. D. O'Shea, "Near Real-Time Comprehension Classification with Artificial Neural Networks: Decoding e-Learner Non-Verbal Behavior," IEEE Trans. Learn. Technol., vol. 11, no. 1, pp. 5–12, 2018, doi: 10.1109/TLT.2017.2754497.

81. D. S. McNamara, S. A. Crossley, and R. Roscoe, "Natural language processing in an intelligent writing strategy tutoring system," Behav. Res. Methods, vol. 45, no. 2, pp. 499–515, 2013, doi: 10.3758/s13428-012-0258-1.
82. S. K. D’Mello and A. Graesser, “Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features,” *User Model. User-adapt. Interact.*, vol. 20, no. 2, pp. 147–187, 2010, doi: 10.1007/s11257-010-9074-4.

83. K. W. Brawner and A. J. Gonzalez, “Modelling a learner’s affective state in real time to improve intelligent tutoring effectiveness,” *Theor. Issues Ergon. Sci.*, vol. 17, no. 2, pp. 183–210, 2016, doi: 10.1080/1463922X.2015.1111463.

84. S. W. McQuiggan, B. W. Mott, and J. C. Lester, “Modeling self-efficacy in intelligent tutoring systems: An inductive approach,” *User Model. User-adapt. Interact.*, vol. 18, no. 1–2, pp. 81–123, 2008, doi: 10.1007/s11257-007-9040-y.

85. C. Y. Ting, Y. C. Sam, and C. O. Wong, “Model of conceptual change for INQPRO: A Bayesian Network approach,” *Comput. Educ.*, vol. 65, pp. 77–91, 2013, doi: 10.1016/j.cpe.2013.01.013.

86. F. Hirt, E. Werlen, I. Moser, and P. Bergamin, “Measuring emotions during learning: Lack of coherence between automated facial emotion recognition and emotional experience,” *Open Comput. Sci.*, vol. 9, no. 1, pp. 308–317, 2019, doi: 10.1515/comp-2019-0020.

87. R. Liu et al., “Learning linkages: Integrating data streams of multiple modalities and timescales,” *J. Comput. Assist. Learn.*, vol. 35, no. 1, pp. 99–109, 2019, doi: 10.1111/jcal.12315.

88. M. Beggiato, M. Pereira, T. Petzoldt, and J. Krems, “Learning and development of trust, acceptance and the mental model of ACC. A longitudinal on-road study,” *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 35, pp. 75–84, 2015, doi: 10.1016/j.trf.2015.10.005.

89. S. K. D’Mello and A. Graesser, “Language and discourse are powerful signals of student emotions during tutoring,” *IEEE Trans. Learn. Technol.*, vol. 5, no. 4, pp. 304–317, 2012, doi: 10.1109/TLT.2012.10.

90. D. F. Lian and Q. Liu, “Jointly Recommending Library Books and Predicting Academic Performance: A Mutual Reinforcement Perspective,” *J. Comput. Sci. Technol.*, vol. 33, no. 4, pp. 654–667, 2018, doi: 10.1007/s11390-018-1847-y.

91. K. Wauters, P. Desmet, and W. Van Den Noortgate, “Item difficulty estimation: An auspicious collaboration between data and judgment,” *Comput. Educ.*, vol. 58, no. 4, pp. 1183–1193, 2012, doi: 10.1016/j.compedu.2011.10.020.

92. E. Ayers and B. Junker, “IRT modeling of tutor performance to predict end-of-year exam scores,” *Educ. Psychol. Meas.*, vol. 68, no. 6, pp. 972–987, 2008, doi: 10.1177/00164988083218758.

93. S. Alexander and A. Sarrafzadeh, “Interfaces That Adapt like Humans How Do Human Tutors Interpret Non-verbal Behaviour?”, pp. 641–645, 2004.

94. S. K. N. A. Rahim, A. H. Jaafar, S. Masrom, F. Zulkipli, S. R. Ismail, and N. Ahmad, “Intelligent tutoring systems’ measurement and prediction of students’ performance using predictive function,” *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 1 Special Issue, pp. 187–192, 2020, doi: 10.30534/ijeter/2020/2981.12020.

95. E. Mousavinasaie, S. R. Niakan Kalhori, M. Rajkhashan, L. Keikha, and M. Ghazi Saeedi, “Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods,” *Interact. Learn. Environ.*, vol. 29, no. 1, pp. 142–163, 2021, doi: 10.1080/10494820.2018.1558257.

96. A. Karaci, “Intelligent tutoring system model based on fuzzy logic and constraint-based student model,” *Neural Comput. Appl.*, vol. 31, no. 8, pp. 3619–3628, 2019, doi: 10.1007/s00521-017-3311-2.

97. A. H. Basori, A. Tenriawaru, and A. B. F. Mansur, “Intelligent Avatar on E-Learning using Facial Expression and Haptic,” *TELEKOMNIKA (Telecommunication Comput. Electron. Control)*, vol. 9, no. 1, p. 115, 2011, doi: 10.12928/telemkwa.v9i1.676.

98. J. Zhu, Y. Zang, H. Qiu, and T. Zhou, “Integrating Temporal Information into Knowledge Tracing: A Temporal Difference Approach,” *IEEE Access*, vol. 6, pp. 27302–27312, 2018, doi: 10.1109/ACCESS.2018.2833874.

99. S. Chaffar and C. Frasson, “Inducing optimal emotional state for learning in intelligent tutoring systems,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3220, pp. 45–54, 2004, doi: 10.1007/978-3-540-30139-4_5.

100. J. A. Ruiz-pérez-Valiente, P. J. Muñoz-Merino, and C. Delgado Kloos, “Improving the prediction of learning outcomes in educational platforms including higher level interaction indicators,” *Expert Syst.*, vol. 35, no. 6, pp. 1–11, 2018, doi: 10.1111/exsy.12298.

101. T. S. Ashwin and R. M. R. Guddeti, “Impact of inquiry interventions on students in e-learning and classroom environments using affective computing framework,” *User Model. User-adapt. Interact.*, vol. 30, no. 5, pp. 759–801, 2020, doi: 10.1007/s11257-019-09254-3.

102. S. Manning and A. Dix, “Identifying students’ mathematical skills from a multiple-choice diagnostic test using an iterative technique to minimise false positives,” *Comput. Educ.*, vol. 51, no. 3, pp. 1154–1171, 2008, doi: 10.1016/j.compedu.2007.10.010.

103. V. Rus, P. M. McCarthy, A. C. Graesser, and D. S. McNamara, “Identification of Sentence-to-Sentence Relations Using a Textual Entailer,” *Res. Lang. Comput.*, vol. 7, no. 2, pp. 209–229, 2009, doi: 10.1007/s11168-009-9065-y.

104. G. Padrón-Rivera, G. Rebolloledo-Mendez, P. P. Parra, and N. S. Huerta-Pacheco, “Identification of action units related to affective states in a tutoring system for mathematics,” *Educ. Technol. Soc.*, vol. 19, no. 2, pp. 77–86, 2016.

105. R. Zatarain Cabada, H. Rodríguez Rangel, M. L. Barron Estrada, and H. M. Cardenas Lopez, “Hyperparameter optimization in CNN for learning-centered emotion recognition for intelligent tutoring systems,” *Soft Comput.*, vol. 24, no. 10, pp. 7593–7602, 2020, doi: 10.1007/s00500-019-04387-4.

106. Y. Zhang, C. Lin, and M. Chi, “Going deeper: Automatic short-answer grading by combining student and question models,” *User Model. User-adapt. Interact.*, vol. 30, no. 1, pp. 51–80, 2020, doi: 10.1007/s11257-019-09251-6.

107. J. D. Gobert, M. Sao Pedro, J. Raziuddin, and R. S. Baker, “From Log Files to Assessment Metrics: Measuring Students’ Science Inquiry Skills Using Educational Data Mining,” *J. Learn. Sci.*, vol. 22, no. 4, pp. 521–563, 2013, doi: 10.1080/10508406.2013.837391.

108. I. Zipiri, J. A. Elorriaga, A. Arruarte, and A. D. De Ilarraza, “From human to automatic summary evaluation,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3220, pp. 432–442, 2004, doi: 10.1007/978-3-540-30139-4_41.
109. K. Muramatsu, E. Tanaka, K. Watanuki, and T. Matsui, “Framework to describe constructs of academic emotions using ontological descriptions of statistical models,” Res. Pract. Technol. Enhanc. Lear., vol. 11, no. 1, pp. 1–18, 2016, doi: 10.1186/s41039-016-0029-1.

110. A. Koduru, H. B. Valiveti, and A. K. Budati, “Feature extraction algorithms to improve the speech emotion recognition rate,” Int. J. Speech Technol., vol. 23, no. 1, pp. 45–55, 2020, doi: 10.1007/s10772-020-09672-4.

111. C. Y. Ting and S. Phon-Amnuaisuk, “Factors influencing the performance of Dynamic Decision Network for INQPRO,” Comput. Educ., vol. 52, no. 4, pp. 762–780, 2009, doi: 10.1016/j.compedu.2008.12.003.

112. J. M. Harley et al., “Examining the predictive relationship between personality and emotion traits and students’ agent-directed emotions: towards emotionally-adaptive agent-based learning environments,” User Model. User-adapt. Interact., vol. 26, no. 2–3, pp. 177–219, 2016, doi: 10.1007/s11257-016-9169-7.

113. M. Liu and Q. Zhang, “Evaluation of student performance with predicted learning curve based on grey models for personalized tutoring,” Int. J. Emerg. Technol. Learn., vol. 14, no. 13, pp. 157–171, 2019, doi: 10.3991/ijet.v14i13.9880.

114. E. Brajković and D. V. T. Volarić, “Evaluation of methods for sentence similarity for use in intelligent tutoring system,” Adv. Sci. Technol. Eng. Syst., vol. 3, no. 5, pp. 1–5, 2018, doi: 10.25046/aj030501.

115. A. Asselman, M. Khalidi, and S. Aamour, “Evaluating the impact of prior required scaffolding items on the improvement of student performance prediction,” Educ. Inf. Technol., vol. 25, no. 4, pp. 3227–3249, 2020, doi: 10.1007/s10639-019-10077-3.

116. K. A. Assielou, C. T. Haba, B. T. Gooré, T. L. Kadjo, and K. D. Yao, “Emotional impact for predicting student performance in intelligent tutoring systems (ITS),” Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 7, pp. 219–225, 2020, doi: 10.14569/IJACSA.2020.0110728.

117. A. Koşir and G. Strle, “Emotion elicitation in a socially intelligent service: The typing tutor,” Computers, vol. 6, no. 2, pp. 1–19, 2017, doi: 10.3390/computers6020014.

118. S. D. Craig, S. D’Mello, A. Witherspoon, and A. Graesser, “Emote aloud during learning with AutoTutor: Applying the Facial Action Coding System to cognitive-Affective states during learning,” Cogn. Emot., vol. 22, no. 5, pp. 777–788, 2008, doi: 10.1080/02699930701516759.

119. S. Asian, N. Alyuz, E. Okur, S. E. Mete, E. Oktay, and A. A. Esme, “Effect of emotion-aware interventions on students’ behavioral and emotional states,” Educ. Technol. Res. Dev., vol. 66, no. 6, pp. 1399–1413, 2018, doi: 10.1007/s11423-018-9589-7.

120. S. M. R. Abidi et al., “Educational sustainability through big data assimilation to quantify academic procrastination using ensemble classifiers,” Sustain., vol. 12, no. 15, pp. 1–23, 2020, doi: 10.3390/su12156074.

121. X. Sun, X. Zhao, B. Li, Y. Ma, R. Sutcliffe, and J. Feng, “Dynamic Key-Value Memory Networks With Rich Features for Knowledge Tracing,” IEEE Trans. Cybern., pp. 1–7, 2021, doi: 10.1109/TCYB.2021.3051028.

122. M. Eryilmaz and A. Adabashi, “Development of an intelligent tutoring system using bayesian networks and fuzzy logic for a higher student academic performance,” Appl. Sci., vol. 10, no. 20, 2020, doi: 10.3390/app10196638.

123. K. Kularbphettong, P. Kedsiribut, and P. Roonrakwit, “Developing an Adaptive Web-based Intelligent Tutoring System Using Mastery Learning Technique,” Procedia - Soc. Behav. Sci., vol. 191, pp. 686–691, 2015, doi: 10.1016/j.sbspro.2015.04.619.

124. O. Alzoubi, S. K. D’Mello, and R. A. Calvo, “Detecting naturalistic expressions of nonbasic affect using physiological signals,” IEEE Trans. Affect. Comput., vol. 3, no. 3, pp. 298–310, 2012, doi: 10.1109/T-AFFC.2012.24.

125. A. Hussain, A. R. Abbasi, and N. Afzulpurkar, “Detecting & interpreting self-manipulating hand movements for student’s affect prediction,” Human-centric Comput. Inf. Sci., vol. 2, no. 1, pp. 1–18, 2012, doi: 10.1186/2192-1962-2-14.

126. W. Chen, “Designing agents that recognise and respond to players’ emotions,” Smart Innov. Syst. Technol., vol. 14, pp. 191–200, 2012, doi: 10.1007/978-3-642-29934-6_19.

127. S. Ismail and S. Abdulla, “Design and implementation of an intelligent system to predict the student graduation AGPA,” Aust. Educ. Comput., vol. 30, no. 2, 2015.

128. C. Perrotta and N. Selwyn, “Deep learning goes to school: toward a relational understanding of AI in education,” Learn. Media Technol., vol. 45, no. 3, pp. 251–269, 2020, doi: 10.1080/17439884.2020.1686017.

129. K. R. Koedinger, S. D.Mello, E. A. Mclaughlin, Z. A. Pardos, and C. P. Rosé, “Data mining and education,” Wiley Interdiscip. Rev. Cogn. Sci., vol. 6, no. 4, pp. 333–353, 2015, doi: 10.1002/wcs.1350.

130. M. Pratama, E. Lughoffer, M. J. Er, S. Anavatti, and C. P. Lim, “Data driven modelling based on Recurrent Interval-Valued Metacognitive Scaffolding Fuzzy Neural Network,” Neurocomputing, vol. 262, no. 4, pp. 2017–2027, 2017, doi: 10.1016/j.neucom.2016.10.093.

131. L. Meng, M. Zhang, W. Zhang, and Y. Chu, “CS-BKT: introducing item relationship to the Bayesian knowledge tracing model,” Interact. Learn. Environ., vol. 0, no. 0, pp. 1–11, 2019, doi: 10.1080/10494820.2019.1629600.

132. A. R. Noriega, R. J. Ramirez, J. J. Tapia, V. H. Castillo, and S. Jimenez, “Construction of conditional probability tables of Bayesian networks using ontologies and Wikipedia,” Comput. y Sist., vol. 23, no. 4, pp. 1275–1289, 2019, doi: 10.13053/CyS-23-4-2705.

133. J. Melesko and V. Novickij, “Computer adaptive testing using upper-confidence bound algorithm for formative assessment,” Appl. Sci., vol. 9, no. 20, 2019, doi: 10.3390/app9204303.

134. F. Alqahtani and N. Ramzan, “Comparison and efficacy of synergistic intelligent tutoring systems with human physiological response,” Sensors (Switzerland), vol. 19, no. 3, 2019, doi: 10.3390/s19030460.

135. M. J. Ventura et al., “Combining computational models of short essay grading for conceptual physics problems,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 3220, pp. 423–431, 2004, doi: 10.1007/978-3-540-30139-4_40.
136. T. Zhang, M. Hasegawa-Johnson, and S. E. Levinson, “Cognitive state classification in a spoken tutorial dialogue system,” *Speech Commun.*, vol. 48, no. 6, pp. 616–632, 2006, doi: 10.1016/j.specom.2005.09.006.

137. X. Tao, Q. Niu, M. Jackson, and M. Ratcliffe, “Classification of video lecture learners’ cognitive and negative emotional states using a Bayesian belief network,” *Filomat*, vol. 32, no. 5, pp. 1823–1829, 2018, doi: 10.2298/FIL1805823T.

138. M. O. Z. San Pedro, R. S. J. D. Baker, and M. M. T. Rodrigo, “Carelessness and affect in an intelligent tutoring system for mathematics,” *Int. J. Artif. Intell. Educ.*, vol. 24, no. 2, pp. 189–210, 2014, doi: 10.1080/09695229.2009.1001573.

139. P. E. Antoniou et al., “Biosensor real-time affective analytics in virtual and mixed reality medical education serious games: Cohort study,” *JMIR Serious Games*, vol. 8, no. 3, pp. 1–16, 2020, doi: 10.2196/17823.

140. A. Yazidi, A. AbolpourMofrad, M. Goodwin, H. L. Hammer, and E. Amrten, “Balanced difficulty task finder: an adaptive recommendation method for learning tasks based on the concept of state of flow,” *Cogn. Neurodyn.*, vol. 14, no. 5, pp. 675–687, 2020, doi: 10.1007/s11571-020-09624-3.

141. Y. He, S. C. Hui, and T. T. Quan, “Automatic summary assessment for intelligent tutoring systems,” *Comput. Educ.*, vol. 53, no. 3, pp. 890–899, 2009, doi: 10.1016/j.compedu.2009.05.008.

142. A. Abdi, N. Idris, R. M. Alguieiev, and R. M. Aliguiyev, “Automatic summarization assessment through a combination of semantic and syntactic information for intelligent educational systems,” *Inf. Process. Manag.*, vol. 51, no. 4, pp. 340–358, 2015, doi: 10.1016/j.ipm.2015.02.001.

143. A. Kapoor, W. Burleson, and R. W. Picard, “Automatic prediction of frustration,” *Int. J. Hum. Comput. Stud.*, vol. 65, no. 8, pp. 724–736, 2007, doi: 10.1016/j.ijhcs.2007.02.003.

144. M. Lintean, V. Rus, and R. Azevedo, “Automatic detection of student mental models based on natural language student input during metacognitive skill training,” *J. Artif. Intell. Educ.*, vol. 21, no. 3, pp. 169–190, 2012, doi: 10.3233/JAI-2012-222.

145. S. Cetintas, L. Si, Y. P. Xin, and C. Hord, “Automatic detection of off-task behaviors in intelligent tutoring systems with machine learning techniques,” *IEEE Trans. Learn. Technol.*, vol. 3, no. 3, pp. 228–236, 2010, doi: 10.1109/TLT.2009.44.

146. S. D’Mello and A. Graesser, “Automatic detection of learner’s affect from gross body language,” *Appl. Artif. Intell.*, vol. 23, no. 2, pp. 123–150, 2009, doi: 10.1080/088395109032631745.

147. S. K. D’Mello, S. D. Craig, A. Witherspoon, B. McDaniel, and A. Graesser, “Automatic detection of learner’s affect from conversational cues,” *User Model. Useradapt. Interact.*, vol. 18, no. 1–2, pp. 45–80, 2008, doi: 10.1007/s11257-007-9037-6.

148. M. H. Myers, “Automatic detection of a student’s affective states for intelligent teaching systems,” *Brain Sci.*, vol. 11, no. 3, pp. 1–15, 2021, doi: 10.3390/brainsci11030331.

149. S. Hutt et al., “Automated gaze-based mind wandering detection during computerized learning in classrooms,” vol. 29, no. 4. Springer Netherlands, 2019.

150. W. Westera, M. Dascalu, H. Kurvers, S. Ruseti, and S. Trausan-Matu, “Automated essay scoring in applied games: Reducing the teacher bandwidth problem in online training,” *Comput. Educ.*, vol. 123, pp. 212–224, 2018, doi: 10.1016/j.compedu.2018.05.010.

151. A. Behera, P. Matthew, A. Keidel, P. Vangorp, H. Fang, and S. Canning, “Associating Facial Expressions and Upper-Body Gestures with Learning Tasks for Enhancing Intelligent Tutoring Systems,” *Int. J. Artif. Intell. Educ.*, vol. 30, no. 2, pp. 236–270, 2020, doi: 10.1007/s40593-020-00195-2.

152. P. J. Muñoz-Merino, R. González Novillo, and C. Delgado Kloos, “Assessment of skills and adaptive learning for parametric exercises combining knowledge spaces and item response theory,” *Appl. Soft Comput. J.*, vol. 68, pp. 110–124, 2018, doi: 10.1016/j.asoc.2018.03.045.

153. L. N. Brown and A. M. Howard, “Assessment of engagement for intelligent educational agents: A pilot study with middle school students,” *Comput. Educ.*, vol. 5, no. 4, pp. 96–106, 2014.

154. A. Mousavi, M. Schmidt, V. Squires, and K. Wilson, “Assessing the Effectiveness of Student Advice Recommender Agent (SARA): the Case of Automated Personalized Feedback,” *Int. J. Artif. Intell. Educ.*, 2020, doi: 10.1007/s40593-020-00210-6.

155. S. Mccarthy and D. S. McNamara, “and Hierarchical Machine Learning Approaches to Text Difficulty Classification,” 2020.

156. J. Makhlof and T. Mine, “Analysis of click-stream data to predict STEM careers from student usage of an intelligent tutoring system,” *J. Educ. Data Min.*, vol. 12, no. 2, pp. 1–18, 2020, doi: 10.5281/zenodo.4008050.

157. N. Tomasevic, N. Gvozdenovic, and S. Mrones, “An overview and comparison of supervised data mining techniques for student exam performance prediction,” *Comput. Educ.*, vol. 143, 2020, doi: 10.1016/j.compedu.2019.103676.

158. X. M. Hu, J. Zhang, H. S. H. Chung, O. Liu, and J. Xiao, “An intelligent testing system embedded with an ant-colony-optimization-based test composition method,” *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 39, no. 6, pp. 659–669, 2009, doi: 10.1109/TSMCC.2009.2021952.

159. S. Yassine et al., “An intelligent diagnostic framework: A scaffolding tool to resolve academic reading problems of Thai first-year university students,” *Comput. Educ.*, vol. 39, no. 1, pp. 3–13, 2013, doi: 10.1016/j.compedu.2014.10.003.

160. R. Conejo, E. Guzmán, J. L. Perez-De-La-Cruz, and B. Barros, “An empirical study on the quantitative notion of task difficulty,” *Expert Syst. Appl.*, vol. 41, no. 2, pp. 594–606, 2014, doi: 10.1016/j.eswa.2013.07.084.

161. K. Muñoz, P. M. Kevitt, T. Lunney, J. Noguez, and L. Neri, “An emotional student model for game-play adaptation,” *Entertain. Comput.*, vol. 2, no. 2, pp. 133–141, 2011, doi: 10.1016/j.entcom.2010.12.006.

162. A. Sarrafzadeh, S. T. V. Alexander, and J. Shanbehzadeh, “Affective Tutoring System for Better Learning,” *Int. J. Mob. Blended Learn.*, vol. 1, no. 1, pp. 61–77, 2009, doi: 10.4018/jmbl.2009101015.

163. T. S. Ashwin and R. M. R. Guddeti, “Affective database for e-learning and classroom environments using Indian students’ faces, hand gestures and body postures,” *Futur. Gener. Comput. Syst.*, vol. 108, pp. 334–348, 2020, doi: 10.1016/j.future.2020.02.075.
164. J. Hoey, T. Schröder, and A. Alhothali, "Affect control processes: intelligent affective interaction using a partially observable Markov decision process," *Artif. Intell.*, vol. 230, pp. 134–172, 2016, doi: 10.1016/j.artint.2015.09.004.

165. M. Feng, N. Heffeman, and K. Koedinger, "Addressing the assessment challenge with an online system that tutors as it assesses," *User Model. User-adapt. Interact.*, vol. 19, no. 3, pp. 243–266, 2009, doi: 10.1007/s11257-009-9063-7.

166. M. Arevalillo-Herréz, L. Marco-Giménez, D. Arnau, and J. A. González-Calero, "Adding sensor-free intention-based affective support to an Intelligent Tutoring System," *Knowledge-Based Syst.*, vol. 132, pp. 85–93, 2017, doi: 10.1016/j.knosys.2017.06.024.

167. B. Paassen, B. Mokbel, and B. Hammer, "Adaptive structure metrics for automated feedback provision in intelligent tutoring systems," *Neurocomputing*, vol. 192, pp. 3–13, 2016, doi: 10.1016/j.neucom.2015.12.108.

168. K. Wauters, P. Desmet, and W. Van Den Noortgate, "Adaptive item-based learning environments based on the item response theory: Possibilities and challenges," *J. Comput. Assist. Learn.*, vol. 26, no. 6, pp. 549–562, 2010, doi: 10.1111/j.1365-2729.2010.00368.x.

169. J. A. Okpo, J. Masthoff, M. Dennis, and N. Beacham, "Adapting exercise selection to performance, effort and self-esteem," *New Rev. Hypermedia Multimed.*, vol. 24, no. 3, pp. 193–227, 2018, doi: 10.1080/13614568.2018.1477999.

170. L. de-La-Fuente-Valentín, A. Pardo, F. L. Hernández, and D. Burgos, "A visual analytics method for score estimation in learning courses," *J. Univers. Comput. Sci.*, vol. 21, no. 1, pp. 134–155, 2015, doi: 10.3217/jucs-021-01-0134.

171. N. Nagashree and N. V. Pujari, "A Tutor Assisting Novel Electronic Framework for Qualitative Analysis of a Question Bank," *Comput. Human Behav.*, vol. 65, pp. 9–13, 2016, doi: 10.1016/j.chb.2016.08.008.

172. L. Chen, N. Tokuda, and P. Hou, "A table look-up parser in online ILTS applications," *Comput. Assist. Lang. Learn.*, vol. 18, no. 1–2, pp. 49–62, 2005, doi: 10.1080/09588220500132357.

173. A. A. Soofi and M. U. Ahmed, "A systematic study of domains, techniques, delivery modes and validation methods for Intelligent Tutoring Systems," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 4, pp. 289–316, 2012, doi: 10.1080/09588220500132357.

174. A. Abdi et al., "A new approach to personalization: Integrating e-learning and m-learning," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 26, no. 4, pp. 351–352, 2017, doi: 10.1007/s11423-012-9250-9.

175. A. Smith, S. Leeman-Munk, A. Shelton, B. Mott, E. Wiebe, and J. Lester, "A Multimodal Assessment Framework for Integrating Student Writing and Drawing in Elementary Science Learning," *IEEE Trans. Learn. Technol.*, vol. 12, no. 1, pp. 3–15, 2019, doi: 10.1109/TLT.2018.2799871.

176. J. M. Harley, F. Bouchet, M. S. Hussain, R. Azevedo, and R. Calvo, "A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system," *Comput. Human Behav.*, vol. 48, pp. 615–625, 2015, doi: 10.1016/j.chb.2015.02.013.

177. K. A. Smith, M. Dennis, J. Masthoff, and N. Tintarev, "A methodology for creating and validating psychological stories for conveying and measuring psychological traits," *User Model. User-adapt. Interact.*, vol. 29, no. 3, pp. 573–618, 2019, doi: 10.1007/s11257-019-09219-6.

178. L. Subirats, L. Pérez, C. Hernández, S. Fort, and G. M. Sacha, "A granularity-based intelligent tutoring system for zooarchaeology," *Appl. Sci.*, vol. 9, no. 22, pp. 1–17, 2019, doi: 10.3390/app9224960.

179. E. Verdú, M. J. Verdú, L. M. Reguera, J. P. De Castro, and R. García, "A genetic fuzzy expert system for automatic question classification in a competitive learning environment," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7471–7478, 2012, doi: 10.1016/j.eswa.2012.01.115.

180. T. Tenório, L. I. Bittencourt, S. Isotani, A. Pedro, and P. Ospina, "A gamified peer assessment model for on-line learning environments in a competitive context," *Comput. Human Behav.*, vol. 64, pp. 247–263, 2016, doi: 10.1016/j.chb.2016.06.049.

181. M. Badaracco and L. Martínez, "A fuzzy linguistic algorithm for adaptive test in Intelligent Tutoring System based on competences," *Expert Syst. Appl.*, vol. 40, no. 8, pp. 3073–3086, 2013, doi: 10.1016/j.eswa.2012.12.023.

182. V. Tsiriga and M. Virvou, "A framework for the initialization of student models in web-based intelligent tutoring systems," *User Model. User-Adapt. Interact.*, vol. 14, no. 4, pp. 289–316, 2004, doi: 10.1023/B:USER.0000043396.14788.cc.

183. P. A. Jaques and R. M. Vicari, "A BDI approach to infer student’s emotions in an intelligent learning environment," *Comput. Educ.*, vol. 49, no. 2, pp. 360–384, 2007, doi: 10.1016/j.compedu.2005.09.002.

184. H. J. Vos, "A Bayesian sequential procedure for determining the optimal number of interrogatory examples for concept-learning," *Comput. Human Behav.*, vol. 23, no. 1, pp. 609–627, 2007, doi: 10.1016/j.chb.2004.11.002.

185. M. Krechovsky, B. Mardell, M. Rivard, and D. Wilson, *Visible learners: Promoting Reggio-inspired approaches in all schools*. John Wiley & Sons, 2013.
Figure 1
Evaluating and grading process. Adapted from [6]

Figure 2
Intelligent Tutoring System
Basic components. Adapted from [10] and [11].

Figure 3
Funnel system.

Figure 4
Production per year
Figure 5

Most relevant sources.

Figure 6

Total citation of the main sources
Figure 7

Main Authors

Figure 8

Percentage of participation based on Keywords Plus.
Figure 9

Co-occurrence of words in keyword Plus from all sources

Figure 10

Co-occurrence of words in keyword Plus from all sources.
Figure 11

Co-occurrence of words in keyword Plus from all sources

Figure 12

Metanode preprocessing
Figure 13

Topic Extractor

Figure 14

Reducing dimensionality, assigning Colors
Figure 15
Metanode interactive visualization.

Figure 16
Metanode Tagging
Figure 17

Inspect Topics and Terms.
Figure 18

Inspect Topics and Terms exploration
Figure 19

Inspect Topics & Terms, highlighting topic 3, related to the assessment

| Document                                                                 | Assigned topic | Weight |
|--------------------------------------------------------------------------|----------------|--------|
| "adapt exercise selection performance effort self-esteeem"              | topic_3        | 171.00 |
| "adapt exercise selection performance effort self-esteeem"              | topic_3        | 183.00 |
| "adapt exercise selection performance effort self-esteeem"              | topic_3        | 210.00 |
| "adapt exercise selection performance effort self-esteeem"              | topic_3        | 241.00 |
| "adapt exercise selection performance effort self-esteeem"              | topic_3        | 221.00 |
| "adapt exercise selection performance effort self-esteeem"              | topic_3        | 249.00 |
| "adapt exercise selection performance effort self-esteeem"              | topic_3        | 322.00 |
| "adapt exercise selection performance effort self-esteeem"              | topic_3        | 369.00 |
| "adapt exercise selection performance effort self-esteeem"              | topic_3        | 644.00 |

Figure 20

Inspect Topics and Terms as a word cloud
Figure 21

Number of publications per year

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- BibiometrixExportFile20210428.xlsx