IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 24, 2020, accepted February 7, 2020, date of publication February 10, 2020, date of current version February 18, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2973026

A Framework for Automated Formative Assessment in Mathematics Courses

JULIÁN MORENO AND ANDRÉS F. PINEDA
Department of Computer and Decision Sciences, National University of Colombia at Medellín, Medellín 050034, Colombia
Corresponding author: Andrés F. Pineda (afpinedac@unal.edu.co)

This work was supported by the Administrative Department of Science, Technology and Innovation (COLCIENCIAS), Colombia.

ABSTRACT In general terms, the aims of formative assessment are to gather and analyze information about the progress of students, with the intention of improving instruction in real time. This can also help teachers to identify those elements that students are struggling with, and therefore to determine where adjustments must be made. There are, however, some issues that must be addressed before this approach can become more widespread. In this research, and with mathematics as our specific scope, we focus on two of these issues: we present a framework to help teachers implement assessment items that are highly dynamic and flexible, but that allows them at the same time to model common misconceptions and to provide suitable feedback. Following this, we carried out an experiment to determine (i) whether the students perceived such feedback to be useful; (ii) whether they actually used it; and ultimately (iii) whether it produced a difference in their performance. To find the answers to these questions, we used an online 'pre-calculus' course with 458 students and collected data directly from those students via both a perception survey and from their interaction with the LMS. Based on these data, we performed several analyses, including process mining and statistical hypothesis tests. The results demonstrated that although many of the students had a positive opinion of the feedback received, it was not always followed. Nevertheless, when it was (or at least when students said it was), the students performed better.

INDEX TERMS Formative assessment, learning analytics, immediate feedback, automatic assessment, self-regulated learning, learning strategies, e-learning.

I. INTRODUCTION

Learning mathematics is typically challenging for most students [1], [2]. In fact, there is a great deal of evidence for the low performance of students throughout these courses at all educational levels, from primary school to college [3]–[8].

On the teaching side, several authors remark that the use of traditional approaches in face-to-face environments imply serious challenges for teachers [9]–[11]. They mention that there is a major problem with feeling overwhelmed when dealing with many students at once, having to grade each student's progress but also solving each individual question. In many cases, this process needs to be done quickly, and teachers try to shorten the explanations with the aim of supporting all students as fairly as possible. According to Radmehr and Drake, this procedure promotes passive learning and presents mathematics as a known field of knowledge that students simply need to assimilate, instead of being a dynamic subject with errors and false trails that plays an important role in the development of new knowledge [5]. This statement is confirmed by Rakoczy et al., who remark that such passive teaching methods also produce lower pass rates than those that engage students in active learning [12]. They also mention that in this case, grades provide little information about the relationship between actual performance and the learning goal.

It is not surprising then that many researchers and educators are using computers and other devices to perform assessments inside and outside the classroom. According to Pacheco-Venegas et al., using such technology in this scenario changes the way student learning is conceived [13]. In particular, it transforms from being characterized as a process of transmission of knowledge from teachers to students to being a process in which students construct their own knowledge, accessing resources and activities at their own

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

pace and according to their particular needs. More importantly, however, they are not forced to wait for the teacher's feedback after submitting an assignment, which often represents an obstacle to the independence of the students in terms of organizing their own learning. Other authors also remark that this approach has other advantages such as reducing administration costs, increasing interactivity and interaction logs, and improving the interfaces [14], [15].

However, in order to be successful, computer-supported assessment must be carried out as part of instructional design. One of the most widely accepted approaches is formative assessment (FA), which can be defined as ''all those activities undertaken by teachers, and by their students in assessing themselves, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged'' [16].

Several authors claim that FA has several benefits for both teachers and students [6], [17], [18]. For teachers, FA provides them with information about the effectiveness of their teaching strategies and allows them to be adjusted, even in real-time. In this case, the authors remark that teachers must know how to respond to the information obtained through assessment, and respond effectively to the students' needs, for example by identifying the most complicated topics or problems. For students, on the other hand, FA provides feedback about their progress in order to help them determine how to close any gaps between their performance and the targeted learning goals. FA also helps them to identify areas to which they need to devote effort, and whether they need to adjust their thinking in any way.

In both cases, it becomes clear that the main element in FA is appropriate feedback, which is defined by Núñez-Peña et al. as ''the information with which a learner can confirm, add to, overwrite, tune, or restructure information in memory, whether that information is domain knowledge, meta-cognitive-knowledge, beliefs about self and tasks, or cognitive tactics and strategies'' [19]. According to Hattie and Timperley, feedback can be provided either at task level (information on task performance), process level (information on processes required to master the task), the self-regulatory level (information on the regulation of the action), and the self level (information on the learner as a person, which is not related to task performance) [20]. Isabwe et al. add that in order for feedback to be effective, students must be convinced that they are likely to succeed if they use it [21]. In fact, according to Pinger et al., there is broad consensus that the effects on students' learning are not achieved automatically, but that students must understand, accept, and actively process the information through such feedback [22]. They add that the characteristics of feedback as well as contextual factors influence the way learners process the information received, and therefore the direction and magnitude of the feedback effects.

Nevertheless, several authors agree that designing effective FA in classrooms can be very difficult and time-consuming for teachers [7], [12], [22]. For this reason, in this paper we propose a framework called Nigma to help teachers to implement, at least preliminary, FA in mathematics courses. For ''preliminary'', we mean that our proposal focuses only in the automated feedback for common misconceptions with the hope that such information become useful for students as well as for teachers.

There are several research studies that propose models and tools with a similar approach in mathematics, such as ViLLE [10], MathDIP [13], Numbas [23], [24], MathCAL [25], MapleTA. [13], [26]–[29], STACK [30], [31] (which is a Moodle plugin), WeBWorK [32]–[34], WebMathematica [35] and Petite [36]. However, this paper is differentiated from these in three respects.

Firstly, the proposed framework includes a formal language for defining any number of random and interrelated parameters for use within the assessment items, in the form of text, charts, equations, or any combination of these. This feature enables the creation of highly flexible and dynamic assessment items in mathematics courses that not only empower the feedback modeled by teachers, but that also help to address some other important issues related to computer-supported assessment, such as plagiarism.

Secondly, Nigma is embedded into a learning management system (LMS) called TICademia, so all data about the students' interaction with the assessment items can be collected and analyzed, not only for the sake of the students themselves but also for the teacher.

Thirdly, we present a validation of the framework in a real educational context, in order to determine the students' response to the feedback provided by the assessment items. To achieve this, we perform an experimental study to answer the following research questions:

RQ1: Do the students follow the feedback they receive?

RQ2: Do the students perceive such feedback as useful?

RQ3: Is there any difference between the performance of the students who follow the feedback and those who don't?

The remainder of this paper is organized as follows: Section II describes the framework in detail, while Section III presents the results of validation and a discussion. Finally, Section IV concludes the paper and describes some directions for future research.

II. FRAMEWORK

As previously mentioned, the FA must be an integral part of the instructional design, and should be reflected not only in the assessment items but also in the structure of the courses. Embracing this principle, the course model used here is presented in Fig. 1. This structure is simple enough to define a course as a set of modules that each contain linearized resources and corresponding assessment items, but also enable us to adopt what Eddy et al. identify as some of the essential components that must be present in the classroom for FA to be effective: clear learning goals (defined per module), appropriate learning tasks (defined as a subsequence of resources and corresponding assessment items), and assessment followed by the appropriate feedback [37].

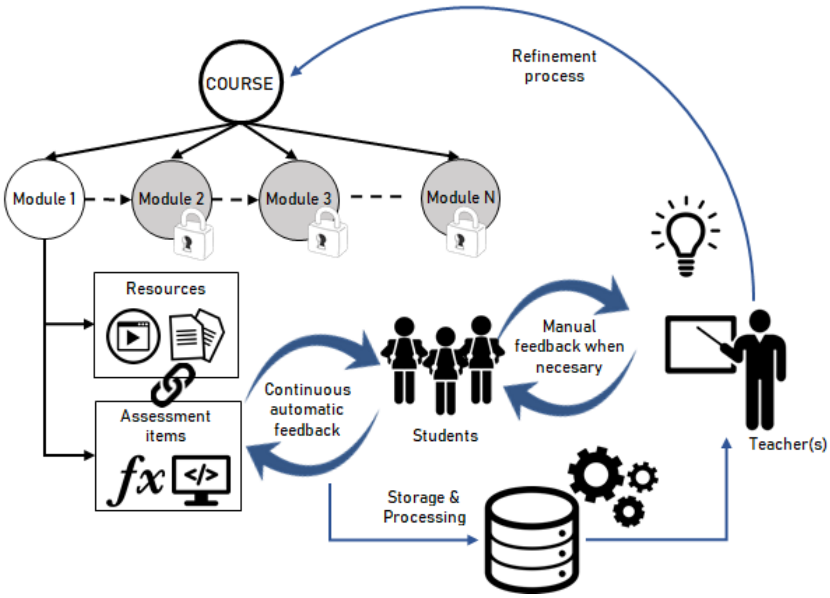


FIGURE 1. Course model.

```
<parameters>
    ?pop1 = C(Bogotá, Lima, Buenos Aires, Caracas)
    ?pop2 = C(Medellín, Quito, Córdoba)
    ?pop3 = C(Cartagena, Cusco, Mendoza, Asunción)
    ?x = U(6, 10, 0.1)
    ?y = U(2, 4, 0.1)
    ?n = E(2, 4, 5, 6, 8)
    ?z = ?x/?n
</parameters>
```

FIGURE 2. Example of the definition of parameters.

In this model, and therefore in the framework, the core of the FA is the assessment items. In order to provide immediate feedback, Nigma only considers types of questions with automatic responses, particularly multiple choice and completion. In both cases, they are composed of three parts: parameters, the question body, and answers.

A. PARAMETERS

An assessment item may contain zero or more parameters, which are random values used in the question body. They can be either numerical or categorical, and in the latter case may or may not be interrelated. For both categorical and numerical parameters, Nigma considers a uniform distribution; the possible numerical values can come either from a discrete range with a certain step size or from a specific dataset.

Fig. 2 shows an example in an XML format of seven parameters: *?pop1*, *?pop2* and *?pop3* are categorical (C), while ?x and ?y are numerical, within a specific range (U), ?n is numerical, from a specific dataset (E), and ?z is a calculated parameter.

B. QUESTION BODY

As its name implies, this refers to the body of the question per se, i.e. the text and other elements that will be presented to the students. In Nigma, this is defined using an embedded

```
</formulation>
    <chart>
        <type>piechart</type>
        <data>
            [[?pop1,?x], [?pop2,?y], [?pop3,?z]]
        </data>
        <label>"Population (in millions)"</label>
    </chart>
    </br>According to the figure, which of the
    following statements is true?
</formulation>
```

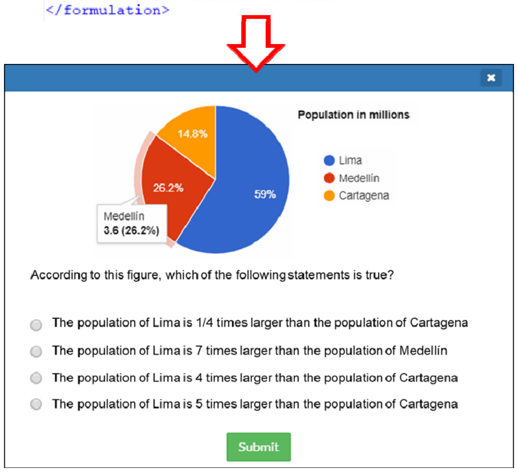


FIGURE 3. Example of a question body and its corresponding presentation.

CKeditor[1] library, allowing teachers to format it in the same way as in a word processor.

Within this body, the parameters can be used not only in the text but also to include dynamic charts and dynamic equations. Charts are created using the Google Charts[2] library, so they are not only graphical but interactive. Fig. 3 shows an example in XML format of a question body using the parameters presented in Fig. 2. An instance of what a student sees

[1]https://ckeditor.com/
[2]https://developers.google.com/chart/

```
<answers>
    <correct>
        The population of ?pop1 is ?n times larger
        than the population of ?pop3
    </correct>
    <incorrect>
        <option>The population of ?pop1 is 1/?n times
        larger than the population of ?pop3</option>
        <feedback>Look that ?z is 1/?n smaller than ?x,
        instead of ?x being 1/?n larger than ?z</feedback>
        <link>VID1275 t=0:8:35</link>
    </incorrect>
    </incorrect>
        <option>The population of ?pop1 is #(?n+1) times
        larger than the population of ?pop3</option>
    </incorrect>
    <incorrect>
        <option>The population of ?pop1 is 6 times
        larger than the population of ?pop2</option>
    </incorrect>
</answers>
```

FIGURE 4. Example of the definition of answers.

is presented also in the bottom part. Notice that in this case, when a student focuses over a pie piece, it gets highlighted and shows additional data.

C. ANSWERS

In order to provide proper feedback, teachers must model the correct answer for each assessment item, and also as many common misconceptions as they are aware of. Depending on the type of question (multiple-choice or completion), these misconceptions may or may not be presented as options. Fig. 4 shows an example in XML format of four answers: the correct answer, one misconception, and two misleading answers.

Each misconception must have a corresponding feedback message for the student, and may also include links to resources which may be referenced in general or specify a specific page, slide, time range, etc., depending on the resource format. For example, in the misconception presented in Fig. 4 (the first incorrect block), the feedback message contains a link to a video starting at a specific time.

But that feedback is not only useful for students. As shown in Figure 1, there is also a cycle on the teacher's side. Since all misconceptions' information can be collected, the teacher can be warned about recurring problems and thus take corrective actions in real time. For example, if a significant percentage of students are having the same misconception in an assessment item, and after checking the corresponding educational resource, the misconception still persists, the teacher should review that resource to determine if it contains some mistake or if the didactic used to present such a resource must be adapted.

Another example of the three parts of an assessment item and a corresponding instantiation for its presentation to students are presented in Fig. 5. This example corresponds to a completion question, and its formulation does not contain a chart but a dynamic equation using random parameters. Equations are created using the MathJax[3] library, meaning that teachers can use LaTex syntax directly for their definition.

In this case, all parameters are numeric: $?a$ is defined with a specific dataset, whereas $?b$, $?c$ and $?d$ have a specific range, and $?e$ is calculated. Notice that $?b$ is random but is also related to $?a$.

It is important to mention that in completion questions a number of significant digits must be defined in order to

[3]https://www.mathjax.org

```
<parameters>
    ?a = E(-4, -3, -2, -1, 1, 2)
    ?b = U(?a+5, ?a+8, 1)
    ?c = U(3, 5, 1)
    ?d = U(2, 5, 1)
    ?e = ?d(?c+1)
</parameters>
</formulation>
    What is the value of the following integral:</br>
    <equation>\int_{?a}^{?b} ?e x^{?c} \partial x</equation>
</formulation>
<answers>
    <correct>?d(pow(?b,?c+1)-pow(?a,?c+1))</correct>
    <incorrect>
        <option>?d(pow(?b,?c+1)-pow(?a,?c+1))</option>
        <feedback>
            You have to evaluate the integral in the
            upper limit first and then, substract
            the lower limit
        </feedback>
    </incorrect>
    <incorrect>
        <option>?d(pow(?b,?c+1)-pow(?a,?c+1))/(?c+1)</option>
        <feedback>
            When integrating in a polynomial, you raise the
            power by a unit, then you must divide the entire
            result by that same value
        </feedback>
    </incorrect>
</answers>
```

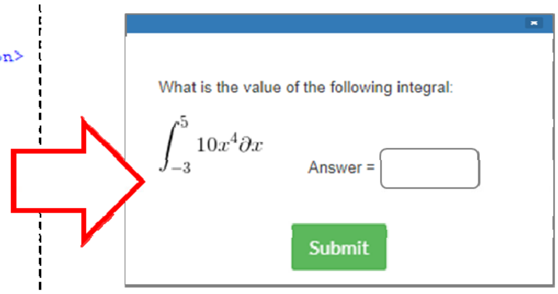


FIGURE 5. Example of a complete definition of an assessment item and its presentation.

TABLE 1. Recorded events.

Interaction	Description	Occurrences
ATTEMPT (AT)	The student answered a new assessment item (i.e. the identification of this item differs from the one for the last attempt)	38,107
CORRECT_RESPONSE (CR)	The student gave a correct response	37,385
INCORRECT_RESPONSE (IR)	The student gave an incorrect response	26,408
VIEW_SUGGESTED_RESOURCE (VR)	The student viewed a suggested resource after feedback	6,968
TRIAL_ERROR_ATTEMPT (TE)	The student submitted a new answer in less than 10 seconds since the last submission	1,498
SAME_INCORRECT_RESPONSE (SI)	The student gave the same incorrect response	3,049
SESSION_ENDED	The attempt ended	38,107

compare the numerical answer of the student with regard to the correct answer modeled by the teacher or to the misconceptions. In the example presented in Fig. 5 there are two misconceptions modeled.

III. VALIDATION

To answer the three research questions, we conducted an experiment using a 'pre-calculus' online course at the National University of Colombia. There were $N = 458$ enrolled students (247 males, and 211 females, aged between 16 and 23 years). The experiment was carried out over a four-week period, with two course modules of two weeks each. There were 48 assessment items in total and 36 resources, all of them designed by a group of four professors who were previously trained on the Nigma model. At the end of the fourth week, the students carried out a general test of the two modules and answered a perception survey.

In order to answer RQ1, and to determine whether or not the students followed the feedback received, we focused on the observable behavior of the students and specifically on the data collected by the LMS regarding the interactions with all assessment items and the corresponding suggested resources. A summary of the event logs that were collected is presented in Table 1.

Using such data, we used a process mining approach with Disco®software to obtain the process mining diagram presented in Fig. 6. This diagram shows the flow of the students' actions, as a total aggregate, each time they face an assessment item. We then took each of these flows and summarized them in the interaction matrix, as presented in Table 2. The acronyms used in the rows and columns headers correspond to those presented in table 1. We drew evidence from this data as to whether or not the students followed the feedback offered.

As a first finding, it seems that in at least 16.48% of the attempts, the feedback was not followed, since incorrect responses were followed by either a trial-and-error response (5.67%) or the same incorrect response (10.8%). In the first case, the most probable explanation is that in a multiple-choice question, students simply tried it again, selecting an option at random. Nevertheless, this percentage was low, possibly because the answer options are arranged randomly each time in the Nigma model, but most likely because the random parameters may significantly change the formulation of the assessment item. In the second case, the most probable explanation is that for a completion question, even with different parameters, the students followed the same procedure, thereby leading them to the same error.

TABLE 2. Interaction matrix.

	CR	IR	VR	TE	SI	SE
AT	65.64%	34.36%	-	-	-	-
IR	26.15%	40.16%	15.67%	5.67%	10.80%	1.54%
VR	27.81%	38.88%	29.55%	-	2.81%	0.95%
TE	87.58%	-	7.88%	-	-	4.54%
SI	72.61%	-	21.42%	-	-	5.97%

As a second finding, it seems that at least in 15.67% of the attempts, the feedback was followed, since incorrect responses resulted in viewing the suggested resources. However, this resulted in a correct response only 27.81% of the time.

Beyond these two scenarios, there was a large "gray" area in which there was no way to determine whether or not the students modified their procedure. In 26.15% of cases, incorrect responses were followed by correct ones. It is unlikely that students just guessed the right answer, but there is no way to determine whether they followed the feedback received. On the other hand, 40.16% of the time, incorrect responses were followed by further failures. In these cases, there is no way to determine whether students did not follow the feedback, or whether they did, corrected what was wrong but then made an additional mistake.

In order to answer RQ2, and to determine whether or not the students perceived the received feedback as useful, we focused on the perception survey conducted at the end of the fourth week, the questions for which are presented in Table 3. The first perception question (PQ1) tried to determine whether students found the feedback appropriate, while the second (PQ2) checked whether they actually followed it. to some extent, and presuming the honesty of the students, PQ2 allowed us to determine the students' behavior in the gray areas identified in the analysis of RQ1.

The results of the survey are summarized in Fig. 7, and demonstrate two facts. Firstly, there was a correspondence between the responses obtained in PQ1 and PQ2, i.e. students generally followed the feedback at almost the same rate at which they found it appropriate. Secondly, there were more students "for" than "against" feedback: 49.48% of the students gave a score of 4 or 5 in PQ1, as against 26.3% who replied 1 or 2, while 47.18% of the students replied 4 or 5 in PQ2 as against 30.07% who replied 1 or 2.

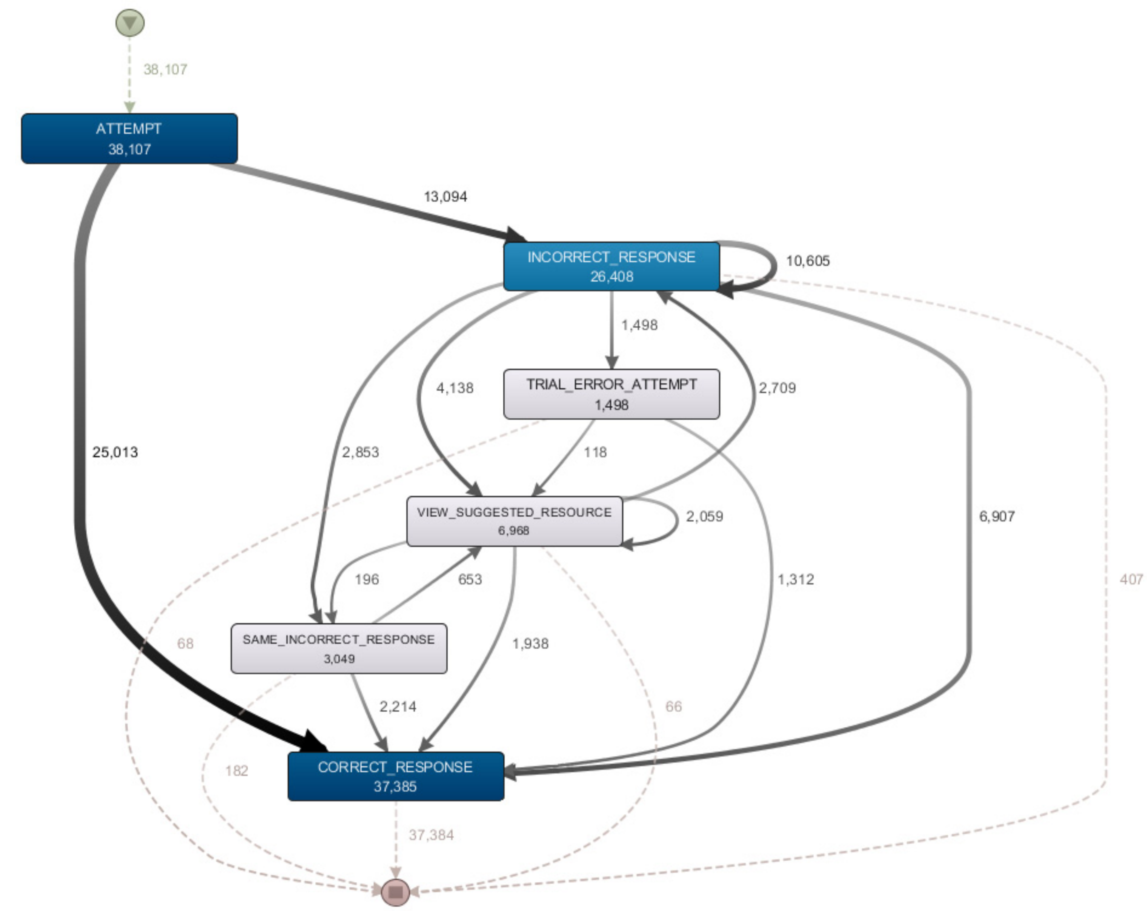


FIGURE 6. Process mining diagram.

TABLE 3. Perception survey applied to students.

Question	Formulation
PQ1	In general, was the feedback correct when you gave a wrong answer to exercise; that is, did it identify the problem you were having? 1: Never 2: Almost never 3: Sometimes 4: Almost always 5: Always
PQ2	When receiving feedback, did you reconsider your answer accordingly or view the suggested resource(s) in order to understand what you did wrong? 1: I always ignored the message and tried the exercise again in my own way 2: I almost always ignored the message and tried the exercise again in my own way 3: I sometimes followed the message and sometimes ignored it 4: I almost always followed the message, including viewing the suggested material when present, to rethink my answer 5: I always followed the message, including viewing the suggested material when present, to rethink my answer

Finally, in order to answer RQ3, and to determine whether there was a difference in the performance of the students who allegedly followed the feedback and those who did not, we compared the results of the perception survey, particularly

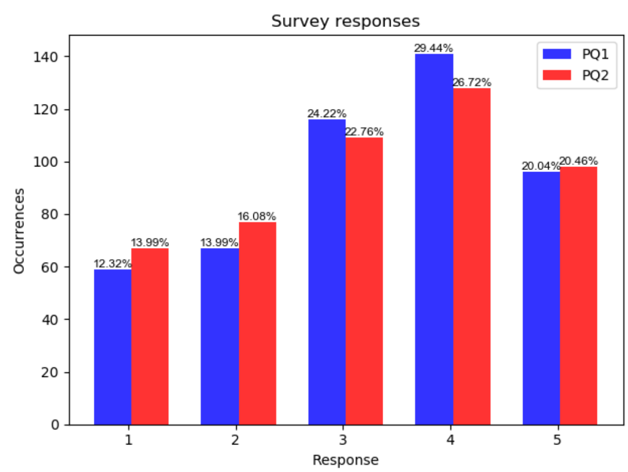


FIGURE 7. Perception survey results.

PQ2, with the grade obtained by the students in the test performed at the end of the fourth week. Fig. 8 presents a comparison, and in general terms this shows a directly proportional relationship between the feedback "acceptance" and the test grade.

To complement this descriptive analysis, we performed a hypothesis test. To do this, we separate students into three categories: "Low feedback acceptance" (L) for those who

TABLE 4. ANOVA test results.

	Sum of squares	Degrees of freedom	Root mean square	F	F-critical
Inter-groups	94.79	2	47.39	53.05	3.01
Intra-groups	425.23	476	0.89		
Total	250.02				

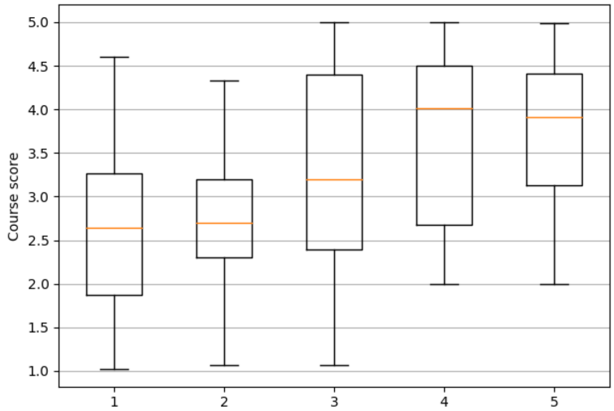


FIGURE 8. Data distribution for PQ2 results versus test grade.

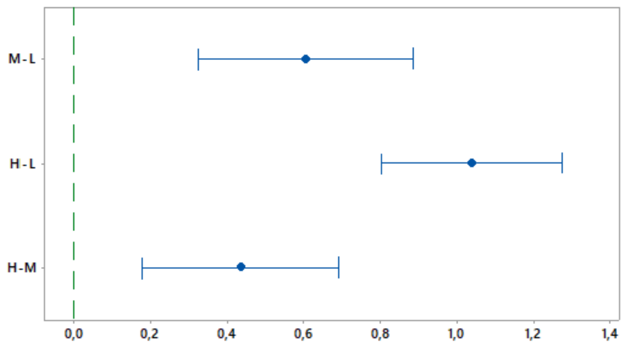


FIGURE 9. Tukey test results.

replied to PQ2 with a score of 1 or 2, "Medium feedback acceptance" (M) for those who replied to PQ2 with a score of 3, and "High feedback acceptance" (H) for those who replied to PQ2 with a score of 4 or 5. Then, we proposed the following hypothesis:

$$H_0 : \mu_L = \mu_M = \mu_H$$
$$H_1 : \mu_L \neq \mu_M \neq \mu_H \quad (1)$$

We performed an ANOVA test with a significance level of $\alpha = 0.05$, for which the results are presented in Table 4. Since $F \geq F_{critical}$ ($53.05 \geq 3.01$), the null hypothesis is rejected in favor of the alternative. That means that a significant difference exists between the performance in the three categories of students.

In view of this result, we also performed a Tukey test with a significance level of 0.05 to verify the alternative hypothesis, for which the results are summarized in Fig. 9.

Since none of the three intervals contain zero, it could be said that the corresponding means are significantly different. There is therefore evidence for better performance by the students with "High feedback acceptance" ($\mu = 3.71$, $N = 226$) compared to those with "Medium feedback acceptance" ($\mu = 3.28$, $N = 109$) and those with "Low feedback acceptance" ($\mu = 2.67$, $N = 144$).

IV. CONCLUSION

The aim of this study was to contribute to the use of formative assessment by (i) developing a framework for mathematics courses in which teachers can define highly dynamic and flexible assessment items, and (ii) performing a preliminary validation of such a framework via an experiment in a real educational scenario.

A total of 458 students participated in this experiment, and three sources of data were analyzed: the LMS event logs of the interactions with all assessment items, the results of a perception survey, and the grades of a final test.

Based on the first source, the results revealed that in nearly one in six cases, the feedback modeled by teachers is followed by the students, and in nearly another one in six, it is not. In the remaining two thirds of cases, it is hard to determine what they actually did.

Nevertheless, according to the second source, nearly half of the students had a positive opinion of the feedback received. Moreover, statistical tests carried out on the data from the third source demonstrated that on average, this group of students obtained higher performance in the final test compared to the group that had a negative opinion.

These results are promising, and coincide with the findings of other research regarding the usefulness of automated FA with regard to student performance and perception [3], [10], [19], [27], [35]. In particular, we agree with Gikandi, Morrow & Davis who, after a meta-analysis in which the review 18 empirical studies, state that such approach can foster a learner focus and enhance student engagement with valuable learning experiences [38]. Nevertheless, we also coincide with them when claim that there are several issues, besides automated feedback, that must be addressed to achieve a more authentic FA process. Having said that, rather than a goal accomplished, we see the obtained results as a starting point for further studies. Among those studies, we would like to analyze how teachers use the information of the recurrent misconceptions to modify learning contents or provide clarifications, and then, how those actions affect students. In the same way, but processing the information not aggregated by misconception but by student, a profile of his/her could be modeled and present early alerts to the teacher when necessary. Also, we would like to analyze the evolution of students' behavior over an entire course or even a group of courses, particularly focusing on how they use feedback as they get the results of periodic testing. Finally, from a technical perspective, we would like to extend the formal language of the framework, for example to include other types of questions such as ordering or pairing, or more distributions for parameters such as normal or binomial.

REFERENCES

[1] H. C. Chu, J. M. Chen, and C. L. Tsai, "Effects of an online formative peer-tutoring approach on students' learning behaviors, performance and cognitive load in mathematics," *Interact. Learn. Environ.*, vol. 25, no. 2, pp. 203–219, 2017.

[2] G. A. Nortvedt, L. Santos, and J. Pinto, "Assessment for learning in Norway and Portugal: The case of primary school mathematics teaching," *Assessment Educ., Princ., Policy Pract.*, vol. 23, no. 3, pp. 377–395, Jul. 2016.

[3] J. Lucariello, M. T. Tine, and C. M. Ganley, "A formative assessment of students' algebraic variable misconceptions," *J. Math. Behav.*, vol. 33, pp. 30–41, Mar. 2014.

[4] K. E. Fong and T. Melguizo, "Utilizing additional measures of high school academic preparation to support students in their math self-assessment," *Community College J. Res. Pract.*, vol. 41, no. 9, pp. 566–592, Sep. 2017.

[5] F. Radmehr and M. Drake, "An assessment-based model for exploring the solving of mathematical problems: Utilizing revised Bloom's taxonomy and facets of metacognition," *Stud. Educ. Eval.*, vol. 59, pp. 41–51, Dec. 2018.

[6] J. Phelan, K. Choi, D. Niemi, T. Vendlinski, E. Baker, and J. Herman, "The effects of POWERSOURCE assessments on middle-school students' math performance," *Assessment Educ., Princ., Policy Pract.*, vol. 19, no. 2, pp. 211–230, 2012.

[7] J. Phelan, K. Choi, T. Vendlinski, E. Baker, and J. Herman, "Differential improvement in student understanding of mathematical principles following formative assessment intervention," *J. Educ. Res.*, vol. 104, no. 5, pp. 330–339, Jul. 2011.

[8] S. Trenholm, L. Alcock, and C. Robinson, "An investigation of assessment and feedback practices in fully asynchronous online undergraduate mathematics courses," *Int. J. Math. Educ. Sci. Technol.*, vol. 46, no. 8, pp. 1197–1221, Nov. 2015.

[9] I. G. Rojas and R. M. C. Garcia, "Towards efficient provision of feedback supported by learning analytics," in *Proc. IEEE 12th Int. Conf. Adv. Learn. Technol.*, Jul. 2012, pp. 599–603.

[10] E. Kurvinen, E. Lokkila, R. Lindén, E. Kaila, and M. Laakso, "Automatic assessment and immediate feedback in third grade mathematics," in *Proc. Ireland Int. Conf. Educ.*, 2016, pp. 15–23.

[11] D. Tempelaar, B. Rienties, J. Mittelmeier, and Q. Nguyen, "Student profiling in a dispositional learning analytics application using formative assessment," *Comput. Hum. Behav.*, vol. 78, pp. 408–420, Jan. 2018.

[12] K. Rakoczy, B. Harks, E. Klieme, W. Blum, and J. Hochweber, "Written feedback in mathematics: Mediated by students' perception, moderated by goal orientation," *Learn. Instruct.*, vol. 27, pp. 63–73, Oct. 2013.

[13] N. D. Pacheco-Venegas, G. López, and M. Andrade-Aréchiga, "Conceptualization, development and implementation of a Web-based system for automatic evaluation of mathematical expressions," *Comput. Educ.*, vol. 88, pp. 15–28, Oct. 2015.

[14] J. Petrović, P. Pale, and B. Jeren, "Online formative assessments in a digital signal processing course: Effects of feedback type and content difficulty on students learning achievements," *Educ. Inf. Technol.*, vol. 22, no. 6, pp. 3047–3061, Nov. 2017.

[15] A. Barana and M. Marchisio, "Ten good reasons to adopt an automated formative assessment model for learning and teaching mathematics and scientific disciplines," *Procedia-Social Behav. Sci.*, vol. 228, pp. 608–613, Jul. 2016.

[16] P. Black and D. William, "Assessment and classroom learning," *Assessment Educ., Princ., Policy Pract.*, vol. 5, no. 1, pp. 7–74, 1998.

[17] K. Rakoczy, P. Pinger, J. Hochweber, E. Klieme, B. Schütze, and M. Besser, "Formative assessment in mathematics: Mediated by feedback's perceived usefulness and students' self-efficacy," *Learn. Instruct.*, vol. 60, pp. 154–165, Apr. 2019.

[18] R. E. Bennett, "Formative assessment: A critical review," *Assessment Educ., Princ., Policy Pract.*, vol. 18, no. 1, pp. 5–25, Feb. 2011.

[19] M. I. Núñez-Peña, R. Bono, and M. Suárez-Pellicioni, "Feedback on students' performance: A possible way of reducing the negative effect of math anxiety in higher education," *Int. J. Educ. Res.*, vol. 70, pp. 80–87, 2015.

[20] J. Hattie and H. Timperley, "The power of feedback," *Rev. Educ. Res.*, vol. 77, no. 1, pp. 81–112, 2007.

[21] G. M. N. Isabwe, F. Reichert, M. Carlsen, and T. A. Lian, "Using assessment for learning mathematics with mobile tablet based solutions," *Int. J. Emerg. Technol. Learn.*, vol. 9, no. 2, p. 29, Apr. 2014.

[22] P. Pinger, K. Rakoczy, M. Besser, and E. Klieme, "Implementation of formative assessment-effects of quality of programme delivery on students' mathematics achievement and interest," *Assessment Educ., Princ., Policy Pract.*, vol. 25, no. 2, pp. 160–182, 2018.

[23] B. Foster, C. Perfect, and A. Youd, "A completely client-side approach to e-assessment and e-learning of mathematics and statistics," *Int. J. e-Assessment*, vol. 2, no. 2, pp. 1–12, 2012.

[24] T. Carrol, K. Mulchrone, Á. Ní She, J. Crowley, and D. Casey, *Transitioning to E-Assessment in Mathematics Education*. Cork, Ireland: National Forum for the Enhancement of Teaching and Learning in Higher Education, 2016.

[25] K.-E. Chang, Y.-T. Sung, and S.-F. Lin, "Computer-assisted learning for mathematical problem solving," *Comput. Educ.*, vol. 46, no. 2, pp. 140–151, Feb. 2006.

[26] I. J. Arnold, "Cheating at online formative tests: Does it pay off?" *Internet Higher Educ.*, vol. 29, pp. 98–106, Apr. 2016.

[27] J. Hannah, A. James, and P. Williams, "Does computer-aided formative assessment improve learning outcomes?" *Int. J. Math. Educ. Sci. Technol.*, vol. 45, no. 2, pp. 269–281, Feb. 2014.

[28] M. Pauna, "Calculus courses' assessment data," *J. Learn. Anal.*, vol. 4, no. 2, pp. 12–21, 2017.

[29] M. Pezzino, "Online assessment, adaptive feedback and the importance of visual learning for students. The advantages, with a few caveats, of using MapleTA," *Int. Rev. Econ. Educ.*, vol. 28, pp. 11–28, May 2018.

[30] C. J. Sangwin, "Inequalities, assessment and computer algebra," *Int. J. Math. Educ. Sci. Technol.*, vol. 46, no. 1, pp. 76–93, Jan. 2015.

[31] C. J. Sangwin, "Assessing elementary algebra with STACK," *Int. J. Math. Educ. Sci. Technol.*, vol. 38, no. 8, pp. 987–1002, Dec. 2007.

[32] A. R. Lucas, "Using WeBWorK, a Web-based homework delivery and grading system, to help prepare students for active learning," *Primus*, vol. 22, no. 2, pp. 97–107, 2012.

[33] N. Engelke, G. Karakok, and A. Wangberg, "Engaging students in the classroom with WeBWorK CLASS," *Primus*, vol. 26, no. 6, pp. 570–584, 2016.

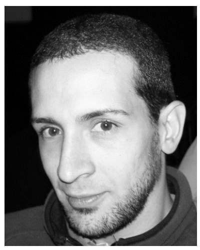
[34] S. Hauk, R. A. Powers, and A. Segalla, "A comparison of Web-based and paper-and-pencil homework on student performance in college algebra," *Primus*, vol. 25, no. 1, pp. 61–79, Jan. 2015.

[35] M. Kawazoe and K. Yoshitomi, "E-learning/e-assessment systems based on webMathematica for university mathematics education," *MSOR Connections*, vol. 15, no. 2, p. 17, Jul. 2017.

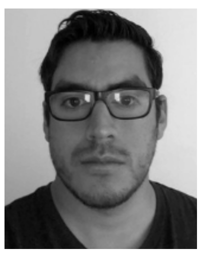
[36] B. Grugeon-Allys, F. Chenevotot-Quentin, J. Pilet, and D. Prévit, "Online automated assessment and student learning: The PEPITE project in elementary algebra," in *Uses of Technology in Primary and Secondary Mathematics Education*. Cham, Switzerland: Springer, 2018, ch. 13, pp. 245–266.

[37] C. M. Eddy, P. Harrell, and L. Heitz, "An observation protocol of short-cycle formative assessment in the mathematics classroom," *Investigations Math. Learn.*, vol. 9, no. 3, pp. 130–147, Jul. 2017.

[38] J. Gikandi, D. Morrow, and N. Davis, "Online formative assessment in higher education: A review of the literature," *Comput. Educ.*, vol. 57, no. 4, pp. 2333–2351, Dec. 2011.



JULIÁN MORENO received the bachelor's, master's, and the Ph.D. degrees in computer science and informatics from the National University of Colombia, in 2007, 2009, and 2013, respectively, and the Ph.D. degree in computers and education from the Federal University of Rio Grande do Sul. He is currently an Associate Professor with the National University of Colombia. He is the author of over 30 scientific articles in refereed international conferences and journals. He has led several software development projects in education and gamification. His research interests are focused on computers and education, data structures, statistics, and learning analytics.



ANDRÉS F. PINEDA received the bachelor's and master's degrees in computer science and informatics from the National University of Colombia at Medellín, Medellín, in 2013 and 2015, respectively, where he is currently pursuing the Ph.D. degree with the Computers and Informatics Program. His research interests are computers and education, learning analytics, software engineering, and methodologies to improve the impact of ICT on mathematics and programming classrooms.