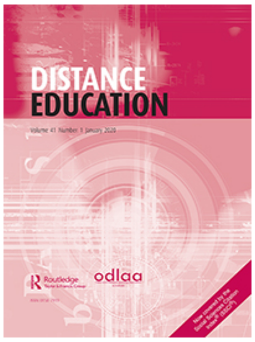


Distance Education

ISSN: 0158-7919 (Print) 1475-0198 (Online) Journal homepage: https://www.tandfonline.com/loi/cdie20

EduZinc: A tool for the creation and assessment of student learning activities in complex open, online and flexible learning environments

David Becerra-Alonso, Isabel Lopez-Cobo, Pilar Gómez-Rey, Francisco Fernández-Navarro & Elena Barbera

DISTANCE EDUCATION
2020, VOL. 41, NO. 1, 86–105
https://doi.org/10.1080/01587919.2020.1724769

Routledge
Taylor & Francis Group

Check for updates

EduZinc: A tool for the creation and assessment of student learning activities in complex open, online and flexible learning environments

David Becerra-Alonso[a], Isabel Lopez-Cobo[b], Pilar Gómez-Rey[b], Francisco Fernández-Navarro,[a] and Elena Barbera[c]

[a]Department of Quantitative Methods, Universidad Loyola Andalucía, Seville, Spain; [b]Department of Communication & Education, Universidad Loyola Andalucía, Seville, Spain; [c]Studies of Psychology and Education, Universitat Oberta de Catalunya, Barcelona, Spain

ABSTRACT

This article describes the development of an application for the grading and provision of feedback on educational processes. The too, named *EduZinc*, enables instructors to go through the complete process of creating and evaluating the activities and materials of a course. The application enables for the simultaneous management of two teaching-related aspects: (a) creation of individualized learning products (activities, tests and exams) and (b) automatic grading (for every learning product; automated creation of student, class, and competency-based reports; and delivery of personalized reports to students, instructors and tutors). The system also has a series of warnings in place to notify instructors and tutors when a student is falling behind. As a means to reward the efforts made during the course, the program keeps relevant statistics, notifying when a student is excelling in the course.

ARTICLE HISTORY

Received 1 February 2019
Accepted 30 January 2020

KEYWORDS

automatic grading tool; content creation; self-learning techniques; student-centred learning

Introduction

Decades of research on assessment at university level have highlighted the benefits of using assessment as the basis for the analysis of student progress and achievements (Ion et al., 2019). Studies have stressed the importance of challenges in assessing key competences and cross-sectional skills (Pellegrino, 2017; Pepper, 2011). They have also acknowledged the challenges around the use of different assessment techniques (Boud & Soler, 2016). Traditionally, the process of monitoring and evaluating student learning has been carried out manually. Over the last decades, however, the development of electronic systems to evaluate student learning has become more frequent (Paiva et al., 2017).

The advent of these tools is closely related to the shift from an information society to a knowledge society (Gallardo-Echenique et al., 2015). Higher education systems within the knowledge society are characterised mainly by the personalization of the learning process, the need for flexible educational systems, and the incorporation of technology in the classroom. Firstly, students require a learning context adapted to their personal needs and abilities.

CONTACT Pilar Gómez-Rey pgomez@uloyola.es
Supplemental data for this article can be accessed https://doi.org/10.1080/01587919.2020.1724769.

Another important component within contemporary educational systems is the creation of personalized course materials and learning for individualized professional development opportunities (especially important for adult learners) (Anshari et al., 2016). Secondly, students demand immediate feedback either because they are used to communicate with their peers through social media or because of professional reasons (Balakrishnan & Gan, 2016; Milligan & Littlejohn, 2017). Technological developments have also enabled the emergence of open, flexible, and technology-enhanced learning environments (such as online education programs and massive open online courses) that require the use of electronic systems (such as automatic grading tools) to evaluate and personalize the learning of the large number of students enrolled in their educational programs (Hew & Cheung, 2014). For all the above-mentioned reasons, the educational and computer science research communities have recently focused their attention on the development of these tools.

These tools for automated grading can be broadly divided in two groups: qualitative and quantitative. Qualitative tools generally employ a technique (called latent semantic analysis) to assess whether a set of learning outcomes (produced by the students), and the terms they contain are correlated with the "ideal" learning outcome (produced by the instructor) and its associated terms. Quantitative tools are of interest for the evaluation of subjects related to science, technology, engineering, and mathematics. The main advantage of this second group of systems is that they provide an objective score of the students' performance (De la Peña et al., 2012).

The tools discussed thus far are either too advanced or do not offer enough flexibility. Motivated by this fact, we developed an automated grading tool, named *EduZinc*, which is suitable for advanced material creation as well as usable by those with limited or no coding experience. The distinguishing feature of this software is that it enables the simultaneous management of the following teaching-related aspects: (a) creation of personalized learning products (activities, tests, and exams) and (b) automated grading for each learning product; automated creation of student, class, and competency-based reports; and the delivery of personalized reports to students and instructors. Thus, compared to other proposals, *EduZinc* is an open and flexible intelligent system that provides competency-based feedback. This feedback is not only generalized for the whole class but also for each student (as students receive different sets of exercises with different parameters).

Literature review

This section is devoted to a discussion of some of the most commonly automated evaluation systems used. We explored two types of automated grading tools organized according to their input data. First, we present an exploration of automated assessment systems for qualitative data. This is followed by an examination of existing automated grading tools for quantitative data.

Qualitative grading tools focus on automated essay evaluation systems (Page & Petersen, 1995; Sijimol & Varghese, 2018). When assessing qualitative texts with human graders, there are human variables to be considered, such as knowledge, emotion, and energy. One way to show the effect of these variables is to consider the deviation between scores provided by several instructors on the same work. Also, providing a judgment of text coherence and quality can be tedious and time-consuming. The main motivations behind developing qualitative computer-assisted assessment tools are to accurately assess the correctness of students' writings and to decrease the time in which students receive feedback. These automated measurement tools are capable of identifying similarities between texts based on previous knowledge acquired by a given model. Latent semantic analysis is the most common technique used for automatic essay grading (Farrús et al., 2010). Other approaches, such as n-gram co-occurrence and BLEU, have been also implemented to support the automatic assessment of essays (He et al., 2009). It is worth mentioning that previous research work has one thing in common: the outcome provided by the tool is subjective. These systems provide only a preliminary evaluation of the learning outcome (Farrús et al., 2010).

In the literature, we found two types of tools under the umbrella of quantitative evaluation: tools for evaluating programming code and tools for evaluating exercises related to mathematics and other scientific subjects. Following this categorisation, below is some research linked to the computer science field (specifically coding). The importance of the topic has motivated authors to perform a systematic literature review of automated feedback generation for programming exercises (Gupta & Gupta, 2017; Keuning et al., 2016). This type of tool automatically assesses programming tasks and focuses on qualifying the code developed by the student. These tools are under constant review and upgrade. Their evolution is aligned with that of programming languages. Some automated teaching assistants are related to the use of machine code and assembly language (Hollingsworth, 1960). Moreover, other tools run grading scripts based on utilities provided by the operating systems (Benford et al., 1994). Furthermore, there are systems that are based on Web technologies (Edwards & Perez-Quinones, 2008); T tools that enable granular evaluation by applying regression methods and patterns (Al Shamsi & Elnagar, 2012; Srikant & Aggarwal, 2014), and others integrated in learning management systems (Pohuba et al., 2014; Rodríguez-del Pino et al., 2012; Suleman, 2008). Table 1 presents, in alphabetical order, an overview of some of these assessment tools with information related to the programming languages that the platform supports, the work mode in which the systems are run, and the type of feedback generated by the tool.

Within quantitative tools for mathematics and science, many web-based systems collect and assess practical exercises from science and engineering courses. These applications share the use of computer algebra systems as the engine and implementation language (Mavrikis & Maciocia, 2003; Pacheco-Venegas et al., 2015; Sangwin, 2004). Over time, researchers have improved the functionalities of these tools by adding features such as the automation of personalized assignments and the parameterization of exercises for each student (Carlos et al., 2015. As an example, Goodle GMS (nowadays called Doctus) is a specific tool that brings all the abovementioned characteristics together (De la Peña et al., 2012). From an educational point of view, individualization and parameterization of assignments have a double benefit. First, providing individualized exercises to the students prevents them from copying from each other and other forms of cheating (plagiarism). Second, parameterization enables an instructor to implement self-learning techniques to enable students to continuously measure their knowledge base. Doctus focusses on the online grading management system. Thus, Doctus operates from the online evaluator, whereas *EduZinc* begins offline directly with the creation of learning products.

Table 1. Automatic assessment tools.

Tool (reference)	Supported programming languages	Work mode	Type of feedback
Algo+ (Bey et al., 2018)	C++	Web application	The feedback is given by the instructor according to the reference solution that is closer to the output submitted by the student.
Automata (https://www.aspiringminds.com/technology/automata)	C, C++, Java, Python	Online application	Grade according to the distance of the solution of the candidate with respect to the possible solution provided by the expert (based on machine learning)
BOSS (Joy et al., 2005)	C	Software package	Error messages
CAP (Sapena et al., 2013)	Java	Web application	List of errors of the evaluated task. Code solution. Statistical analysis of presented tasks.
Ceilidh (Benford et al., 1994)	C, Pascal, C++	Desktop application	Report of areas where marks have been lost and gained; exercise solution
CourseMarker (Higgins et al., 2003)	Java, C++	Standalone	Comments on code explanation
EPFL grader (Bey et al., 2018)	C++	Web application	Output on how the code performed in the unit test
IT VBE (Skūpas, 2013)	Pascal, C++	Plugin	Semi-automated testing
JAssess (Yusof et al., 2012)	Java	Moodle plugin	Semi-automated testing
JavaBrat (Patil, 2010)	Java, Scala	Moodle plugin	Based on test cases
Pythia (Combéfis & Paques, 2015)	Fortran, C++	Web application	Unit-testing grader specifically designed for education (intelligent feedback)
RoboLIFT (Allevato & Edwards, 2012)	Java	Standalone	Unit testing
VPL (Rodríguez-del Pino et al., 2012)	Ada, C, C++, Fortran, Java, Pascal, Prolog, SQL, Scheme	Moodle plugin	Unit testing
Web-CAT (Edwards & Perez-Quinones, 2008)	Java C++, Pascal	Web application	Output for the test cases

EduZinc *explained*

EduZinc is focussed on easy ways for instructors to create personalized learning material. While the content creator creates material, the tool provides the necessary information for its automated evaluation. The material is associated with the source code required for the solution of the corresponding exercise. Thus, the content can be focussed, one exercise at a time, in both the way the exercise will look and the way in which it will be evaluated. Figure 1 shows a basic flowchart with the main elements that constitute *EduZinc*. Each one of them will be explained in detail in this section. (The tool code and its implementation are available upon request.)

Roles: Student, instructor, and content creator

EduZinc differentiates between roles, depending on the use and purpose at different points in the course:

- **Content creators** (resource material creators) create the material prior to the beginning of the course. They have access to the templates for exercises and can compile

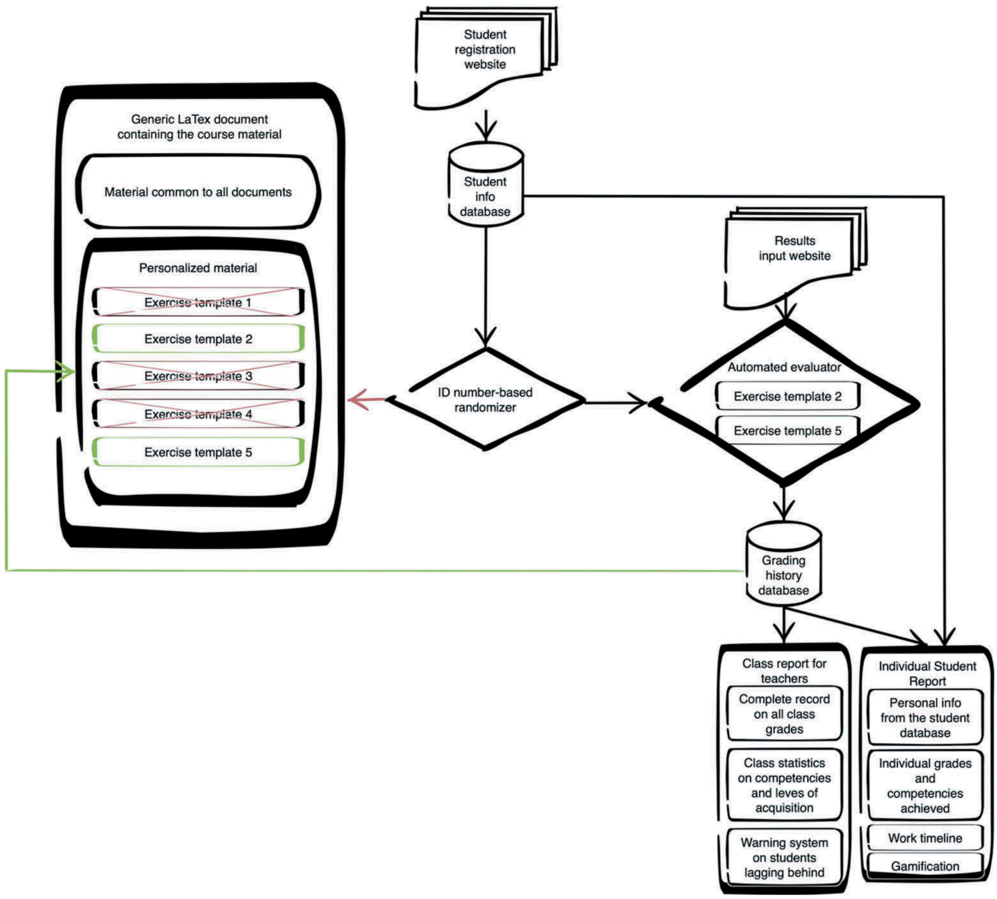


Figure 1. The *EduZinc* flowchart.

a complete Content Creator's Guide. This guide shows all exercises that can be assigned to a student. The content creator is also responsible for the assignment of the competencies associated to each template.

- **Students** register on the first day of class. They immediately receive the Student's Guide, which includes their personalized exercises. During the course, they submit their solutions to the exercises on the submissions website, and as a response to that, both their Student's Guide and their Student's Report are upgraded with grades and information that can facilitate further submissions.
- **Instructors** receive an Instructor's Guide with the solutions specific to each Student's Guide. It can be used to tutor students whenever they have questions about the exercises they have to solve. Instructors also receive a daily report on the entire class. This report provides individual information about the students and general statistics about the overall progress in competences.

We took inspiration from other research work for the design of the abovementioned roles. As suggested in distance education environments, there are some courses designed by experts (instructional designers) but taught by a different person (instructors); while in other courses, the same person is in charge of both tasks (Gómez-Rey, Barbera, et al., 2018). In *EduZinc*, tasks related to the role of content creators and instructors are distinct.

Registration website

To register as a student in the system, a website was set up to gather both generic and specific information (see Figure 2). Of all the inputs, an ID number (linked to the student's national identity card or foreign identity number) and an email address are compulsory for students to obtain the personalized course material.

Student's Guide

A generic LaTeX template creates a Student's Guide that has material common to all students as well as personalized material. This material is created using templates specifically designed to create real-time evaluation for the students. Some of these templates are pre-designed to create (for example) quizzes, numerical exercises, multiple-choice questions, open-ended questions, fill-in-the-blanks questions, and coding evaluation.

All students have a guide with the same number of exercises of a certain kind, chosen at random from a collection of templates. The code not only chooses the exercises to be shown but also randomizes the parameters of the exercises, according to the wishes of the content creator. Therefore, even when two students receive the same exercise, the parameters are unlikely to be the same.

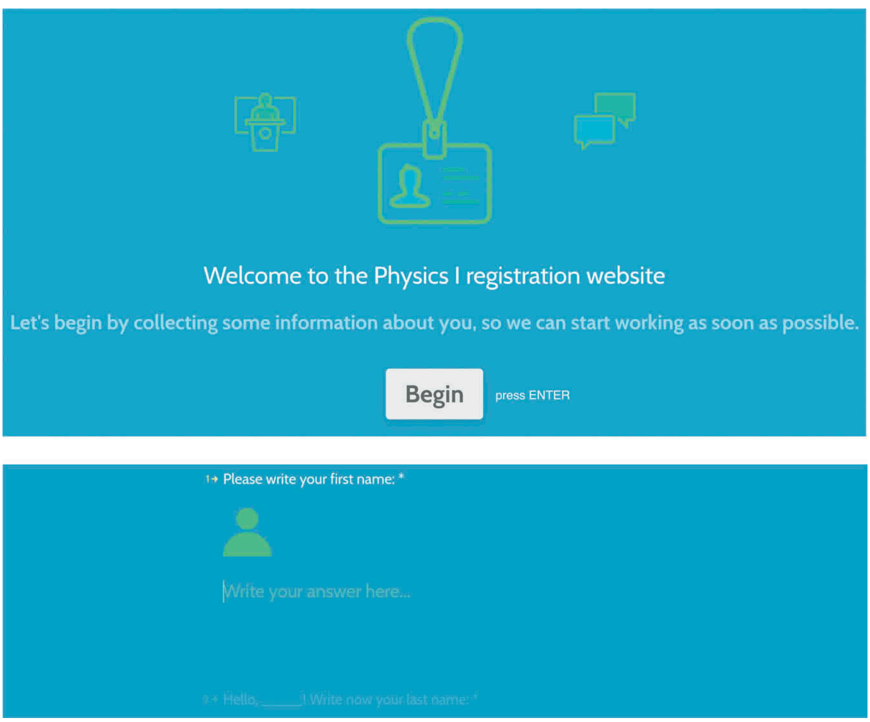


Figure 2. Student registration website.

Results input website

Every result on a specific exercise is submitted via a common website: the results input website (see Figure 3). This website enables the input needed for any of the templates created. Every template shown is identified with a unique hash. Thus, students input their ID and the hash of the specific exercise they intend to submit. Inputs include numerical results, text, formulae, and equations (in LaTeX), code, and multiple-choice answers.

An automated evaluator reads the results for a particular student and exercise and checks if they are correct. The corresponding grades and competencies are assigned to the student and stored in a grading history database. This database is used for two tasks: first, to update the Student's Guide, with information on each of the individualized exercises; second, to create the Student's Report and the Instructor's Report.

Type of exercises in the Student's Guide

EduZinc can generate two types of exercises: (a) exercises where students cannot submit multiple attempts (single-pass exercises) and (b) exercises where students can submit as

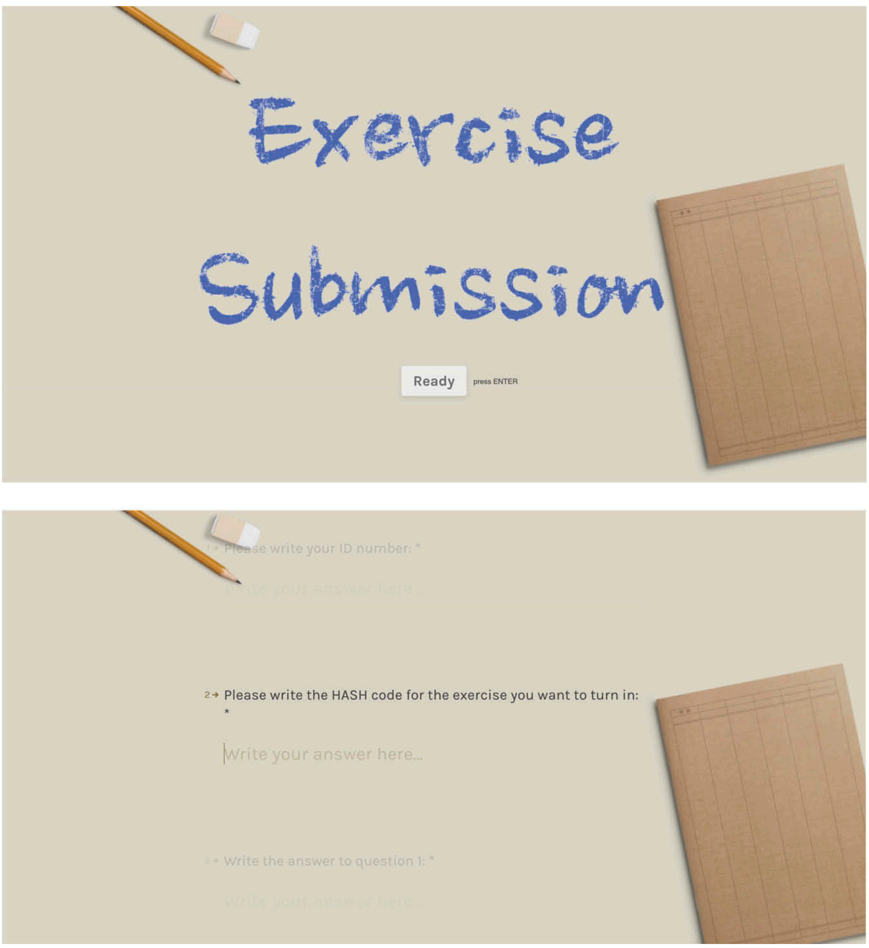


Figure 3. Results input website.

many times as they need or want to (multiple-pass exercises). Unlike other automatic grading tools, *EduZinc* offers both kinds, yet it proves particularly useful on multiple-pass exercises. These exercises can be used as a strategy to reinforce students' learning.

Instructor's Report

The Instructor's Report must include the academic record with the achievements and grades of all students. It must update as new information comes in, graphically showing the wave of submissions and results. This, of course, is a time saver on one of the most common jobs for an instructor: filling out the chart with all the grades.

Templates and exercises are associated with specific skills, competencies, and achievements. These must, in turn, feed into a general report that may guide the instructor whenever the class is lagging in particular skills. Often, even when a generic problem in the class is detected, the root competency affecting it is not easy to spot. Armed with this information, instructors can redirect the direction of their lessons.

The Instructor's Report also includes a section for automated warnings whenever a student stops participating in the course for a certain amount of time. The section also informs when certain students excel in their learning processes, enabling the instructor to make curricular adaptations to help students reach their maximum level of competency.

Student's Report

The Student's Report includes information related to a student's achievement (similar to a portfolio). Thus, it details the student's individual progress (student's marks) as well as the competencies already achieved or to be achieved. Text, images, and videos submitted by students are linked on the report as a means to prove their qualification. Additionally, every report includes an icon that identifies the student anonymously. Only each student knows the icon (randomly set by the system) shown in the report. Using those icons, a ranking of the best students in class is shown for the purposes of gamification. Figure 4 illustrates an anonymized example of a Student's Report.

Content creation

With *EduZinc*, the aim is to produce templates that cover most kinds of exercises. Following is an explanation of the most common automated template. Every exercise under this template requires the student to go to the results input website and type short text or numerical results. The template uses the following fill-in fields (bold indicates compulsory fields):

- **Exercise name**: All templates must have a unique name. From this name, hash codes, links, and QR codes are created.
- **Main exercise text**: In this field, the text of the exercise and the questions posed must be specified. A non-expert content creator can write the exercises in plain text. However, this field understands LaTeX code, thus enriching the compiled text. A number of LaTex and PythonTeX commands have been created for this field. These commands are intended to personalize parameters within the text.

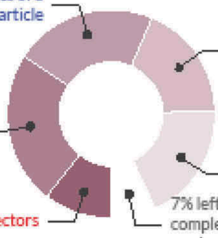
[STUDENT NAME HERE]

Electromechanical engineering student

Android studentemail@email.com Seville, Spain
ID: 12312312 eInstruction: 01

EDUZINC EXERCISES

IN-CLASS QUESTIONNAIRES



EDUZINC SUBMISSION TIMELINE

Date	Hash	Questions	Grade	Grade Today
8-1	72898f2	✓✗✓✓✓✗	6.66	6.90
8-1	a91ac0e	✓✗	5	5.17
6-1	a91ac0e	✓	5	5.25
3-1	91f0e82	✓✓	10	10.7
3-1	23736e5	✓✓	10	10.7
3-1	8be265b	✓	10	10.7
2-1	3354211	✓✓✗✓✓✓	8.33	8.98
2-1	91f0e82	✗✗	0	0
2-1	3354211	✓✓✗✓✓✓	8.33	8.98
2-1	c3ae7ef	✓	10	10.7
2-1	4998277	✓✗	5	5.39
28-12	4d81c4a	✓✓✓✓✓✓✓	10	11.1
28-12	4998277	✗✗	0	0
28-12	4d81c4a	✗✗✓✓✓✓✓	7.14	7.95
28-12	0d980cb	✓✓✓✓	10	11.1
28-12	c2d0d82	✓✓✓✓	10	11.1
28-12	46c1e90	✓✓✓✓✓✓	10	11.1
28-12	4998277	✗✗	0	0
26-12	4d81c4a	✗✗✓✓✓✓✓	7.14	8.06
26-12	0d980cb	✓✓✓✓	10	11.2
26-12	c2d0d82	✓✓✓✓	10	11.2

January 31, 2019

MY LIFE PHILOSOPHY

"Life is the art of drawing without an eraser."

COURSE NEWS

No warnings at this point
Two weeks to the final exam

Correctly registered
for the 2018 fall semester

STRENGTHS

Dreamer | Farsighted | Nerd

Mathematics | Python

SECOND LANGUAGE

English ●●●●●

TOP 10

Rank	Icon
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

Figure 4. Anonymized example of a Student's Report.

- **Parameters**: Within the text of the exercise, the content creator might be interested in diversifying words of values (potential parameters of the problem), thus creating an individualized exercise for those students having to solve it. These values are parameters within the text and must be specified in the "parameters" field. These words or values are randomly chosen for each student. In this way, each student has a different set of exercises.
- **Solution**: This is a text field, which uses LaTeX and benefits from the same LaTeX and PythonTex commands created for the main exercise text field. The intent for the "solution" field is to directly write the solution to the exercise without further explanations on how it is obtained.
- **Step-by-step solution**: Just like the "solution" field, this is a LaTeX field where, if useful, the content creator can detail the steps that lead to the solution. This field will be shown to students only when the exercise is used as an example of how these kinds of exercises are solved.
- **Solver**: This field includes, one by one, all the solutions to the questions posed in the Main exercise text. Each solution has a specific name. If a certain word or text is the answer, it simply goes alongside the solution's name. When the solution is numerical, the field uses the names of the parameters and the names of other solutions to calculate a specific solution. The calculations are made using the common arithmetic provided in Python; NumPy library is available if special functions (e.g., trigonometric, logarithmic) are needed.
- **Group**: Exercises are grouped according to any criteria useful to the syllabus. Exercises within the same group have the same deadlines.
- **Value of individual questions**: Every exercise may pose one or many questions. Some might be more difficult than others, and thus deserve a greater percentage of the overall value. By default, empty brackets [] are available to be filled out.
- **Value of the exercise within its group**: The default value of all exercises in a group is 1.
- **Competencies and levels of acquisition**: Every question of the exercise can be assigned to one or many levels of acquisition. These levels of acquisition belong to competences. This information will feed into both the Student's Report and the Instructor's Report.
- **Hints**: After several attempts at solving an exercise, a student can be helped with a hint. Content creators and instructors can add one or many hints per question.

An example of what an exercise looks like with this template as well as an additional example of an automated template that entails multiple-choice questions can be seen at http://pgomezrey.com/research/supplementary-material/.

Methodology

This section details the experimental framework used to highlight the educational advantages of the proposed application. To do this, we conducted a validation of the application from two perspectives: we considered students' learning performance over the time and we analyzed students' satisfaction with the process of teaching and learning in the last academic year. The instruments used, the participants and the data analysis carried out in this empirical study are described below.

Procedures

Two sources of data were used to assess the effectiveness of *EduZinc*: first, the academic marks obtained by students during 6 academic years (time period 2013–2019); second, an online questionnaire that was administered to the students at the end of the semester in the academic year 2018/19 (see the Results section). Specifically, the questionnaire was provided to students enrolled in the Degree in Electromechanical Engineering and the Degree in Industrial Organization Engineering at the Universidad Loyola Andalucía (Spain), during the fall semester of the 2018/19 academic year. The instructor of the course provided the questionnaire to the students in class in December 2018. Students were invited during the class to voluntarily fill out the questionnaire.

Participants

The participants were students from the Universidad Loyola Andalucía (Spain), a private face-to-face university with campuses in Seville and Córdoba, which was established in the 2013/14 academic year. It is a Catholic university run by the Society of Jesus. The university offers undergraduate as well as postgraduate programs (both in Spanish and English) in areas related to economics, business, social sciences, law, education, and engineering. Specifically, the participants in this study were enrolled in a subject called "Physics for Engineers." The sample used to validate students' learning performance over the time period of 6 academic years consisted of 300 students. Students were aged from 17 to 19 years. The distribution of students disaggregated by academic year is shown in Table 2. However, the sample used to analyse students' satisfaction with the process of teaching and learning were the 51 students enrolled in the academic year 2018/19. The overall response rate to the questionnaire was 64.70%, as 33 students ultimately participated in the study. Furthermore, the same course instructor was responsible for the subject during the period 2013–2019.

Data analysis

Statistical tests were applied to ascertain the significance of differences between students' marks over the 6-year period. The Shapiro-Wilk test was first used to evaluate whether the marks obtained by the students followed a normal distribution. In a second phase, the non-parametric Mann-Whitney-Wilcoxon test was conducted to check if a randomly selected student mark from the tool selected as the control method (*EduZinc*) was greater than a randomly selected mark from a student of the comparison tool (either Human Grader or Doctus). Additionally, a chi-square test was also performed to determine if there were statistical differences between the students who passed the course and those who did not, and between the students who passed with excellent grades and those who

Table 2. Students enrolled in "Physics for Engineers" during the period 2013–2019.

	Academic year					
	2013/14	2014/15	2015/16	2016/17	2017/18	2018/19
Males	39	36	58	44	44	46
Females	4	5	7	6	6	5
Total	43	41	65	50	50	51

passed but did not achieve excellent grades. Finally, a descriptive analysis was conducted based on students' satisfaction with regard to different factors related to *EduZinc*.

Results

Comparing the marks obtained by students during the period 2013–2019

We aimed to validate the proposed application, *EduZinc*, by comparing the marks that students obtained during 6 academic years (time period 2013–2019). It is worth mentioning that the course design, the course content, the syllabus and, ultimately, the degree of difficulty of the subject were similar over the period 2013–2019. However, three types of evaluation systems were carried out during these years. The first evaluator, called Human Grader, was used during 2013–2015; it was mainly the course instructor who evaluated the assignments and assigned a score during the entire process of continuous assessment. The second evaluator, called Doctus, was used during 2015–2017. Doctus is an automatic grading tool (see the Introduction and Literature review sections) and, to the best of our knowledge, it is the automatic grading system most like the one proposed here (De la Peña et al., 2012). Motivated by the limitations observed in Doctus, we created *EduZinc*, which was piloted during the period 2017–2019. Data of the academic years are grouped in elements of 2 years in order to reduce the variability of students' marks according to the tool implemented during this period (Table 3).

Table 3 (descriptive analysis) shows the mean, standard deviation, median, and the distribution of the students' marks, whereas Table 4 reports on the statistics and *p* values associated with the statistical test implemented. From a descriptive point of view, it can be seen that *EduZinc* obtained the best mean (5.40) and the best median (6.16). It is also important to highlight that *EduZinc* had the highest level of excellence (10 out of 101 students passed the course with excellent final grades).

Students evaluated with *EduZinc* (experimental group) achieved a significantly better median mark than those evaluated using the Human Grader system (Table 4). The median mark of the students evaluated using *EduZinc* was also greater than that for students evaluated with Doctus (although these differences were not significant according to the Mann-Whitney-Wilcoxon test, p value = 0.64).

Furthermore, we carried out two chi-square tests to analyse the differences in proportions. The first chi-square test enabled us to check if there were differences in the proportions of students who passed the course and the proportions of students who did not pass the course. The second chi-square test enabled us to detect if there were differences in the proportion of students who passed the course with excellence and the proportion of students who did not pass the course with excellence. In all cases, *EduZinc* achieved better results (proportionally) as all the chi-square values obtained in the comparisons were lower than 0.005. As can be seen in Table 4, the average of students' marks per academic year improved with the introduction and improvement of automatic tools proposed in the subject. Our hypothesis, based on our teaching experience with *EduZinc*, is that due to the daily evaluation and feedback, students are a lot more engaged in the subject. They dedicate more time to improve their results and are motivated to go further in the skills and competencies proposed to them (Steen-Utheim & Hopfenbeck, 2019).

Table 3. Statistical analysis of the marks obtained: mean, standard deviations, median, and the distribution of the students' marks.

Human Grader (2013–2015)			Doctus (2015–2017)			*EduZinc* (2017–2019)		
M	*SD*	*Mdn*	*M*	*SD*	*M*	*M*	*SD*	*Mdn*
3.05	2.07	2.55	4.49	2.90	5.12	5.40	2.77	6.16
Frequencies of students' scoresa								
63 (NP)				94 (NP)			44 (NP)	
21 (P)				21 (P)			47 (P)	
0 (PE)				0 (PE)			10 (PE)	

[a]NP (No pass), P (Pass), PE (Pass with excellence)

Table 4. *Statistical analysis of the marks obtained*: z *statistic and* p *value of the Mann-Whitney-Wilcoxon test and χ^2 statistic and* p *value of the χ^2 test (for the two cases explored).*

Mann-Whitney-Wilcoxon test		
Experimental group: *EduZinc* (2013–2015)	Human Grader (2013–2015)	Doctus (2015–2017)
z^2 statistic	0.97	0.46
p value	0.00*	0.64
χ^2-square test (No pass, Pass)		
Experimental group: *EduZinc* (2017–2019)	Human Grader (2013–2015)	Doctus (2015–2017)
χ^2 statistic	18.58	33.96
p value	0.00*	0.00*
χ^2 square test (No excellence, Excellence)		
Experimental group: *EduZinc* (2017–2019)	Human Grader (2013–2015)	Doctus (2015–2017)
χ^2 statistic	8.79	11.93
p value	0.00*	0.00*

*Significant differences were found for $\alpha = 0.05$.

Additionally, the competencies detailed on the Instructor's Report enabled the instructor to steer the emphasis toward those skills that are more difficult for the class as a whole. From there, the instructor can modify the sequencing of content, generate tutorials or supporting material, and identify those competencies that are more complex for students.

Analysing students' perceptions in the academic year 2018/19

To obtain insight into students' satisfaction toward the teaching and learning process, we conducted an exhaustive analysis on the questionnaire provided to the students. Students were asked to rank their satisfaction with regard to several factors involved in the educational process (elements related to *EduZinc*) according to a 10-item Likert scale (with 1 being *not at all satisfied* and 10 *extremely satisfied*). Specifically, four questions were examined in detail. The first question relates to the Student's Guide: "I believe that the Physics Student' Guide was useful for my learning." The second question concerns the Student's Report: "I felt the Student's Report was easy to understand and use." The third question has to do with the in-class questionnaires (self-assessment questionnaires) the students completed during their classes: "I liked the self-questionnaires made during the learning process as a part of my evaluation." The fourth question is about the instructor's role during the teaching-learning process: "The instructor was involved in the subject during the entire course." The statistical results obtained in this part are explained and interpreted in educational terms (Table 5).

Table 5. Online questionnaire—descriptive statistics results, with best result in boldface.

Student's evaluation	Student's Guide (%)	Student's Report (%)	In-class questionnaires (%)	Instructor's role (%)
[1–2]	0 (0%)	0 (0%)	0 (0%)	0 (0%)
[3–4]	1 (3.03%)	1 (3.03%)	0 (0%)	0 (0%)
[5–6]	3 (9.09%)	5 (9.09%)	1 (3.03%)	1 (3.03%)
[7–8]	**17 (51.52%)**	**14 (51.52%)**	13 (39.39%)	2 (6.06%)
[9–10]	12 (36.36%)	13 (36.36%)	**19 (57.58%)**	**30 (90.90%)**
Total	33 (100%)	33 (100%)	33 (100%)	33 (100%)

From a descriptive point of view, the abovementioned *EduZinc* elements met with a general satisfaction. Specifically, out of 10 points, 51.52% of the students scored the Student's Guide exercises and the Student's Report as either 7 or 8. Furthermore, 57.58% of the students scored the in-class questionnaire as either 9 or 10.

The Student's Guide is the all-in-one source of information for the course. Although it contains the information typically found in a textbook, it is a lot more than that as it changes daily. The adaptive content, the information on how to move through the course, the links to more materials, or the access to the input website, for instance—all of these features make it the main tool for the course. Yet for many students this was the first time they were exposed to this kind of system, and it took a slight adaptation in some cases. The general acceptance of the Student's Guide over the last few years is, however, positive. In future upgrades of *EduZinc*, the Student's Report will most likely be incorporated with the Student's Guide. Interaction with the Student's Report has shown it is more useful in context with the rest of the material. Nonetheless, students thrive as they see the daily reward for their work. In general, students were satisfied with the Student's Guide, specially stressing its ease of use. However, they have suggested they want more gamification in it.

On the other hand, the more remarkable results are related to the instructor's role and its implication in the tracking of students' learning. *EduZinc* promotes a more fluid communication between instructors and students, which improve students' engagement with the subject. The atmosphere created through the *EduZinc* system encourages students to ask for the tutoring sessions they need. The instructor, on the other hand, has not only early warnings on the state of the class but also is encouraged with the progress of good students, thriving on helping them on a more individual basis. This creates the feedback loop that eventually makes more than 90% of the class highly satisfied with the instructor's involvement.

An educational data mining analysis

The objective was to test several educational data mining (EDM) techniques in the context of the software presented. Thus, we created a classification model that relates the binary variable (Pass/No pass) with the different evaluation items employed during the course. The data used to test the EDM models were those generated by the *EduZinc* software (in the period 2017–2019).

The EDM models were extracted from the KNIME tool (Rangra & Bansal, 2014). The metric used to evaluate the goodness-of-fit of the different models was the correct classification rate. The EDM models tested were decision trees, naïve Bayes, neural networks, and support vector machines. The parameters of these models were cross-validated as suggested by Rangra and Bansal. The data were partitioned using a 10-fold cross-validation in order to test the robustness of the result generated.

The model with the best accuracy in the generalization set was the support vector machines, which achieved an overall accuracy of 86.7%. The good performance achieved by the model enabled us to interpret the variable that influenced the most in the classification. To that end, we used the global sensitivity analysis proposed by Fernández-Navarro et al. (2017).

The global sensitivity analysis enabled us to determine those activities with greater influence in the course. Thus, this analysis could help instructors to design new activities (not redundant to the previous ones) in content with more impact in the course. The exercises in Units 3 and 4 have an important impact on the explanation of the variable associated to the students' final grades, whereas the exercises in Unit 7 have less impact on the previously mentioned variable. These results are consistent with the perspective of the course instructor as exercises in Unit 7 were very long, and students had more difficulties solving them. This fact reduces significantly the discriminative power of the input variable linked to the exercises in Unit 7.

Discussion and conclusions

There are many motivations to use software that enables both the design of intelligent feedback and automated evaluation. The Bologna process has been understood as a more serious requirement that instructors be involved in the learning process of their students (Van der Wende, 2000). The common traditional model where teacher-centered education is followed by a final exam is no longer an option in many cases. This places the responsibility on instructors for creating a system that enables information to flow in both directions between them and students. The advent of technology is the opportunity we should seize if we are ever to meet Bologna's requirements.

A concrete example where we can take advantage of software is in summative assessment through automatic grading. There are many tools for automated grading available for instructors of all levels. Intelligent diagnostic feedback was introduced in the educational community a long time back and it is still a topic of interest for the research community (Gupta & Gupta, 2017; Hollingsworth, 1960; Page & Petersen, 1995; Sijimol & Varghese, 2018; Simsek et al., 2013). Some see these tools as the holy grail of independent learning, which changes the role of instructor to that of a guide on the side (Tomas & Jessop, 2019). This might be true in some areas, as many elaborate online courses have come to show. Then again, a good instructor-student learning interaction is yet to be superseded by any kind of automation.

These instruments are widely used not only in face-to-face environments but also in online scenarios. The reason can be seen in a shift from an information society to a knowledge society. This structural change in society has had important consequences for the educational system. For instance, education and professional preparation have moved from an industrial model to a model that requires continuous learning (Gallardo-Echenique et al., 2015). Thus, the ways of learning in the knowledge society have been significantly expanded and refocussed. An important example of these new ways of learning is open, flexible, and technology-enhanced learning environments (online education programs, in general, and massive open online courses, in particular) (Hew & Cheung, 2014). In this regard, the increasing number of students who are enrolled in this type of scenario justifies the development of such tools. Moreover, on the instructor's side, these electronic systems have facilitated both the creation of personalized course materials and students' evaluations (Anshari et al., 2016). Consequently, instructors can dedicate more time to other aspects of the teaching-learning process.

The advantages afforded to assessing and grading automatically are diverse. First, much of what students do online is saved, archived, and stored. This presents huge opportunities for instructors and students in the sense of tracking progress and generating sources of evidence of the student's learning. Second, the flexible nature of online learning can be leveraged to support diverse approaches to learning. Students have the opportunity to do their work when it suits them.

Motivated by these reasons, we developed *EduZinc* as an open and flexible automated grading application for both distance and face-to-face courses. The tool not only has a great flexibility in the creation of content but also provides intelligent feedback. This is particularly important in the context of distance education. Students drift; motivation and discipline are often hard to maintain (Gómez-Rey et al., 2016). To counter that, many tools offer online interaction. Some offer easy creation of personalized material. Some also include warnings for both instructors and students as the course moves on. Fewer have exercises that evolve as answers come in. Even fewer suggest new material to cover areas where the student needs additional support. Some have great flexibility to customize the rules of all the abovementioned features. Only a few go beyond grading to assessing the causes of better and worse learning processes, looking at competences in context. *EduZinc* has all these features and is ready to expand as new templates are proposed.

The traditional fixation on grading has removed many, when approaching technology, from the real potential: not automated grading, but automated evaluation. We now have the in-detail competencies that anticipate and steer the course toward those specific parts with which the class as a whole struggles the most. Individual assessment of skills and competencies is now also possible.

These are some of the lessons learned during the development and implementation of *EduZinc*. The first is that we do not have to be too ambitious in the creation of our material. *EduZinc* is easy to use in both the creation of learning materials and the daily interaction with students. The first simple achievement that *EduZinc* has brought is that, because of the daily feedback and follow-up, students who would typically do none of the exercises of the course (or wait until the last days before the due date) now do all or most of the compulsory ones. The *EduZinc* Student's Reports become the motivation for them not to lag behind, while the Instructor's Reports provide early warnings on students who need support.

This places students in a completely different frame of mind with regard to the course. First, they are motivated as they see progress and reward every day. Second, they learn enough basics to challenge themselves to tackle exam-level exercises and other more demanding course objectives. Third, they ask for tutoring sessions on a regular basis, making the individual interaction with the instructor a spontaneous process. Finally, although they are aware of their individual responsibility, they share their personalized exercises with their classmates to work together toward the solutions needed. If someone other than the student completes the work, it eventually shows in the abilities the student must show toward the end of the semester.

The results presented here are the testimony of a subject ("Physics for Engineers") where we showed a way to structure a course to increase students' engagement. It is the customization, the automation, and the provision of daily feedback that create this atmosphere and this degree of involvement from both instructors and students. This is an emergent behavior not foreseen merely by looking at the syllabus of the subject.

Study limitations

Although this study presents an automatic grading tool with certain advantages with respect to other automated summative assessment systems introduced in the literature, it is important to highlight that this research work in its current form has a potential limitation. Our study used two forms of evidence to assess the merits of *EduZinc*: the marks obtained by students during the period 2013–2019 and the students' perceptions in the academic year 2018/19. Taking into account that the proposed tool was recently created, the authors do not have a wide range of documentation to justify the extent to which the students' success (students' marks) corresponds to the introduction of the tool and the extent to which the students' success (students' marks) is related to either non-teaching or teaching-related input variables. Thus, motivated by the study of Gómez-Rey, Fernández-Navarro, et al. (2018), future research will present a classification model that relates the technology—*EduZinc*—and the abovementioned variables to the students' final mark. This kind of data will help us validate the importance of the proposal.

Notes on contributors

David Becerra-Alonso obtained his PhD from the School of Computing at the University of the West of Scotland in 2010, where he worked on dynamical chaotic systems. He currently holds a position as associate professor at the Universidad Loyola Andalucía. His research interests include dynamical systems and machine learning methods.

Isabel Lopez-Cobo holds a PhD in education from the Universidad de Córdoba and is a lecturer in education at the Universidad Loyola Andalucía, Spain. Her work focusses on improving the effectiveness of schools, information and communications technology, skill assessment of university professorate, and the creation of learning communities in disadvantaged environments.

Pilar Gómez-Rey holds an MSc in business administration and an MSc in education and information and communications technology. She received her PhD in education and information and communications technology in 2017. She is an assistant professor at the Universidad Loyola Andalucía, Spain. Her research is focussed on measuring quality in online and distance scenarios.

Francisco Fernández-Navarro received a PhD in computer science and artificial intelligence in 2011. He was a research fellow in computational management with the European Space Agency, The Netherlands, and is currently working as associate professor at the Universidad Loyola Andalucía, Spain.

Elena Barbera holds a PhD in educational psychology. She is a senior researcher and director of the PhD program "Education and ICT" at Universitat Oberta Catalunya and the head of the e-DUS research group (Distance School and University education). Her research is focussed on educational psychology and self-regulation in e-learning.

ORCID

Isabel Lopez-Cobo http://orcid.org/0000-0003-3579-0877
Pilar Gómez-Rey http://orcid.org/0000-0002-1883-4279
Francisco Fernández-Navarro http://orcid.org/0000-0002-5599-6170

References

Allevato, A., & Edwards, S. H. (2012, February 29 – March 3). RoboLIFT: Engaging cs2 students with testable, automatically evaluated android applications. In L. A. Smith King, D. R. Musicant,

T. Camp, & P. Tymann (Eds.), *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education* (pp. 547–552). Association for Computing Machinery. https://doi.org/10.1145/2157136.2157293

Al Shamsi, F., & Elnagar, A. (2012). An intelligent assessment tool for students' Java submissions in introductory programming courses. *Journal of Intelligent Learning Systems and Applications*, *4*(1), 59–69. https://doi.org/10.4236/jilsa.2012.41006

Anshari, M., Alas, Y., Sabtu, N. P. H., & Hamid, M. S. A. (2016). Online learning: Trends, issues and challenges in the Big Data Era. *Journal of e-Learning and Knowledge Society*, *12*(1), 121–134. https://doi.org/10.20368/1971-8829/1003

Balakrishnan, V., & Gan, C. L. (2016). Students' learning styles and their effects on the use of social media technology for learning. *Telematics and Informatics*, *33*(3), 808–821. https://doi.org/10.1016/j.tele.2015.12.004

Benford, S., Burke, E., Foxley, E., Gutteridge, N., & Zin, A. M. (1994). Ceilidh as a course management support system. *Journal of Educational Technology Systems*, *22*(3), 235–250. https://doi.org/10.2190/Y62D-CYEX-Q4D8-R8DQ

Bey, A., Jermann, P., & Dillenbourg, P. (2018). A comparison between two automatic assessment approaches for programming: An empirical study on MOOCs. *Journal of Educational Technology & Society*, *21*(2), 259–272. https://www.j-ets.net/collection/published-issues/21_2

Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, *41*(3), 400–413. https://doi.org/10.1080/02602938.2015.101813

Carlos, S., De la Peña, D. M., & Gomez-Estern, F. (2015). Virtual assistant for individualized practical training on controller design. *International Federation of Automatic Control (IFAC)-PapersOnLine*, *48*(29), 205–210. https://doi.org/10.1016/j.ifacol.2015.11.238

Combéfis, S., & Paques, A. (2015, July 14–17).Pythia reloaded: An intelligent unit testing-based code grader for education. In T. Xie, A. Orso, J. Bishop, & N. Tillmann (Eds.), *CHESE 2015: Proceedings of the 1st International Workshop on Code Hunt Workshop on Educational Software Engineering* (pp. 5–8). Association for Computing Machinery. https://doi.org/10.1145/2792404.2792407

De la Peña, D. M., Gómez-Estern, F., & Dormido, S. (2012). A new Internet tool for automatic evaluation in control systems and programming. *Computers & Education*, *59*(2), 535–550. https://doi.org/10.1016/j.compedu.2011.12.016

Edwards, S. H., & Perez-Quinones, M. A. (2008, June 30 – July 2). Web-cat: Automatically grading programming assignments. In J. Amillo, C. Laxer, E. M. Ruiz, A. Young (Eds.), *ITiCSE '08: Proceedings of the 13th Annual Conference on Innovation and Technology in Computer Science Education* (pp. 328–328). Association for Computing Machinery. https://doi.org/10.1145/1384271.1384371

Farrús, M., Costa-Jussà, M. R., Cobo, G., García, D., Villarejo, L., & Banchs, R. E. (2010, September 21–23). Automatic evaluation of continuous assessment tests. In S. Ishikawa, & G.R. Weir (Eds.), *Proceedings of the 6th ICTATLL International Conference* (pp. 173–180). ICTATLL. http://hdl.handle.net/10230/30932

Fernández-Navarro, F., Carbonero-Ruz, M., Becerra-Alonso, D., & Torres-Jiménez, M. (2017). Global sensitivity estimates for neural network classifiers. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(11), 2592–2604. https://doi.org/10.1109/TNNLS.2016.2598657

Gallardo-Echenique, E. E., de Oliveira, J. M., Marqués-Molias, L., Esteve-Mon, F., Wang, Y., & Baker, R. (2015). Digital competence in the knowledge society. *MERLOT Journal of Online Learning and Teaching*, *11*(1), 1–16. https://jolt.merlot.org/vol11no1/Gallardo-Echenique_0315.pdf

Gómez-Rey, P., Barbera, E., & Fernández-Navarro, F. (2016). Measuring teachers and learners' perceptions of the quality of their online learning experience. *Distance Education*, *37*(2), 146–163. https://doi.org/10.1080/01587919.2016.1184396

Gómez-Rey, P., Barbera, E., & Fernández-Navarro, F. (2018). Students' perceptions about online teaching effectiveness: A bottom-up approach for identifying online instructors' roles. *Australasian Journal of Educational Technology*, *34*(1), 116–131. https://doi.org/10.14742/ajet.3437

Gómez-Rey, P., Fernández-Navarro, F., Barbera, E., & Carbonero-Ruz, M. (2018). Understanding student evaluations of teaching in online learning. *Assessment & Evaluation in Higher Education*, *43*(8), 1–14. http://doi.org/10.1080/02602938.2018.1451483

Gupta, S., & Gupta, A. (2017, December 22–23).E-assessment tools for programming languages: A review. In A. Jaiswal, V. Kumar, Z. Lu, & N. Rajput (Eds.), *Proceedings of the First International*

Conference on Information Technology and Knowledge Management (Vol. 14, pp. 65–70). ICITKM. https://doi.org/10.15439/2017KM31
He, Y., Hui, S. C., & Quan, T. T. (2009). Automatic summary assessment for intelligent tutoring systems. *Computers & Education*, *53*(3), 890–899. http://doi.org/10.1016/j.compedu.2009.05.008
Hew, K. F., & Cheung, W. S. (2014). Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. *Educational Research Review*, *12*, 45–58. http://doi.org/10.1016/j.edurev.2014.05.001
Higgins, C., Hegazy, T., Symeonidis, P., & Tsintsifas, A. (2003). The coursemarker cba system: Improvements over ceilidh. *Education and Information Technologies*, *8*(3), 287–304. http://doi.org/10.1023/A:1026364126982
Hollingsworth, J. (1960, October). Automatic graders for programming classes. *Communications of the ACM*, *3*(10), 528–529. https://doi.org/10.1145/367415.367422
Ion, G., Sánchez Martí, A., & Agud Morell, I. (2019). Giving or receiving feedback: which is more beneficial to students' learning? *Assessment & Evaluation in Higher Education*, *44*(1), 124–138. http://doi.org/10.1080/02602938.2018.148488
Joy, M., Griffiths, N., & Boyatt, R. (2005). The boss online submission and assessment system. *Journal on Educational Resources in Computing (JERIC)*, *5*(3), 1–27. http://doi.org/10.1145/1163405.1163407
Keuning, H., Jeuring, J., & Heeren, B. (2016, July 11–13).Towards a systematic review of automated feedback generation for programming exercises. In *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education* (pp 41–46). Association for Computing Machinery. http://doi.org/10.1145/2899415.2899422
Mavrikis, M., & Maciocia, A. (2003). Wallis: A web-based file for science and engineering students studying mathematics. In M. Seppala, S. Xambo, & O. Capriotti (Eds.), *Workshop of Advanced Technology for Mathematics Education in the 11th International Conference on Artificial Intelligence in Education* (pp. 505–512). Springer.
Milligan, C., & Littlejohn, A. (2017). Why study on a MOOC? The motives of students and professionals. *The International Review of Research in Open and Distributed Learning*, *18*(2), 92–102. https://doi.org/10.19173/irrodl.v18i2.3033
Pacheco-Venegas, N. D., López, G., & Andrade-Aréchiga, M. (2015). Conceptualization, development and implementation of a web-based system for automatic evaluation of mathematical expressions. *Computers & Education*, *88*, 15–28. https://doi.org/10.1016/j.compedu.2015.03.021
Page, E. B., & Petersen, N. S. (1995). The computer moves into essay grading: Updating the ancient test. *Phi Delta Kappan*, *76*(7), 561–566. https://www.questia.com/library/journal/1G1-16765769/the-computer-moves-into-essay-grading-updating-the
Paiva, R. C., Ferreira, M., & Frade, M. (2017). Intelligent tutorial system based on personalized system of instruction to teach or remind mathematical concepts. *Journal of Computer Assisted Learning*, *33*(4), 370–381. https://doi.org/10.1111/jcal.12186
Patil, A. (2010). *Automatic grading of programming assignments* [Master's thesis, San José State University]. https://doi.org/10.31979/etd.vnt7-hgnd
Pellegrino, J. W. (2017). Teaching, learning and assessing 21st century skills. In S. Guerriero (Ed.), *Pedagogical knowledge and the changing nature of the teaching profession* (pp. 223–251). OECD Publishing. https://doi.org/10.1787/9789264270695-en
Pepper, D. (2011). Assessing key competences across the curriculum—and Europe. *European Journal of Education*, *46*(3), 335–353. https://doi.org/10.1111/j.1465-3435.2011.01484.x
Pohuba, D., Dulik, T., & Janku, P. (2014, May 14–16).Automatic evaluation of correctness and originality of source codes. In *Proceedings of the 10th European Workshop on Microelectronics Education* (pp. 49–52). IEEE. http://doi.org/10.1109/EWME.2014.6877393
Rangra, K., & Bansal, K. L. (2014). Comparative study of data mining tools. *International Journal of Advanced Research in Computer Science and Software Engineering*, *4*(6), 216–223. http://www.iosrjournals.org/iosr-jce/papers/Vol19-issue4/Version-3/J1904035661.pdf
Rodríguez-del Pino, J. C., Rubio Royo, E., & Hernández Figueroa, Z. (2012, July 16–19).A virtual programming lab for Moodle with automatic assessment and anti-plagiarism features. In H. R. Arabnia, A. Bahrami, J. Cabiria, & M. Neal (Eds.), *EEE'12: Proceedings of the 2012 International conference on e-Learning, e-Business, Enterprise Information Systems, e-Government* (pp. 1–6). CSREA

Press. https://www.semanticscholar.org/paper/A-Virtual-Programming-Lab-for-Moodle-with-automatic-Rodríguez-del-Pino-Royo/a818e121ad140fcb323cb269c54b84d7cd70b354

Sangwin, C. (2004). Assessing mathematics automatically using computer algebra and the internet. *Teaching Mathematics and Its Applications: An International Journal of the IMA*, *23*(1), 1–14. http://doi.org/10.1093/teamat/23.1.1

Sapena, Ó., Galiano, M., Prieto, N., & Llorens, M. (2013, July). Aprender, enseñar y evaluar con CAP, un corrector automático de tareas de programación [Learn, teach and evaluate with CAP, an automatic corrector of programming tasks]. In S. B. M. Marqués Andrés, J. M. Badía Contelles, & S. Barrachina Mir (Eds.), *Proceedings of the XIX Jornadas sobre la Enseñanza Universitaria de la Informática* (pp. 343–350). Publicacions de la Universitat Jaume I. Servei de Comunicació i Publicacions. http://jenui2013.uji.es/file/actas_jenui2013.pdf

Sijimol, P., & Varghese, S. M. (2018). Handwritten short answer evaluation system (HSAES). *International Journal of Scientific Research in Science and Technology*, *4*(2), 1514–1518. http://ijsrst.com/paper/2702.pdf

Simsek, D., Buckingham Shum, S., Sandor, A., De Liddo, A., & Ferguson, R. (2013, April 8–12).XIP Dashboard: Visual analytics from automated rhetorical parsing of scientific metadiscourse. In D. Suthers, K. Verbert, E. Duval, & X. Ochoa (Eds.), *1st International Workshop on Discourse-Centric Learning Analytics* (pp.1–10). Society for Learning Analytics Research. https://www.solaresearch.org/events/lak/lak13/dcla13/

Skūpas, B. (2013). *A method for semi-automatic evaluation and testing of programming assignments* [Unpublished doctoral dissertation]. Vilnius University.

Srikant, S., & Aggarwal, V. (2014, August 24–27).A system to grade computer programming skills using machine learning. In S. Macskassy & C. Perlich (Eds.), *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1887–1896). Association for Computing Machinery. https://doi.org/10.1145/2623330.2623377

Steen-Utheim, A., & Hopfenbeck, T. N. (2019). To do or not to do with feedback: A study of undergraduate students' engagement and use of feedback within a portfolio assessment design. *Assessment & Evaluation in Higher Education*, *44*(1), 80–96. https://doi.org/10.1080/02602938.2018.1476669

Suleman, H. (2008, October 6–8).Automatic marking with Sakai. In R. Botha & C. Cilliers (Eds.), *Proceedings of the 2008 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on it research in Developing Countries: Riding the wave of technology* (pp. 229–236). Association for Computing Machinery. https://doi.org/10.1145/1456659.1456686

Tomas, C., & Jessop, T. (2019). Struggling and juggling: A comparison of student assessment loads across research and teaching-intensive universities. *Assessment & Evaluation in Higher Education*, *44*(1), 1–10. https://doi.org/10.1080/02602938.2018.1463355

Van der Wende, M. C. (2000). The Bologna Declaration: Enhancing the transparency and competitiveness of European higher education. *Journal of Studies in International Education*, *4*(2), 3–10. https://doi.org/10.1080/713669277

Yusof, N., Zin, N. A. M., & Adnan, N. S. (2012). Java Programming Assessment Tool for Assignment Module in Moodle e-learning system. *Procedia – Social and Behavioral Sciences*, *56*, 767–773. https://doi.org/10.1016/j.sbspro.2012.09.714