

Received July 5, 2021, accepted July 24, 2021, date of publication July 28, 2021, date of current version August 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3100890

A Systematic Review of the Effects of Automatic Scoring and Automatic Feedback in Educational Settings

MARCELO GUERRA HAHN¹, (Senior Member, IEEE),
SILVIA MARGARITA BALDIRIS NAVARRO^{1,2}, LUIS DE LA FUENTE VALENTÍN¹,
AND DANIEL BURGOS^{1,2}, (Senior Member, IEEE)

¹Escuela Superior de Ingeniería y Tecnología (ESIT), Universidad Internacional de La Rioja (UNIR), Logroño, 26006 La Rioja, Spain

²Research Institute for Innovation & Technology in Education (UNIR iTED), Universidad Internacional de La Rioja (UNIR), Logroño, 26006 La Rioja, Spain

Corresponding author: Marcelo Guerra Hahn (marcelo.guerrahahn261@comunidadunir.net)

This work was supported in part by the PLeNTaS Project “Proyectos I+D+i 2019” under Grant PID2019-111430RB-I00, and in part by the PL-NETO Project “Proyecto PROPIO UNIR” under Grant B0036.

ABSTRACT Automatic scoring and feedback tools have become critical components of online learning proliferation. These tools range from multiple-choice questions to grading essays using machine learning (ML). Learning environments such as massive open online courses (MOOCs) would not be possible without them. The usage of this mechanism has brought many exciting areas of study, from the design of questions to the ML grading tools’ precision and accuracy. This paper analyzes the findings of 125 studies published in journals and proceedings between 2016 and 2020 on the usages of automatic scoring and feedback as a learning tool. This analysis gives an overview of the trends, challenges, and open questions in this research area. The results indicate that automatic scoring and feedback have many advantages. The most important benefits include enabling scaling the number of students without adding a proportional number of instructors, improving the student experience by reducing the time between submission grading and feedback, and removing bias in scoring. On the other hand, these technologies have some drawbacks. The main problem is creating a disincentive to develop innovative answers that do not match the expected one or have not been considered when preparing the problem. Another drawback is potentially training the student to answer the question instead of learning the concepts. With this, given the existence of a correct answer, such an answer could be leaked to the internet, making it easier for students to avoid solving the problem. Overall, each of these drawbacks presents an opportunity to look at ways to improve technologies to use these tools to provide a better learning experience to students.

INDEX TERMS Education, feedback, inclusive learning, literature reviews, machine learning.

I. INTRODUCTION

Automatic scoring and feedback consist of calculating grades on students’ work and providing personalized feedback using technological tools that do not require human participation [1]. These tools play a significant role in online learning. Like massive open online courses (MOOCs), many new learning environments would not be possible without them [2]. Automatic scoring has been a tool for a while, and multiple-choice tests have been available for a long time. Large-scale multiple-choice tests have been possible since the

introduction of the Scantron. This tool continues to be used today [3].

With the rapid growth of technology and internet access, the use of automatic scoring and feedback has accelerated [4], [5]. The benefits for institutions and instructors of the usage of these tools are apparent. The institutions and instructors acquire the ability to increase students per instructor and provide fast and consistent results [6]. However, all these advantages come with potential drawbacks.

Multiple areas of study can use automatic scoring. An early implementation of immediate automatic feedback is multiple-choice questions. Generating the questions in multiple-choice format simplifies any issues related to

The associate editor coordinating the review of this manuscript and approving it for publication was Dongxiao Yu¹.

interpreting the answers [7]. Unit testing evaluation of programming assessments was also an early entry into the area, given its simplicity of usage and implementation [8]–[10]. With the broader availability of machine learning in education [11], the field expanded to include the grading of short essays [12] and long essays [13]. It also started to address other more complex problems, such as grading code correctness [14].

With the expansion of automatic scoring and feedback as a tool, several issues have emerged. From a technical perspective, devices based on machine learning need data, potentially in substantial amounts, to be accurate [15]. From an educational perspective, authors, including Bancroft [16], affirm that automatically scored tests, for example, multiple-choice tests, “do not test anything more than just straight recall of facts.” Given these potential issues, studies on automatic feedback, problem set up, and its effects on student’s education and experience are still being produced [17].

This review attempts to expand the literature on the effects of using automatic scoring and feedback as a learning tool, emphasizing its impact on the students’ learning experience. To achieve that goal, it focuses on these research questions:

-RQ1 What types of automatic scoring and automatic feedback are in use?

-RQ2 What are the positive effects on education goals of using automatic feedback and automatic scoring?

-RQ3 What are the positive effects on the student experience of using automatic feedback and automatic scoring?

-RQ4 What are the adverse effects on educational goals and student experience using automatic feedback and automatic scoring?

-RQ5 What type of evaluation was carried out to measure the effect of automatic scoring and feedback on student academic performance?

-RQ6 What improvements can be made to mitigate the adverse effects in RQ4?

Below, we list most tools and application fields to present current automatic scoring and feedback (RQ1). We present most tools currently being used, even though some are on their way to obsolesce, to show the field’s evolution. Concerning the positive effects (RQ2), we look at the experiences and opportunities these technologies have enabled, focusing on the student (RQ3). We then shift focus to the problems these tools may introduce from a learning perspective and a student experience perspective (RQ4). We also analyze the size evaluations conducted to examine the effect of effect on the academic performance of students (RQ5). Having studied the adverse effects, we look at possible improvements that mitigate the findings on RQ4 (RQ6).

The following ten sections of this paper continue with a discussion of previous studies, the process used to carry out the review, the most relevant findings, and interpretation of those findings, and potential avenues for future research.

II. RELATED WORK

Multiple papers have investigated the state of automatic scoring and feedback. These papers tend to focus on ways to

TABLE 1. Related works.

| Study | Purpose | Findings |
|-------|---|---|
| [27] | This paper conducted a review of 146 publications from 2007 to 2018. It looked at the overall state of the usage of artificial intelligence in education. | Multiple levels of education use types of artificial intelligence. Automatic scoring is among these types—however, there are others, like student performance prediction, material curation, and course adaptation. The paper predicts that artificial intelligence will continue to expand its place in education and that multiple improvements will happen in the next 20 years. |
| [28] | This work reviewed 93 works produced between 2015 and 2019. It analyzed the automatic generation of questions as a way to improve their quality. | The review found that this is an area under development. Multiple different approaches are being used. However, the complexity of standardizing ways to measure effectiveness makes it hard to show improvements produced by new methods. |
| [29] | This review analyzed 44 papers published between 2003 and 2016. It investigates datasets, machine learning techniques, commonly used features, and quality results. | The study found that some of these tools are being used. The technology is under continuous evolution, and that the publication of data sets has opened the field for further research and improvement. |

use the tools and the quality of their output. Few of them concentrate on the educational effects of using the technologies. Table 1 provides an overview of some of the studies and their findings. Among their conclusions is that automatic feedback is being used more in structured questions that require well-defined answers. These questions include multiple-choice [18], fill-in-the-blank [19], or those with a solution presented in a structured language, i.e., a mathematical formula [20] or a program [21], [22]. The main positive effects of automatic feedback include the students using the feedback for improvement [23], increased student engagement [24], [25], and reduction of instructor bias [26]. Despite its benefits, automatic scoring is only one of the potential uses for machine learning and can be expanded to encompass others, including performance prediction, material curation, and course adaptability [27]. However, automatic feedback has some drawbacks, including the complexity of measuring the feedback quality compared to a manual grader [28].

This work looks at reviewing a broader set of papers compared to Table 1, focusing on examining the effect of feedback on the student experience and identifying opportunities to improve the automatic feedback mechanisms from a student experience perspective.

The previous reviews do not determine the effect of automatic scoring and feedback on students’ performance. In this paper, we advance this issue.

III. METHODS

This study follows the steps described in [30] and involves three stages: 1) planning, 2) conducting, and 3) reporting.

IV. PLANNING

This part of the work included creating a strategy to select the most relevant result that helped address the research questions. We performed an iterative search using Web of Science as a platform for the search given the quality of its database and the iterative filtering capabilities [31].

We searched for the terms “automatic scoring’ AND education,” “automatic grading’ AND education,” “automatic feedback’ AND education,” and “machine learning’ AND education.” The search included only works published from 2016 to mid-2020. The first query returned 15 papers for the first terms, the second 19, the third 27, and the fourth 1233. We restricted the last one by refining the search results by looking for “scoring” or “feedback.” The refined result of all the search refinements included 256 papers which were then manually filtered by title using the inclusion criteria.

V. INCLUSION CRITERIA

The works selected for review fell under the following parameters:

- 1-The study focuses on the use of technologies in education.
- 2-The study helps answer at least one of the research questions.
- 3-The study was published after peer review.

After the criteria were applied, the list of works was reduced to 125.

VI. CONDUCTING THE REVIEW AND REPORTING

After completing the planning, a content analysis was carried out. As mentioned by [32], content analysis allows one to find the research trends by analyzing the articles’ content and grouping them according to the shared characteristics. We created a collection form to code the information relevant to answering the research questions. The columns included in that form are shown in Table 2. Each paper was thoroughly reviewed by three of the authors using a shared Excel file. A simple majority (2 votes) was needed for the value to be selected for the categorical values. For the open-ended questions also two votes/appearances were used to keep an answer. This analysis was used to group the papers, and the groupings were used to answer the research questions.

VII. FINDINGS

This section shows the current trends in automatic feedback and scoring. First, we present general findings, followed by the analysis of each research question.

A. GENERAL FINDINGS

From the selected papers, 12% are from 2016, 16% from 2017, 27% from 2018, 28% from 2019, and 17% from the early part of 2020. This trend suggests a likely increase in interest in the subject. Figure 1 shows this behavior.

TABLE 2. Coding criteria.

| Code | Values |
|---|---|
| <i>Related to automatic scoring and/or feedback</i> | Yes/No |
| <i>Publication Year</i> | 2016-2020 |
| <i>Education</i> | Yes/No |
| <i>Research Question</i> | 1/2/3/4/5/6/Multiple |
| <i>Automated Feature</i> | Scoring/Feedback/Both |
| <i>One time experience</i> | Yes/No |
| <i>Experiment</i> | Yes/No |
| <i>Field of Education</i> | Subject Independent/Science, Math, Computer/Arts, and Humanity/Medicine |
| <i>Technology Area</i> | Structured Answer/Short Free Form/Long Free Form/Others |
| <i>Technology Type</i> | Static/Dynamic |
| <i>Positive Educational Effects</i> | Open-Ended |
| <i>Positive Student Experience</i> | Open-Ended |
| <i>Negative Educational Effects</i> | Open-Ended |
| <i>Negative Student Experience</i> | Open-Ended |

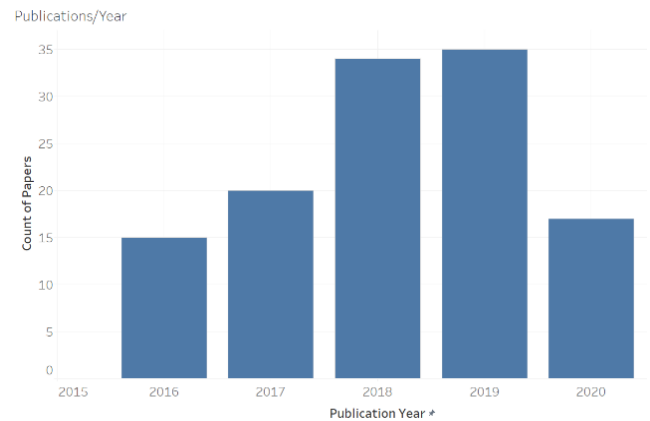


FIGURE 1. Number of studies over time.

B. EDUCATIONAL LEVEL

According to the International Standard Classification of Education [33], most of the work reviewed was at the bachelor’s or equivalent level (92% of papers), with small numbers at the early education (2% of papers), such as Saha’s study of automatic grading of explanatory answers in middle school [34] and secondary education (6% of papers) such as Anohah’s analysis of high school computing science courses [35].

This distribution supports the fact that technologies used for automatic grading and feedback require information technology and a medium understanding of language and mathematics. Despite this, some of the works dealt with teaching topics in early education, including handwriting [36] and basic math [20]. See Figure 2.

C. FIELD OF EDUCATION

Most of the papers addressed works that have effects across disciplines, e.g., using student data to predict

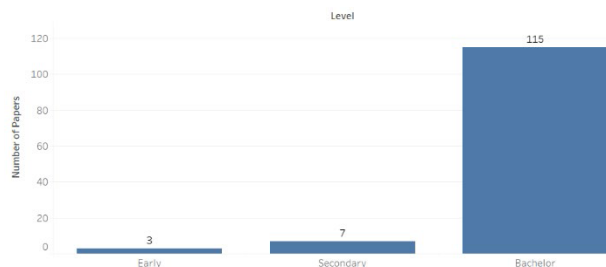


FIGURE 2. Educational level.

performance [26]. For the discipline-specific ones and using the International Standard Classification of Education (ISCED) [33], most of the work fell into the categories of the sciences [37], including areas like geology [38], mathematics [20], computer science [24], computer networking [39] (47% of papers). This is followed by cross-disciplinary applications (32% of the papers), followed by art and humanities (21% papers). This set is completed by applications in medicine where virtual reality and other technologies are being used to support immersive practical experiences such as virtual artificial intelligent assistants [40], surgical skill assessments [41], [42], physiotherapy training [43], and clinical skills [44]. A couple of papers target areas such as music where immediate feedback is also used to improve the student experience in general musical learning [45] and instrument learning [46].

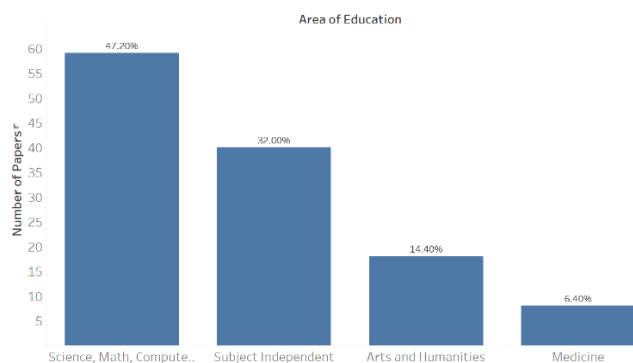


FIGURE 3. Field of education.

The following sections will describe the answers according to each research question.

D. RQ1 WHAT TYPES OF AUTOMATIC SCORING AND AUTOMATIC FEEDBACK ARE IN USE?

The types of automatic scoring and feedback can be divided into two dimensions. The first is the input-form, and the second is the mechanism used for auto-grading and generating feedback.

From the input perspective, the primary forms are structured. These include mathematics [47], code [22], and controlled environments such as simulations [40]. Other inputs include a short free form (e.g., a short sentence [12]), long free form (e.g., an essay [13]). The main mechanisms are

static or dynamic. Static ones include comparing the answer to a key or set of keys [16] or running a fixed set of unit test cases [48]. Dynamic ones include comparing the answer to other student answers or using machine learning to learn expected grades from past answers.

The tools used to produce grades and feedback dynamically include ontologies [7], neural networks to identify possible solutions [42], [47], machine learning used to identify learning paths [49], [50], and machine learning used to determine student’s risk of falling [51].

Table 3 summarizes the tools and techniques in each area identified in the works analyzed.

TABLE 3. Mechanisms.

| Input | Static | Dynamic |
|---|--------|---------|
| Structured | 5% | 46% |
| Short Free Form | - | 2% |
| Long Free Form | - | 14% |
| Others (i.e., virtual reality, handwriting, etc.) | - | 33% |

E. RQ2 WHAT ARE THE POSITIVE EFFECTS ON EDUCATION GOALS OF USING AUTOMATIC FEEDBACK AND AUTOMATIC SCORING?

Most of the analyzed papers concluded that there were educational advantages to automatic feedback and scoring. The most commonly mentioned advantages included bias reduction [52] and grading consistency [26], the ability for the instructor to shift focus away from grading into other activities [49], and allowing more students to participate in the learning experiences [53]. Table 4 summarizes the benefits.

TABLE 4. Benefits.

| Benefit | Papers |
|--|--------|
| Reduction in bias and increase inconsistency of the grading | 68% |
| Instructors to focus on other activities instead of grading, i.e., working with students struggling with the material. | 21% |
| Ability to provide education to a larger number of students at the same time (i.e., MOOCs) | 12% |

F. RQ3 WHAT ARE THE POSITIVE EFFECTS ON THE STUDENT EXPERIENCE OF USING AUTOMATIC FEEDBACK AND AUTOMATIC SCORING?

Very few of the works focused on the student experience, instead emphasizing learning experiences. They primarily highlighted students’ positive reception of features such as immediate grading with the allowance of multiple submissions [54]. Together with this, several works focused on the ability to create custom learnings paths based on student performance [55], [26], [53]–[58], and the ability to flag students at high risk of not succeeding [51], [59]–[63], [61]. Table 5 summarizes the benefits found in the review.

TABLE 5. Benefits.

| Benefit | Papers |
|--|--------|
| <i>Ability to iterate over problems instead of only having one opportunity to get the correct answer</i> | 51% |
| <i>Ability to create personalized paths and learn at the student's own pace.</i> | 6% |
| <i>Ability to be warned of at-risk status</i> | 6% |

G. RQ4 WHAT ARE THE ADVERSE EFFECTS ON EDUCATIONAL GOALS AND STUDENT EXPERIENCE OF USING AUTOMATIC FEEDBACK AND AUTOMATIC SCORING?

Most of the papers focused on the usability of the tools and techniques they work with and evaluated ways to replace current practices with equivalent or better methods, with very few adverse side effects. When detected, these side effects included students losing the social aspect of learning [64]. This was replaced by human-computer interactions and students learning to work within the system (e.g., creating multiple accounts in a MOOC to gain access to the answers [65]).

Very few studies assessed the adverse effects on student experience, although some included sections on potential issues that need to be further studied. These issues included students learning to solve the assessment questions without understanding the underlying concepts [66]. This phenomenon is not exclusive to automatic feedback, as studies have shown that the number of past tests studied is a strong indicator of future tests [67]. Other potential adverse effects included loss of human interaction and lack of interpersonal skills while solving problems [64], and lack of personalized feedback that could help outlier students [68], especially struggling students [69].

H. RQ5 WHAT TYPE OF EVALUATION WAS USED TO MEASURE THE EFFECT OF AUTOMATIC SCORING AND FEEDBACK ON STUDENT ACADEMIC PERFORMANCE?

As shown in Table 6, some of the papers contained experiments related to the tool's quality or the algorithm. Examples include [70]–[74] (36%). Others included a one-group experiment (34%) and an individual case study (29%).

TABLE 6. Evaluation type.

| Benefit | Papers |
|------------------------------|--------|
| <i>Algorithm Performance</i> | 36% |
| <i>One Group Experiment</i> | 34% |
| <i>Individual Case Study</i> | 29% |

These results show the need for further experiments following this pattern to understand better the actual effects of automatic feedback and scoring on student academic performance.

I. RQ6 WHAT IMPROVEMENTS ARE BEING MADE TO MITIGATE THE ADVERSE EFFECTS IN RQ4?

Some of the effects cannot be easily mitigated, i.e., bringing back the student-professor interaction [29]. Automatically, automatic systems are becoming popular are impossible to keep large MOOCs [66]. Chatbot technologies could eventually help this area provide a more personalized experience [75], [76]. Specific learning can be mitigated by generating dynamic problems unique for each student [77], [78]. From a student perspective, work can be developed to improve the design and delivery of the automatic feedback to improve the experience, including finding a way to personalize the feedback [79]. Finally, it is essential to mention the need for more long-term studies to understand the impact of feedback on the students' experience.

VIII. DISCUSSION

The results of this review suggest that automatic scoring and feedback is an area undergoing constant improvements as technology evolves and data becomes available. The use of automatic scoring and feedback has led to the expansion of MOOCs and online courses and the ability to support large students in the same program [53]. Automated scoring and feedback are not only present both in MOOCs [80] and other systems where the scale requires it but also in smaller settings as a tool to support learning, including introductory programming classes [21], [81]–[83]. This new capability has led universities to open their programs to more applicants and allowed more students to go through those programs.

The most common uses of automatic scoring and feedback are in three areas: 1) programming problems through mechanisms such as assisting the student with the coding [84], [85], analyzing coding patterns [86], [87], automatic grading [88]–[92], and customized feedback [84], [86], [85]; 2) short essays [93]; and 3) extended essays [94]. Programming problems are the easiest to use as input for this technology, as they appear in a structured language that computers can understand [95]. Short essays can also be looked at, as their complexity tends to be low [96], while long essays prove the most considerable challenge for this technology [97]. With this in mind, automatic scoring and feedback are being employed broadly in computer science, mathematics [98]–[101], [47], and similarly analytical courses, together with language-learning areas [102]–[105], [36], [97].

With the expansion of these technologies, we expect to see the benefits presented in the works surveyed in this paper materialized beyond the furthering of access, particularly an improvement in grading consistency [106] and the freeing of instructor time to dedicate to other activities [107]. An analysis of the effects of this time re-allocation would provide more information on the end effects of this benefit. We also expect to see an increase in student engagement driven by the ability to solve problems in a more interactive way given the feedback [108]–[110]. Looking at the effects of this engagement on learning is another area of future study. Similarly, a more personalized student experience would be expected to

lead to better matching between the learning experience and the student's learning style [111]–[115].

On the other hand, the potential adverse effects of using automatic feedback cannot be ignored. Students can solve problems without learning the underlying concepts. Using the feedback as a trial-and-error exercise or accessing answers using the internet can cause a very detrimental effect on learning [116]. Similarly, where solutions are not easily defined or grouped, subjects will have a more challenging time implementing these technologies as they are less developed in those areas.

IX. LIMITATIONS

This study addresses only some relevant questions when analyzing the extensive use of automatic scoring and feedback. There are fundamental questions about the quality of content and student privacy [117], for example, which are not considered in the study. The study also does not reveal funding and other possible biases affecting the underlying studies and does not focus on the specific tools used to implement the technologies. The research does not focus on features and requirements for automatic scoring and feedback tools or possible solutions to many challenges.

X. CONCLUSION AND FUTURE WORK

This work presents a systematic review of the literature with an analysis of 125 studies focused on using automatic scoring and feedback. Results indicate that these technologies play an essential role in expanding access to education and are still evolving. The use of these technologies is also growing both in large and small classes in multiple areas. The number of application areas, tools being used, and published works in this area are increasing. This trend is most likely related to a combination of technological advances and the need to serve more students.

This review shows the current state of automatic scoring and feedback and identifies areas of potential improvement and further analysis. Among these areas, the study of the effects on educational quality and student experience is highly relevant.

REFERENCES

- [1] F. Dong and Y. Zhang, "Automatic features for essay scoring—an empirical study," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 1072–1077.
- [2] A. Chauhan, "Massive open online courses (MOOCs): Emerging trends in assessment and accreditation," *Digit. Educ. Rev.*, vol. 25, pp. 7–17, Jul. 2014.
- [3] Y. Cao, L. Porter, S. N. Liao, and R. Ord, "Paper or online? A comparison of exam grading techniques," in *Proc. ACM Conf. Innov. Technol. Comput. Sci. Educ.*, Jul. 2019, pp. 99–104.
- [4] B. Csapó and G. Molnár, "Online diagnostic assessment in support of personalized teaching and learning: The eDia system," *Frontiers Psychol.*, vol. 10, p. 1522, Jul. 2019.
- [5] A. P. Cavalcanti, R. F. L. de Mello, V. Rolim, M. Andre, F. Freitas, and D. Gasevic, "An analysis of the use of good feedback practices in online learning courses," in *Proc. IEEE 19th Int. Conf. Adv. Learn. Technol. (ICALT)*, Jul. 2019, pp. 153–157.
- [6] N. Alruwais, G. Willis, and M. Wald, "Advantages and challenges of using E-assessment," *Int. J. Inf. Educ. Technol.*, vol. 8, no. 1, pp. 34–37, 2018.
- [7] M. Cubric and M. Tomic, "Design and evaluation of an ontology-based tool for generating multiple-choice questions," *Interact. Technol. Smart Educ.*, vol. 17, no. 2, pp. 109–131, Jun. 2020.
- [8] S. H. Edwards and M. A. Perez-Quinones, "Web-CAT: Automatically grading programming assignments," in *Proc. 13th Annu. Conf. Innov. Technol. Comput. Sci. Educ.*, Jun. 2008, p. 328.
- [9] A. Wakatani and T. Maeda, "Web applications for learning CUDA programming," in *Proc. 8th Int. Conf. Inf., Intell., Syst. Appl. (IISA)*, Aug. 2017, pp. 1–5.
- [10] J. DeNero, S. Sridhara, M. Pérez-Quinones, A. Nayak, and B. Leong, "Beyond autograding: Advances in student feedback platforms," in *Proc. ACM SIGCSE Tech. Symp. Comput. Sci. Educ.*, Mar. 2017, pp. 651–652.
- [11] Y. Takahashi and P. Vate-U-Lan, "Toward understanding the impact of artificial intelligence on education: An empirical research in Japan," in *Proc. Eur. Conf. Impact Artif. Intell. Robot., Academic*, 2019, p. 433.
- [12] F. S. Pribadi, T. B. Adji, A. E. Permasari, A. Mulwinda, and A. B. Utomo, "Automatic short answer scoring using words overlapping methods," *AIP Conf. Proc.*, vol. 1818, Mar. 2017, Art. no. 020042.
- [13] K. Taghipour and H. T. Ng, "A neural approach to automated essay scoring," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 1882–1891.
- [14] J. Huang, C. Pieh, A. Nguyen, and L. Guibas, "Syntactic and functional variability of a million code submissions in a machine learning MOOC," in *Proc. AIED Workshops Proc.*, 2013, pp. 1–15.
- [15] M. Zhang, "Contrasting automated and human scoring of essays," *R&D Connections*, vol. 21, no. 2, pp. 1–11, 2013.
- [16] P. Bancroft and K. Woodfors, "Using multiple choice questions effectively in information technology education," *Ascilite*, vol. 4, pp. 948–955, Dec. 2004.
- [17] O. Bulut, M. Cutumisu, A. M. Aquilina, and D. Singh, "Effects of digital score reporting and feedback on students' learning in higher education," *Frontiers Educ.*, vol. 4, p. 65, Jun. 2019.
- [18] S. Narayanan, V. S. Kommuri, S. N. Subramanian, and K. Bijlani, "Question bank calibration using unsupervised learning of assessment performance metrics," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2017, pp. 19–25.
- [19] A. Sahu and P. K. Bhowmick, "Feature engineering and ensemble-based approach for improving automatic short-answer grading performance," *IEEE Trans. Learn. Technol.*, vol. 13, no. 1, pp. 77–90, Jan. 2020.
- [20] J. C. S. Kadupitiya, S. Ranathunga, and G. Dias, "Automated assessment of multi-step answers for mathematical word problems," in *Proc. 16th Int. Conf. Adv. ICT Emerg. Regions (ICTer)*, Sep. 2016, pp. 66–71.
- [21] S. Parihar, Z. Dadachanji, P. K. Singh, R. Das, A. Karkare, and A. Bhattacharya, "Automatic grading and feedback using program repair for introductory programming courses," in *Proc. ACM Conf. Innov. Technol. Comput. Sci. Educ.*, Jun. 2017, pp. 92–97.
- [22] S. Li, X. Xiao, B. Bassett, T. Xie, and N. Tillmann, "Measuring code behavioral similarity for programming and software engineering education," in *Proc. 38th Int. Conf. Softw. Eng. Companion*, May 2016, pp. 501–510.
- [23] M. Marchisio, A. Barana, M. Fioravera, S. Rabellino, and A. Conte, "A model of formative automatic assessment and interactive feedback for STEM," in *Proc. IEEE 42nd Annu. Comput. Softw. Appl. Conf. (COMPSAC)*, Jul. 2018, pp. 1016–1025.
- [24] S. Marwan, G. Gao, S. Fisk, T. W. Price, and T. Barnes, "Adaptive immediate feedback can improve novice programming engagement and intention to persist in computer science," in *Proc. ACM Conf. Int. Comput. Educ. Res.*, Aug. 2020, pp. 194–203.
- [25] H. Guei, T. H. Wei, and I. Wu, "2048-Like games for teaching reinforcement learning," *ICGA J.*, vol. 42, pp. 14–37, Jan. 2020.
- [26] Q. Chen, X. Wang, and Q. Zhao, "Appearance discrimination in grading?—Evidence from migrant schools in China," *Econ. Lett.*, vol. 181, pp. 116–119, Aug. 2019.
- [27] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education—where are the educators?" *Int. J. Educ. Technol. Higher Educ.*, vol. 16, no. 1, pp. 1–27, Dec. 2019.
- [28] G. Kurdi, J. Leo, B. Parsia, U. Sattler, and S. Al-Emari, "A systematic review of automatic question generation for educational purposes," *Int. J. Artif. Intell. Educ.*, vol. 30, no. 1, pp. 121–204, Mar. 2020.
- [29] L. B. Galhardi and J. D. Brancher, "Machine learning approach for automatic short answer grading: A systematic review," in *Proc. Ibero-Amer. Conf. Artif. Intell.*, 2018, pp. 380–391.

- [30] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *J. Syst. Softw.*, vol. 80, no. 4, pp. 571–583, Apr. 2007.
- [31] A. Martín-Martín, E. Orduna-Malea, and E. D. López-Cózar, "Coverage of highly-cited documents in Google scholar, web of science, and scopus: A multidisciplinary comparison," *Scientometrics*, vol. 116, no. 3, pp. 2175–2188, Sep. 2018.
- [32] Y.-H. Ching and Y.-C. Hsu, "Peer feedback to facilitate project-based learning in an online environment," *Int. Rev. Res. Open Distrib. Learn.*, vol. 14, no. 5, pp. 258–276, Dec. 2013.
- [33] *International Standard Classification of Education*, ISCED, London, U.K., 2011.
- [34] S. K. Saha and D. Rao, "Development of a practical system for computerized evaluation of descriptive answers of middle school level students," *Interact. Learn. Environ.*, vol. 4, pp. 1–14, Aug. 2019.
- [35] E. Anohah, "Pedagogy and design of online learning environment in computer science education for high schools," *Int. J. Online Pedagogy Course Design*, vol. 6, no. 3, pp. 39–51, Jul. 2016.
- [36] M. Candeias, M. G. A. D. Reis, J. Escola, and M. J. C. S. Reis, "Using Android tablets to develop handwriting skills: A case study," *Heliyon*, vol. 5, no. 12, Dec. 2019, Art. no. e02970.
- [37] X. Zhai, Y. Yin, J. W. Pellegrino, K. C. Haudek, and L. Shi, "Applying machine learning in science assessment: A systematic review," *Stud. Sci. Educ.*, vol. 56, no. 1, pp. 111–151, Jan. 2020.
- [38] S. M. Sit and M. R. Brudzinski, "Creation and assessment of an active E-learning introductory geology course," *J. Sci. Educ. Technol.*, vol. 26, no. 6, pp. 629–645, Dec. 2017.
- [39] V. Muniyasamy, I. M. Ejalani, and M. Anadhavalli, "Student's performance assessment and learning skill towards wireless network simulation tool-cisco packet tracer," *Int. J. Eng. Technol.*, vol. 14, no. 7, pp. 196–208, 2019.
- [40] N. Mirchi, V. Bissonnette, R. Yilmaz, N. Ledwos, A. Winkler-Schwartz, and R. F. Del Maestro, "The virtual operative assistant: An explainable artificial intelligence tool for simulation-based training in surgery and medicine," *PLoS ONE*, vol. 15, no. 2, Feb. 2020, Art. no. e0229596.
- [41] M. Levin, T. McKechnie, S. Khalid, T. P. Grantcharov, and M. Goldenberg, "Automated methods of technical skill assessment in surgery: A systematic review," *J. Surg. Educ.*, vol. 76, no. 6, pp. 1629–1639, Nov. 2019.
- [42] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Müller, "Accurate and interpretable evaluation of surgical skills from kinematic data using fully convolutional neural networks," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 9, pp. 1611–1617, Sep. 2019.
- [43] M. Jovanović, J. Seiffarth, E. Kutafina, and S. M. Jonas, "Automated error detection in physiotherapy training," *Stud. Health Technol. Informat.*, vol. 248, pp. 164–171, Jan. 2018.
- [44] Y. Sugamiya, T. Otani, R. Nakadate, and A. Takanishi, "Construction of automatic scoring system to support objective evaluation of clinical skills in medical education," in *Proc. 41st Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 4177–4181.
- [45] S. Giraldo, G. Waddell, I. Nou, A. Ortega, O. Mayor, A. Perez, A. Williamson, and R. Ramirez, "Automatic assessment of tone quality in violin music performance," *Frontiers Psychol.*, vol. 10, p. 334, Mar. 2019.
- [46] J. Kosakaya, R. Kawamorita, and M.-F. Hsu, "A cooperative multi-agent-based musical scoring system for Tsugaru and Nambu Shamisen," in *Proc. IEEE Int. Conf. Ind. Eng. Eng. Manage. (IEEM)*, Dec. 2018, pp. 1061–1067.
- [47] N. Wiggins and M. Smith, "Free response evaluation via neural network for an IMathAS system," in *Proc. IEEE Int. Symp. Meas. Control Robot. (ISMCR)*, Sep. 2019, p. 1.
- [48] K. Wang, R. Singh, and Z. Su, "Search, align, and repair: Data-driven feedback generation for introductory programming exercises," in *Proc. 39th ACM SIGPLAN Conf. Program. Lang. Design Implement.*, Jun. 2018, pp. 481–495.
- [49] T. Saito and Y. Watanobe, "Learning path recommendation system for programming education based on neural networks," *Int. J. Distance Educ. Technol.*, vol. 18, no. 1, pp. 36–64, Jan. 2020.
- [50] C. Srisa-An and K. Yongsiriwit, "Applying machine learning and AI on self automated personalized online learning," in *Proc. FSDM*, 2019, pp. 137–145.
- [51] J. L. Harvey and S. A. P. Kumar, "A practical model for educators to predict student performance in K-12 education using machine learning," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Dec. 2019, pp. 3004–3011.
- [52] K. DiCerbo, "Assessment for learning with diverse learners in a digital world," *Educ. Meas., Pract.*, vol. 39, no. 3, pp. 90–93, Sep. 2020.
- [53] L. Singelmann, E. Swartz, M. Pearson, R. Striker, and E. A. Vazquez, "Design and development of a machine learning tool for an innovation-based learning MOOC," in *Proc. IEEE Learn. (LWMOOS)*, Oct. 2019, pp. 105–109.
- [54] E. Faulconer, J. C. Griffith, and H. Frank, "If at first you do not succeed: Student behavior when provided feedforward with multiple trials for online summative assessments," *Teaching Higher Educ.*, vol. 26, no. 4, pp. 586–601, May 2021.
- [55] M. A. Alsuaiket, A. H. Blasi, and K. Altarawneh, "Refining student marks based on enrolled modules assessment methods using data mining techniques," 2020, *arXiv:2009.06381*. [Online]. Available: <http://arxiv.org/abs/2009.06381>
- [56] K. Niemeijer, R. Feskens, G. Krempel, J. Koops, and M. J. S. Brinkhuis, "Constructing and predicting school advice for academic achievement: A comparison of item response theory and machine learning techniques," in *Proc. 10th Int. Conf. Learn. Anal. Knowl.*, Mar. 2020, pp. 462–471.
- [57] S. Cunningham-Nelson, M. Baktashmotlagh, and W. Boles, "Visualizing student opinion through text analysis," *IEEE Trans. Educ.*, vol. 62, no. 4, pp. 305–311, Nov. 2019.
- [58] S. A. Raza, "Predicting collaborative performance at assessment level using machine learning," in *Proc. 2nd Int. Conf. Comput. Appl. Inf. Secur. (ICCAIS)*, May 2019, pp. 1–6.
- [59] M. C. S. Manzanares, R. M. Sánchez, Á. A. González, M. del Camino Escolar Llamazares, and M. Á. Q. Dios, "Detection of at-risk students with Learning Analytics Techniques," *Eur. J. Invest. Health, Psychol. Educ.*, vol. 8, no. 3, pp. 129–142, 2018.
- [60] A. M. Radwan and Z. Cataltepe, "Improving performance prediction on education data with noise and class imbalance," *Intell. Autom. Soft Comput.*, vol. 5, pp. 1–8, Jul. 2017.
- [61] M. Ciolacu, A. F. Tehrani, L. Binder, and P. M. Svasta, "Education 4.0—Artificial intelligence assisted higher education: Early recognition system with machine learning to support students' success," in *Proc. IEEE 24th Int. Symp. Design Technol. Electron. Packag. (SIITME)*, Oct. 2018, pp. 23–30.
- [62] L. P. Alonso, "Learning process analysis using machine learning techniques," *Int. J. Eng. Educ.*, vol. 34, no. 3, pp. 981–989, 2018.
- [63] F. Chen and Y. Cui, "Utilizing student time series behaviour in learning management systems for early prediction of course performance," *J. Learn. Anal.*, vol. 7, no. 2, pp. 1–17, Sep. 2020.
- [64] D. Gamage, I. Perera, and S. Fernando, "MOOCs lack interactivity and collaborativeness: Evaluating MOOC platforms," *Int. J. Eng. Pedagogy*, vol. 10, no. 2, p. 94, Mar. 2020.
- [65] G. Alexandron, J. A. Ruipérez-Valiente, Z. Chen, P. J. Muñoz-Merino, and D. E. Pritchard, "Copying@ scale: Using harvesting accounts for collecting correct answers in a MOOC," *Comput. Educ.*, vol. 108, pp. 96–114, May 2017.
- [66] P. Topali, A. Ortega-Arranz, E. Er, A. Martínez-Monés, S. L. Villagrà-Sobrino, and Y. Dimitriadis, "Exploring the problems experienced by learners in a MOOC implementing active learning pedagogies," in *Proc. Eur. MOOCs Stakeholders*, 2019, pp. 81–90.
- [67] O. O. Adesope, D. A. Trevisan, and N. Sundararajan, "Rethinking the use of tests: A meta-analysis of practice testing," *Rev. Educ. Res.*, vol. 87, no. 3, pp. 659–701, Jun. 2017.
- [68] E. Jensen, M. Dale, P. J. Donnelly, C. Stone, S. Kelly, A. Godley, and S. K. D'Mello, "Toward automated feedback on teacher discourse to enhance teacher learning," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2020, pp. 1–13.
- [69] T. Walls and C. Zwicky, "Transforming curriculum, exploring identity, and cultivating culturally responsive educators," in *Cultural Competence in Higher Education*. Emerald, 2020, pp. 67–82.
- [70] X. Wu, W. Zheng, D. Mu, and N. Li, "UTCPredictor: An uncertainty-aware novel teaching cases predictor," *Comput. Appl. Eng. Educ.*, vol. 27, no. 6, pp. 1518–1530, Nov. 2019.
- [71] A. Bey, D. Bouhineau, and R. Champagnat, "Human scoring versus automatic scoring of computer programs: Does Algo+ score as well as instructors? An experimental study," in *Proc. IEEE 18th Int. Conf. Adv. Learn. Technol. (ICALT)*, Jul. 2018, pp. 355–357.
- [72] J. R. Guasp, M. A. Ramírez, A. J. V. Garayo, J. M. Meseguer-Dueñas, and J. A. Gómez-Tejedor, "Students' perception of auto-scored online exams in blended assessment: Feedback for improvement," *Educación*, vol. 21, no. 2, pp. 79–103, May 2018.

- [73] J. R. Rico-Juan, A.-J. Gallego, and J. Calvo-Zaragoza, "Automatic detection of inconsistencies between numerical scores and textual feedback in peer-assessment processes with machine learning," *Comput. Educ.*, vol. 140, Oct. 2019, Art. no. 103609.
- [74] J. Ursenbach, M. E. O'Connell, J. Neiser, M. C. Tierney, D. Morgan, J. Kosteniuk, and R. J. Spiteri, "Scoring algorithms for a computer-based cognitive screening tool: An illustrative example of overfitting machine learning approaches and the impact on estimates of classification accuracy," *Psychol. Assessment*, vol. 31, no. 11, p. 1377, 2019.
- [75] J. Pereira and M. A. Barcina, "A chatbot assistant for writing good quality technical reports," in *Proc. 7th Int. Conf. Technol. Ecosyst. Enhancing Multiculturalism*, Oct. 2019, pp. 59–64.
- [76] V. Fernoaga, G. A. Stelea, C. Gavrilă, and F. Sandu, "Intelligent education assistant powered by Chatbots," in *Proc. Int. Sci. Conf. Learn. Softw. Educ.*, 2018, pp. 376–383.
- [77] M. A. R. Serra, A. Bikfalvi, J. S. Masó, and J. P. Garcia, "A generic tool for generating and assessing problems automatically using spreadsheets," *Int. J. Emerg. Technol. Learn.*, vol. 13, no. 10, p. 23, Oct. 2018.
- [78] S. A. Andrade Freitas, R. De Cassia Silva, T. F. R. Lucena, E. D. N. Ribeiro, V. C. De Lima, and R. M. S. Da Silva, "Smart quizzes in the engineering education," in *Proc. 49th Hawaii Int. Conf. Syst. Sci. (HICSS)*, Jan. 2016, pp. 66–73.
- [79] D. Azcona, I.-H. Hsiao, and A. F. Smeaton, "Detecting students-at-risk in computer programming classes with learning analytics from students' digital footprints," *User Model. User-Adapted Interact.*, vol. 29, no. 4, pp. 759–788, Sep. 2019.
- [80] M. Santamaría Lancha, M. Hernández, Á. S.-E. Paniagua, J. M. Luzón Encabo, and G. de Jorge-Botana, "Using semantic technologies for formative assessment and scoring in large courses and MOOCs," *J. Interact. Media Educ.*, vol. 2018, no. 1, pp. 1–10, Aug. 2018.
- [81] J. Krugel and P. Hubwieser, "Computational thinking as springboard for learning object-oriented programming in an interactive MOOC," in *Proc. IEEE Global Eng. Educ. Conf. (EDUCON)*, Apr. 2017, pp. 1709–1712.
- [82] S. Marginson, "Do rankings drive better performance?" *Int. Higher Educ.*, vol. 89, pp. 6–8, Apr. 2017.
- [83] B. Jiang, Y. He, R. Chen, C. Hao, S. Liu, and G. Zhang, "Progressive teaching improvement for small scale learning: A case study in China," *Future Internet*, vol. 12, no. 8, p. 137, Aug. 2020.
- [84] A. Rubinstein, N. Parzanchevski, and Y. Tamarov, "In-depth feedback on programming assignments using pattern recognition and real-time hints," in *Proc. ACM Conf. Innov. Technol. Comput. Sci. Educ.*, Jul. 2019, pp. 243–244.
- [85] W. Zheng, Y. Bai, and H. Che, "A computer-assisted instructional method based on machine learning in software testing class," *Comput. Appl. Eng. Educ.*, vol. 26, no. 5, pp. 1150–1158, Sep. 2018.
- [86] C. Bhanuprakash, Y. S. Nijagunarya, and M. A. Jayaram, "Performance analysis of students by evaluating their examination answer scripts by using soft computing techniques," in *Emerging Research in Electronics, Computer Science and Technology*. Springer, 2019, pp. 549–568.
- [87] A. Lobanov, T. Bryksin, and A. Shpilman, "Automatic classification of error types in solutions to programming assignments at online learning platform," in *Proc. Int. Conf. Artif. Intell. Educ.* Springer, 2019, pp. 174–178.
- [88] S. Pape, J. Flake, A. Beckmann, and J. Járjens, "STAGE: A software tool for automatic grading of testing exercises: Case study paper," in *Proc. 38th Int. Conf. Softw. Eng. Companion*, May 2016, pp. 491–500.
- [89] S. Király, K. Nehéz, and O. Hornyák, "Some aspects of grading Java code submissions in MOOCs," *Res. Learn. Technol.*, vol. 25, no. 5, pp. 1–16, Jul. 2017.
- [90] A. Bey, P. Jermann, and P. Dillenbourg, "A comparison between two automatic assessment approaches for programming: An empirical study on MOOCs," *J. Educ. Technol. Soc.*, vol. 21, no. 2, pp. 259–272, 2018.
- [91] A. A. Gamazo, X. B. Illa, M. J. Casany, C. M. Escofet, C. Quer, M. E. G. Rodríguez, Ó. R. Moral, A. U. Tubella, "A software tool for E-assessment of relational database skills," *Int. J. Eng. Educ.*, vol. 32, pp. 1289–1312, Jan. 2016.
- [92] G. Paravati, F. Lamberti, V. Gatteschi, C. Demartini, and P. Montuschi, "Point cloud-based automatic assessment of 3D computer animation courseworks," *IEEE Trans. Learn. Technol.*, vol. 10, no. 4, pp. 532–543, Oct. 2017.
- [93] G. G. Smith, R. Haworth, and S. Žitnik, "Computer science meets education: Natural language processing for automatic grading of open-ended questions in ebooks," *J. Educ. Comput. Res.*, vol. 58, no. 7, pp. 1227–1255, Dec. 2020.
- [94] L. B. Moreira, A. D. Tamariz, and J. V. Fettermann, "The use of texts mining in the support to corrections of discursive questions in a higher education institution," *Texto Livre-Linguagem Tecnol.*, vol. 11, no. 3, pp. 213–227, 2018.
- [95] Z. Huang, Q. Liu, C. Zhai, Y. Yin, E. Chen, W. Gao, and G. Hu, "Exploring multi-objective exercise recommendations in online education systems," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, Nov. 2019, pp. 1261–1270.
- [96] A. Shehab, M. Faroun, and M. Rashad, "An automatic Arabic essay grading system based on text similarity algorithms," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 3, pp. 263–268, 2018.
- [97] M. Liu, Y. Wang, W. Xu, and L. Liu, "Automated scoring of Chinese engineering students' English essays," *Int. J. Distance Educ. Technol.*, vol. 15, no. 1, pp. 52–68, Jan. 2017.
- [98] A. Suresh, T. Sumner, I. Huang, J. Jacobs, B. Foland, and W. Ward, "Using deep learning to automatically detect talk moves in teachers' mathematics lessons," in *Proc. IEEE Int. Conf. Big Data*, Dec. 2018, pp. 5445–5447.
- [99] M. Platz, M. Krieger, E. Niehaus, and K. Winter, "Electronic proofs in mathematics education—A South African Teacher Professional Development (TPD) course informing the conceptualisation of an E-proof system authoring support workshop," in *Proc. IST-Africa Week Conf. (IST-Africa)*, May 2017, pp. 1–9.
- [100] T. Sancho-Vinuesa, R. Masià, M. Fuertes-Alpiste, and N. Molas-Castells, "Exploring the effectiveness of continuous activity with automatic feedback in online calculus," *Comput. Appl. Eng. Educ.*, vol. 26, no. 1, pp. 62–74, Jan. 2018.
- [101] S. Hoch, F. Reinhold, B. Werner, J. Richter-Gebert, and K. Reiss, "Design and research potential of interactive textbooks: The case of fractions," *ZDM*, vol. 50, no. 5, pp. 839–848, Sep. 2018.
- [102] X. A. H. J. Dong, "An exploration of impact factors influencing students' reading literacy in Singapore with machine learning approaches," *Int. J. English Linguistics*, vol. 9, no. 5, pp. 52–65, 2019.
- [103] W. Qu, "Research on the application of automatic scoring system in college English writing," in *Proc. Int. Conf. Econ., Social Sci., Arts, Educ. Manage. Eng.*, 2016, pp. 1–5.
- [104] N. Kim, J.-H. Kim, M. K. Wolters, S. E. MacPherson, and J. C. Park, "Automatic scoring of semantic fluency," *Frontiers Psychol.*, vol. 10, p. 1020, May 2019.
- [105] Y. Huang, X. Yang, F. Zhuang, L. Zhang, and S. Yu, "Automatic Chinese reading comprehension grading by LSTM with knowledge adaptation," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*, 2018, pp. 118–129.
- [106] E. Hegarty-Kelly and D. A. Mooney, "Analysis of an automatic grading system within first year Computer Science programming modules," in *Proc. Conf. Comput. Educ. Pract.*, 2021, pp. 17–20.
- [107] 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, Aug. 2018.
- [108] R. Cobos and J. C. Ruiz-Garcia, "Improving learner engagement in MOOCs using a learning intervention system: A research study in engineering education," *Comput. Appl. Eng. Educ.*, vol. 29, no. 4, pp. 733–749, Jul. 2021.
- [109] K. Grindrod, K. Morris, and R. Killeen, "Assessing performance and engagement on a computer-based education platform for pharmacy practice," *Pharmacy*, vol. 8, no. 1, p. 26, Feb. 2020.
- [110] S. Fu, H. Gu, and B. Yang, "The affordances of AI-enabled automatic scoring applications on learners' continuous learning intention: An empirical study in China," *Brit. J. Educ. Technol.*, vol. 51, no. 5, pp. 1674–1692, Sep. 2020.
- [111] H.-C. Hung, I.-F. Liu, C.-T. Liang, and Y.-S. Su, "Applying educational data mining to explore students' learning patterns in the flipped learning approach for coding education," *Symmetry*, vol. 12, no. 2, p. 213, Feb. 2020.
- [112] C. A. Bacos, "Machine learning and education in the human age: A review of emerging technologies," in *Proc. Sci. Inf. Conf.* Springer, 2019, pp. 536–543.

- [113] F. Duzhin and A. Gustafsson, "Machine learning-based app for self-evaluation of teacher-specific instructional style and tools," *Educ. Sci.*, vol. 8, no. 1, p. 7, Jan. 2018.
- [114] M. Farhan, S. Jabbar, M. Aslam, M. Hammoudeh, M. Ahmad, S. Khalid, M. Khan, and K. Han, "IoT-based students interaction framework using attention-scoring assessment in eLearning," *Future Gener. Comput. Syst.*, vol. 79, pp. 909–919, Feb. 2018.
- [115] A. James, M. Kashyap, Y. H. Victoria Chua, T. Maszczyk, A. M. Nunez, R. Bull, and J. Dauwels, "Inferring the climate in classrooms from audio and video recordings: A machine learning approach," in *Proc. IEEE Int. Conf. Teach., Assessment, Learn. Eng. (TALE)*, Dec. 2018, pp. 983–988.
- [116] T. O'Riordan, D. E. Millard, and J. Schulz, "Is critical thinking happening? Testing content analysis schemes applied to MOOC discussion forums," *Comput. Appl. Eng. Educ.*, vol. 29, no. 4, pp. 690–709, Jul. 2021.
- [117] N. Norouzi and R. Hausen, "Quantitative evaluation of student engagement in a large-scale introduction to programming course using a cloud-based automatic grading system," in *Proc. IEEE Frontiers Educ. Conf. (FIE)*, Oct. 2018, pp. 1–5.
- [118] L. Bayerlein, "Students' feedback preferences: How do students react to timely and automatically generated assessment feedback?" *Assessment Eval. Higher Educ.*, vol. 39, no. 8, pp. 916–931, Nov. 2014.
- [119] H. Wei, H. Li, M. Xia, Y. Wang, and H. Qu, "Predicting student performance in interactive online question pools using mouse interaction features," in *Proc. 10th Int. Conf. Learn. Analytics Knowl.*, Mar. 2020, pp. 645–654.
- [120] S. M. R. Abidi, J. Ni, S. Ge, W. Xiangmeng, H. Ding, W. Zhu, and W. Zhang, "Demystifying help-seeking students interacting multimodal learning environment under machine learning regime," in *Proc. 11th Int. Conf. Graph. Image Process. (ICGIP)*, Jan. 2020, Art. no. 113732.
- [121] B. Hodgkinson, C. Lutteroth, and B. Wunsche, "GIGetFeedback—Towards automatic feedback and assessment for OpenGL 3D modelling assignments," in *Proc. Int. Conf. Image Vis. Comput. New Zealand (IVCNZ)*, Nov. 2016, pp. 1–6.
- [122] O. Mirmotahari, Y. Berg, S. Gjessing, E. Fremstad, and C. Damsa, "A case-study of automated feedback assessment," in *Proc. IEEE Global Eng. Educ. Conf. (EDUCON)*, Apr. 2019, pp. 1190–1197.
- [123] C. A. O. Niculescu, "Intelligent tutoring systems-trends on design, development and deployment," in *Proc. Conf. Learn. Softw. Educ.*, 2016, pp. 280–285.
- [124] O. Kassak, M. Kompan, and M. Bielikova, "Student behavior in a web-based educational system: Exit intent prediction," *Eng. Appl. Artif. Intell.*, vol. 51, pp. 136–149, May 2016.



MARCELO GUERRA HAHN (Senior Member, IEEE) received the bachelor's and master's degrees in computer science from the Universidad de la Republica, Uruguay. He is currently pursuing the Ph.D. degree with the Universidad Internacional de la Rioja. He is studying technologies associated with automatic assignment feedback and their effects on learning achievements and experiences. He is also a Guest Lecturer with the University of Washington and the Director of engineering at Sound Commerce.



SILVIA MARGARITA BALDIRIS NAVARRO received the bachelor's degree in systems and industrial engineering from the Industrial University of Santander (UIS), Colombia, and the master's degree in industrial informatics and automation and the Ph.D. degree in technologies from the University of Girona. She is currently an Associate Professor with the Universidad Internacional de La Rioja, Spain, and Fundación Universitaria Tecnológico Comfenalco, Colombia. Since her early twenties, she has been interested in research on how technologies can facilitate all students' inclusion in the educational systems. She has coordinated and participated in international projects and initiatives in Europe and North/South America, including serving on the editorial boards of high-impact scientific journals.



LUIS DE LA FUENTE VALENTÍN received the Ph.D. degree from the Universidad Carlos III de Madrid, in 2011. He is currently a full-time Associate Professor with the Universidad Internacional de La Rioja (UNIR). He has participated in several national and European public-funded projects, one of them as an investigator-in-charge. He has authored more than 40 articles. His current research interest includes machine learning tools applied to the educational field.



DANIEL BURGOS (Senior Member, IEEE) received the master's degree in artificial intelligence from MIT, the Ph.D. degree in communication, education, anthropology, theology, management, and open science and STEM, the Dr.Eng. degree in computer science, and the D.B.A. degree in business administration. He holds a position of the UNESCO chair in eLearning and the ICDE chair in open educational resources. He currently works as a Full Professor of technologies for education and communication and the Vice Rector for international research at the Universidad Internacional de La Rioja (UNIR). He also works as the Director of the Research Institute for Innovation and Technology in Education (UNIR iTED, <http://ited.unir.net>). He is also a Professor at An-Najah National University, Palestine, an Adjunct Professor at the Universidad Nacional de Colombia (UNAL), Colombia, an Extraordinary Professor at North-West University, South Africa, and a Visiting Professor at Coventry University, U.K. He is or has been involved in more than 60 European and worldwide research and development projects. He works as a Consultant for the United Nations (UNECE), the European Commission and Parliament, and the Russian Academy of Science.

...