Journal on Interactive Systems, 2023, 14:1, doi: 10.5753/jis.2023.3247

Tutoria: a software platform to improve feedback in education

Taciana Pontual Falcão [**Universidade Federal Rural de Pernambuco** | ***taciana.pontual@ufrpe.br***]
Verenna Arêdes [**Universidade Federal Rural de Pernambuco** | ***verenna.aredes@gmail.com***]
Samuel Barbosa Jatobá de Souza [**Universidade Federal de Pernambuco** | ***sbjs@cin.ufpe.br***]
Giuseppe Fiorentino [**Universidade Federal Rural de Pernambuco** |***fiorentinogiuseppebcc@gmail.com***]
José Rodrigues Neto [**Universidade Federal Rural de Pernambuco** | ***rodriguesnt01@gmail.com***]
Gabriel Alves [**Universidade Federal Rural de Pernambuco** | ***gabriel.alves@ufrpe.br***]
Rafael Ferreira Mello [**Universidade Federal Rural de Pernambuco** | ***rafael.mello@ufrpe.br***]

Abstract

Educational feedback is essential to help students learn from their mistakes and self-regulate their learning strategies. However, work overload and lack of time are barriers for educators to give quality and timely feedback, particularly for written assessments. Software tools to support feedback processes typically focus on automatic messages, lacking personalization. We present Tutoria, a software tool that uses artificial intelligence techniques to correct assessments more efficiently while also ensuring that good practices of educational feedback are followed. Tutoria was developed through a user-centered design process, including interviews and prototype validation with undergraduate students and instructors from higher education institutions in different fields of knowledge. Results indicate that the software presents good usability and relevance for educators. We expect that Tutoria can help educators construct personalized written feedback efficiently, allowing them to give quality feedback to large groups within realistic time frames.

Keywords: *Educational feedback, Software tools, Written assessment*

1 Introduction

In learning processes, quality feedback is critical to help students understand the gaps between their current performance and the competencies they are expected to develop (Wiggins, 1998; Sadler, 1989), and thus self-regulate their learning to a successful trajectory. However, quality feedback in education is often overlooked. In the way the educational system is organized, instructors have a hierarchical and intellectual power which makes them the most prominent and reliable source of feedback in the eyes of students. But, although instructors are very effective in identifying errors, conveying quality and timely feedback is not that straightforward. First of all, being a process of communication, the feedback includes issues of discourse, identity, power, control and social relationship that must be taken into account by instructors (Higgins et al., 2001). Instructors struggle to find the right time, tone and content of feedback messages, and they are often complex and challenging to translate into action (Ivanic et al., 2000; Higgins et al., 2001). Secondly, the demands to provide quality and timely feedback encounter two intertwined barriers: work overload and little time available (Ivanic et al., 2000; Higgins et al., 2001). In higher education, instructors struggle to deliver consistent, timely and constructive feedback to meet the needs and expectations of students (Carless et al., 2011; Boud and Molloy, 2013).

Several characteristics are cited in the literature to define quality feedback: more descriptive than evaluative; corrective advice rather than non-specific comments, such as exhortations (e.g., “try harder”); establishing a dialogue between instructors and students; closing the gap between current and desired performance; avoid excessive criticism and encourage motivational beliefs; among others (Wiggins, 1998; Nicol and Macfarlane-Dick, 2006; Sadler, 1989; Freeman and Lewis, 2016). Also very important is to give feedback on time, i.e. close to the delivery of the assignment, and researchers go even further to say that much greater emphasis should be placed on providing feedback for work-in-progress and allowing resubmissions, thus creating opportunities for students to use the feedback to improve their work and their learning (Boud, 2000; Hounsell, 2004). Sending feedback promptly is rare in higher education when students typically move to the next assignment just after they receive feedback on the previous one (Nicol and Macfarlane-Dick, 2006), or, much worse, they only receive feedback for all assignments at the end of the course.

As digital technologies become increasingly integrated into teaching and learning, and many assignments are delivered in a digital format, opportunities to develop software tools to support and facilitate the feedback process are broadened. Typically, however, these tools have focused on the automatic correction of multiple-choice assessments or on sending automated messages from the instructor’s correction (Cavalcanti et al., 2021). These approaches lack personalization and often fail to establish connections with the students as part of the communication process. More recent works shift to the goal of assisting instructors in constructing quality feedback instead (Pardo et al., 2019; Cavalcanti et al., 2020; Tsai et al., 2021).

Within this context, our research question is: How can we scale the process of giving educational feedback while ensuring quality and personalization of the feedback? In this paper, we present Tutoria, a software platform developed through user-centered design (UCD) (Barbosa and Silva, 2010) to

help instructors compose quality feedback messages within a flow of correction of assignments, with a focus on open questions. We hope to help instructors not only be able to give feedback beyond grades but also help them produce feedback messages that are informative and effectively enable students to understand the gaps between actual and expected performances.

The paper is organized as follows. In Section 2, we discuss a little further about the role of feedback in education. In Section 3, we explain the method used to develop our software solution. In Section 4, we discuss instructors and students' main needs and desires regarding the feedback process, identified through interviews. Then, we present the main functionalities of Tutoria (Section 5), followed by the results of user evaluation (Section 6).

2 Educational Feedback

Feedback is a crucial activity in the learning process. It enhances communication between students and educators, clarifying expectations, monitoring the current progress of learners, and moving towards learning goals (Hattie and Timperley, 2007). Several theories seek to define good-quality feedback. For instance, Nicol and Macfarlane-Dick (2006) described as good feedback practice any strategy or content that could enhance students' capacity to self-regulate their learning performance. The authors proposed seven general principles of good feedback encompassing aspects such as: helping to clarify what good performance is; facilitating the development of self-assessment (reflection) in learning; encouraging positive motivational beliefs and self-esteem, among others. Hattie and Timperley (2007) proposed another point of view where educational feedback can be seen as four-level content related to learning tasks, learning process, student self-regulation, and student motivation. For the authors, the level of feedback on tasks is only valuable if combined with the other levels, which are generally missing.

Despite the vast literature on the importance of educational feedback and what constitutes quality feedback, there is substantial evidence showing that in higher education, instructors struggle to deliver consistent, timely and constructive feedback to meet the needs and expectations of students (Carless et al., 2011; Boud and Molloy, 2013). In general, instructors fail to provide feedback that speaks to the needs of individual students. Higgins et al. (2001) argue that assessment feedback is a process of communication and that as such it includes issues of discourse, identity, power, control and social relationship.

Several tools have been developed over time to assist instructors in the feedback process (Gulwani et al., 2014; Marin et al., 2017; Krusche and Seitz, 2018). However, the majority focuses on sending automatic feedback messages, or performing automatic correction using online judges that lack personalization and educational features (Santos and Ribeiro, 2012; Cavalcanti et al., 2021). Recently, research shifted to the goal of assisting instructors in understanding students' behavior (Pereira et al., 2020), and constructing quality, personalized feedback instead of sending automatic messages (Pardo et al., 2019; Cavalcanti et al., 2020; Tsai et al., 2021). Yet, the accountability and quality assurance of the feedback process is still an open issue to be addressed (Winstone and Carless, 2021; Pereira et al., 2020).

Qualitative studies performed through focus groups evaluated what students and instructors from a Brazilian higher education institution perceived as important topics to address in the learning process (Falcao et al., 2019; Falcão et al., 2020). Among other topics, results demonstrated that students are not satisfied with the feedback provided by instructors. On the other hand, instructors reported that they are too overloaded with their academic activities to dedicate the necessary time to provide good quality feedback, feeling frustrated as they recognize its importance. Another study in the same institution showed that the ideal expectations of students and instructors about feedback provision are much higher than their perceptions of what they consider realistic in the context of their institution (Garcia et al., 2021). In other words, instructors would like to provide quality feedback but do not see this happening in the short run, considering the tools and workload they have at present.

3 Method

In order to propose a software platform to assist instructors in giving educational feedback, we adopted a UCD process Barbosa and Silva (2010) with iterative cycles of user research and analysis, ideation, prototyping, and user testing. In the first phase (user research and analysis), we performed semi-structured individual interviews in order to better understand instructors' and students' needs related to the assessment and feedback process in the context of Brazilian higher education. Twenty-two higher education instructors from 9 different fields and 38 undergraduate students from 13 different degree programs, from Brazilian public and private universities, were interviewed. All interviews were performed through video calls using Google Meet. A member of the research team conducted the interview while another member took notes. Interviews with instructors lasted around one hour, while interviews with students lasted about a half-hour.

The interview scripts for instructors included the following topics: methods for evaluating students (e.g. tests, reports, seminars); types of feedback given to students (e.g. oral, written) and its constitutive elements (e.g. explanation for errors, providing the correct solutions, indicating study material); tools for giving feedback; ways of following up students' progress; difficulties in the process of assessment; qualities of good assessment and feedback; characteristics of poor evaluations and their impact for learning; assessment in online teaching; strategies to motivate students. The interview scripts for students included: qualities of good assessment; relevance of elements of feedback; ways instructors evaluate them; their opinion and expectations in the learning process; learning from feedback; engagement; challenges of online learning; platforms used in online courses; opinions about the automatic correction of activities. The questions about online teaching and learning were added due to the sudden migration of classes to this modality because of the Covid-19 pandemic. As universities were going through this adaptation as we performed the research, we decided to investigate the changes

and needs brought by this new context. Qualitative content analysis was performed by the first author on the interview data using Dovetail software[1], separately for instructors and students. We followed an inductive procedure with open and axial coding.

In the second phase (ideation and prototyping), based on meetings of the research and development team, we developed an interface prototype using the Figma software[2]. Moving to the third phase (user testing), this low-fidelity version was evaluated by six higher education instructors, from different fields of knowledge. Through individual video calls using Google Meet, we presented to the instructors the goal of the tool and the screens designed and asked for their opinions, comments and suggestions.

From the instructors' feedback, we circled back to prototyping and evolved the product into a high-fidelity prototype. We then moved to user testing again, this time with a usability test with 10 instructors. All participants were invited to a Google Meet call and were told that the goal of the test was to assess the usability and utility of the platform. They were given access to a fictitious class in Google classroom and were asked to import the assignment available to the platform and perform the correction. After correcting the assignment, instructions were asked to send the feedback to students. No instructions were given as to how the correction is performed or how feedback can be built and sent with Tutoria, as we wanted to evaluate autonomous use and learnability. At the end of the tasks, instructors were encouraged to comment orally on problems they encountered or suggestions they had for the platform. The call was recorded and all contributions were later discussed by the team to decide how to make adjustments to the product. The instructors were also asked to fill out a form with the 5-point Likert System Usability Scale (SUS). The results of all phases of design are discussed in the next sections.

4 Instructors and students' perspectives on feedback

Five categories emerged from the analysis of interviews: assessment; feedback format, contents, and characteristics; and barriers to giving feedback. Each category had codes that were applied to the interviews' data and counted. Table 1 shows all codes in each category, ordered by the number of occurrences. In this section, we present an overall discussion of the findings for each category.

4.1 Instructors' Perspectives

In this section, we present the instructors' perspectives about feedback, centered on the five categories (Table 1). Whenever quotes from instructors are used to illustrate the findings, we label them with the instructor's ID.

The assessment category revealed that written assignments (including exams) are the most popular format for evaluating students. Indeed, written feedback appeared as more common than oral and automatic. However, the interviews indicate that instructors used to prefer giving oral collective feedback in face-to-face class, as it is quick, simple and effective, but were deprived of this possibility by the Covid-19 pandemic and the conditions of emergency online teaching. Some instructors tried to transpose these face-to-face moments to online meetings (individually or in group) so that they could give oral feedback. However, scheduling these meetings is very time consuming, and students' assiduity proved much lower in the online context. As for written feedback, instructors write their comments directly on each students' assignments (typically a PDF file), but giving written feedback to all students proved impossible for many: *"explaining something through writing can be very hard and demanding"* (I10).

On the other hand, instructors are enthusiastic adopters of software tools when dealing with assessments, such as Google Classroom, Moodle, Google collab, Jupyter, Dojo, Trello, Excel, Repl.it, Telegram, Whatsapp, or sometimes tools developed by themselves. Their interest in tools to support communication, teaching and feedback increased with online teaching, but none of them use tools that were specifically developed for giving educational feedback (i.e. including for example features for facilitating the correction of open questions or optimizing the writing of feedback comments). Thus, they face limitations and frustration caused by: poor usability; the need to use different tools for each purpose; and the lack of specific functionalities for giving educational feedback.

The workload was the most cited barrier for giving feedback (*"giving feedback is very tiresome"* - I12; *"if you want to give good feedback, it's a lot of work"* - I10); closely followed by the number of students per class (*"nowadays it's very hard to scale"* - I12; *"there are too many activities, I cannot give feedback"* - I13); and lack of time (*"often, feedback is long and the instructor cannot deliver it in good time"* - I06). Of course, these three aspects are closely related: *"it's a compromise between the number of students, the size of the feedback, and time you have available at that moment of your life"* - I10. As a result, the most common feedback content is grade (16 occurrences out of 65 coded excerpts in the content category): *"in online teaching, I only send the grades"* - I13; *"my feedback is essentially based on grades"* - I08; *"I receive loads of emails from students asking what they got wrong"* - I05. Rubrics, which could help explain the grades given, are not a rule (9 occurrences): *"I add some comments, but I do not give the criteria for grading"* - I08.

However, instructors try to add explanations to their feedback, when possible, pointing out what is wrong or missing, what can be improved, revealing the correct answer, etc. Although instructors recognize the value of comprehensive feedback, they focus on errors more than on positive aspects: *"focusing on errors is more feasible, but I wish I could give more complete feedback. I can't give positive feedback because there are too many assignments to evaluate"* - I12; *"when the answer is correct, I don't say much, just: ok."* - I10.

Several instructors mentioned identifying recurrent errors and sharing them with the class somehow. Some try to develop patterns from recurrent errors, which can be reused in correction. Instructors are divided as to the educational

[1] https://dovetailapp.com
[2] https://www.figma.com/

Table 1. Categories of analysis (Instructors)

Type of assessment	Feedback format	Feedback contents	Feedback characteristics	Barriers
Written assignment	Use of software	Grade	Individual / Personalized	Workload
Project	Written	Explanations	Quality	Number of students
Written exam	Oral	Errors	See peers' answers	Time
Participation	Automatic	Rubric	Two-way	Lack of experience
Product		Recurrent errors	Iterative	Procrastination
Competences		Positive aspects	Immediate / Timely	Online teaching
General idea			Engaging / Motivating	
Seminar			Importance	
Goals			Humanized	
			Contextualized	

value of sharing peers' errors or allowing students to see their peers' answers. Several of them have brought this method from face-to-face teaching, where they used to solve exercises in groups or discuss results of exams, and found it useful for students to learn (also) from their peers' errors: *"a good feedback is when students can see their peers' answers and their mistakes"* - I02; *"if a student learns from their errors, they will learn even more by seeing their peers' errors too"* - I12. Others worry that such situations might expose and make students uncomfortable: *"I learned students hated it and felt embarrassed because all their peers could see their mistakes"* - I06.

On the other hand, the top cited characteristic of feedback was individualization/personalization. This seems to be, by far, what instructors most value for quality feedback, being also associated with other feedback characteristics such as engaging, motivating and humanized: *"you need to give different feedback for the ways each student functions, otherwise, it won't work"* - I10; *"personalized feedback, with appropriate language, can avoid many problems. It's about how to communicate with students in a more humanized way, so that they will learn without creating blockages"* - I08; *"An ideal feedback is fully personalized, the least generic as possible, meeting the exact needs of the student. Students appreciate it when you give more individual attention, a personalized experience"* - I03; *"Good feedback is dialogical, horizontal, empathetic and sensitive"* - I11. Nevertheless, instructors feel they are unable to achieve it: *"the more students I have, the less personalized feedback I give"* - I10; *"I don't feel comfortable with the feedback I give, because it's not fully tailored for each student"* - I06; *"Instructors who try to give personalized feedback face a lot of difficulties"* - I03; *"I don't give individual feedback, rather I try to work on recurrent errors"* - I02.

Other characteristics of feedback less cited were: it should be two-way (from instructors to students and vice-versa), iterative and contextual: *"If students establish direct communication with me, I am open for discussion and revisions of my feedback"* - I02; *"Ideally, feedback should cover all important aspects and allow for resubmission of a corrected version of the assignment"* - I10; *"feedback needs to be constructed together with the person who receives it"* - I06; and timely - which is problematic given that instructors' time is one of the main barriers for feedback: *"Feedback should be quick - students complain when I take too long to give feedback"* - I09.

Given this scenario, where instructors have a clear opinion about the importance of quality feedback but are admittedly unable to attain it, they were divided about automatic feedback using software tools. Although they fear the lack of the human touch, and that it will not be personalized enough, they also admit the impossibility to deliver quality and timely feedback manually. Thus, several were open and curious about tools that would help them improve their feedback, even if this means having a fully or semi-automatic process: *"Humanizing automatic feedback would be ideal, with language that is more personal. Artificial Intelligence can be used for that, for example for automatically posting comments on discussion forums."* - I02.

4.2 Students' Perspectives

In this section, we present an overall discussion of students' opinions about the feedback they receive from instructors, centered on the same five categories (Table 2). Whenever quotes from students are used to illustrate the findings, we label them with the student's ID.

Students' answers confirmed that written assignments and exams are the most common forms of assessment in their majors, although projects were also cited. Feedback given on these activities was either through text or orally (in online meetings or by sending audio files).

However, overall students were mostly dissatisfied with the feedback provided or the lack of it. According to them, they typically receive grades only, sometimes along with an answer sheet showing the correct expected responses (this usually happens for closed questions). Sometimes, assignments are left with no feedback at all. There are also cases where feedback is too late, at the end of the course when nothing can be done, and they do not know what they got wrong: *"Some teachers gave assignments and disappear, took a long time to give the grades, and in some cases, we ended up without any feedback. We didn't know if we were right or wrong, and in case we were wrong there was nothing we could do about it."* - S05. In many cases, students argue that they learn nothing from feedback and complain about its low quality (saying feedback is "bad" or "superficial").

Ideally, students expect to be continuously evaluated, receiving feedback iteratively and frequently throughout the course. Beyond grades, they would very much like to receive

Table 2. Categories of analysis (Students)

Type of assessment	Feedback format	Feedback contents	Feedback characteristics	Barriers
Written assignment	Tools	Explanations	Continuous / Iterative	Workload
Project	Automatic	Grade	Lack of feedback	Number of students
Written exam	Oral	Errors	Quality	
Debates	Written	Correct answers	Individual / Personalized	
		Positive aspects	Frequency	
			Delayed	
			Engaging / Motivating	
			Contextual	
			Humanized	

explanations, in particular about their errors or aspects to improve: *"I like feedback where the instructor tells me what I got wrong."* - S03; *"I would like to receive feedback saying what I got wrong, why it is wrong, and what would be the correct way to do it. The most relevant part is the reason for being wrong."* - S08; *"It would be great if feedback was more descriptive than right or wrong. I would like to see more than a number (the grade), but also what is missing for me to reach a good performance involving the concepts of each activity."* - S02; *"I would like to know if what I said makes sense, and where I could do better. I'd like to know all the instructor thought of what I said."* - S06.

They also mentioned the benefits of personalized feedback, although this seemed like a very distant scenario for them: *"If the instructor notices that the student is struggling with a specific topic, improve the feedback on that topic so that the student can do better."* - S05. One student (S07) said that frequent meetings would be good to show that the instructor "cares" and "is there for you". Only one student acknowledged, as barriers that instructors face in giving quality feedback, the workload, and the high number of students.

Many tools were cited by students, such as Microsoft Teams, Google Meet, Zoom, Google forms, Google Classroom, Moodle, Blackboard, Github, Slack, Discord and Whatsapp, but none of them with a specific focus on feedback. Rather, they are tools instructors have been using in the context of online learning.

5 Development of Tutoria

From the needs identified in the interviews, we designed Tutoria mainly thinking about features for facilitating the correction of open questions and optimizing the writing of feedback comments. Table 3 maps the main findings from the analysis to Tutoria's features.

Assignments can be imported from Google Classroom or Moodle so that the instructor easily sees the questions of each assignment created in the Learning Management System, as well as students' answers. Google Classroom and Moodle were among the most cited tools by instructors and students.

5.1 Early prototype evaluation

In the ideation phase, the project team had meetings to discuss, prioritize the functionalities and design the software platform, which firstly was prototyped using the Figma tool. This first prototype was evaluated by instructors, by showing them the platform and asking for their opinions. Overall, instructors thought the platform to be a useful and interesting tool to help them correct assignments and give feedback. They reinforced the lack of a tool to help them in this task and liked the proposed layout and design. Considering the increasing number of students per class, the personalization of feedback provided by the platform was seen as a way of "re-humanizing" the instructor-student relationship. They also liked being able to follow the progress of correction, which gives a sense of achievement and being able to see statistics about the class performance in the assignments. No main changes were needed, but the instructors gave several suggestions for additional features, some of which were integrated into the high-fidelity prototype, and others listed for future work. The main takeaway from this phase was the confirmation that the prototype was adequate for the target users, and thus we moved on to develop a high-fidelity prototype, presented in the next section.

5.2 Tutoria

Tutoria is a software platform whose main goal is to help instructors correct and compose written feedback for assignments (created and imported from Google classroom or Moodle) so that the instructor easily sees the questions of each assignment created in the Learning Management System, as well as students' answers. After importing the assignment, the instructor can choose to navigate per question or per student (Figure 1). This means the instructor can either correct the complete assignment of each student, or all students' answers to a specific question.

In the correction process, the instructor can highlight parts of the text of the answer and apply tags, which correspond to errors or correct statements. When highlighting a piece of text, the instructor can apply existing tags (that they created previously) or create new tags on-the-fly (Figure 2). When creating a new tag, the instructor can provide a description that justifies why that tag configures a right or wrong piece of answer. This explanation is mandatory, as it will be used to compose the feedback message to students, but its input can be done later if the instructor prefers to create all tags first and describe them at the end. The category of correct statement is an encouragement for instructors to give positive feedback besides pointing out errors, which configures good

Table 3. Categories of analysis versus Tutoria's features.

Category of analysis	Tutoria's feature
Use of software in the correction process	Tutoria as a web platform
Written feedback for written assignment / exam	Correction of open questions and fields for writing feedback
Grading	Field for inputing student's grade
Explanations	Complete explanations composed from the explanations of the tags applied to each student's assignment
Recurrent errors	Reuse of tags for similar errors, with the support of AI
Positive aspects	Category of tags to encourage marking positive aspects of the assignment
Individual / Personalized feedback	Individual email sent to each student with feedback, which can be edited by the instructor
Quality feedback	Compulsory explanation for each tag created and template to help structure feedback
Number of students and time	Reuse of tags making the correction process faster, and feedback message built from template

pedagogical practice that can motivate students by praising them on what they did right, or aspects at which they excelled Nicol and Macfarlane-Dick (2006). Tags can also be created for omissions (i.e. when the student fails to add a relevant point to justify that statement correctly) or for general comments about the answer. In this case, the instructor does not need to highlight the text, they can simply create an extra tag and describe it.

Tutoria's process of correction accelerates the task as the instructor can reuse previous tags with their descriptions. As the correction evolves, the number of new tags tend to decrease, given the usual repetition of errors. Besides the instructor's own choices of tags, Tutoria also suggests tags for text excerpts that are similar to others already tagged (i.e. occurrences of the same error). Tags suggestions are automatically shown in the interface, for the instructor to accept or reject, making the correction process faster. This functionality is implemented through natural language processing techniques. In short, Tutoria uses the content from similar tags created by the instructor in previous activities to measure the semantic similarity with the new student's answer to recommend previous tags. Our previous work provides more details about the measures used in this process (Mello et al., 2022). Another use of Artificial Intelligence (AI) techniques in Tutoria is plagiarism detection among students' answers, flagging to the instructor every case with similarity above a certain threshold (the default is 85%, but this can be configured by the instructor). For automatic plagiarism detection, we measure the similarity of the full students' answers using the partial token set ratio similarity measure (Yujian and Bo, 2007). The instructor can visualize all similar answers and decide if it is an actual case of plagiarism. By always giving the instructor the last word, we maintain their autonomy, while providing automatic features that can accelerate their work.

Tutoria also supports the correction of multiple-choice questions. In this case, the instructor must indicate the correct answers, and input explanations for all items (correct and incorrect) (Figure 3). Tutoria processes all answers based on this information, so it is not necessary for the instructor to navigate through multiple-choice questions.

When the instructor finishes the correction of an assignment, they will see the compilation of all created tags for that specific assignment and have the opportunity to revise the explanations of each tag, and complete missing explanations. When all tag explanations are complete, the instructor will build the template of the feedback message (Figure 4). They will see predefined blocks of text which will group the explanations for tags in each question. These blocks will be automatically personalized for each student from the tags applied to their answers. Additionally, on this screen, the instructor can add other blocks of text such as greetings and closing statements, sentences to connect the feedback blocks for each question, or general comments about the activity. This process is only done once, as the template will be used for all students. For further personalization, the instructor can add variables such as the student's name, which will be replaced by their value in the final message.

Once the template is complete, the instructor can visualize the final feedback messages automatically generated from the template and the tags applied for each student. Instructors are free to make edits to each individual message as they see fit. The instructor can send the feedback messages to all students in a batch or send the message to a specific student only. Again, this gives the instructor autonomy in notifying students.

5.3 Implementation

The development process was implemented mainly in the Python programming language (back-end) and React Native (front-end). Aside from commonly used Python libraries, to implement our analysis method, we also used the following software packages and libraries:

- The main library for the back-end was Django.
- For Natural Language Processing functionalities, we used the `SpaCy` Python library.
- To extract keywords and create Latent semantic analysis (LSA) semantic space, we used TAGMEsemantic annotation tool.
- To implement the classification algorithm and its training and testing, we used `Scikit-learn` Python library.

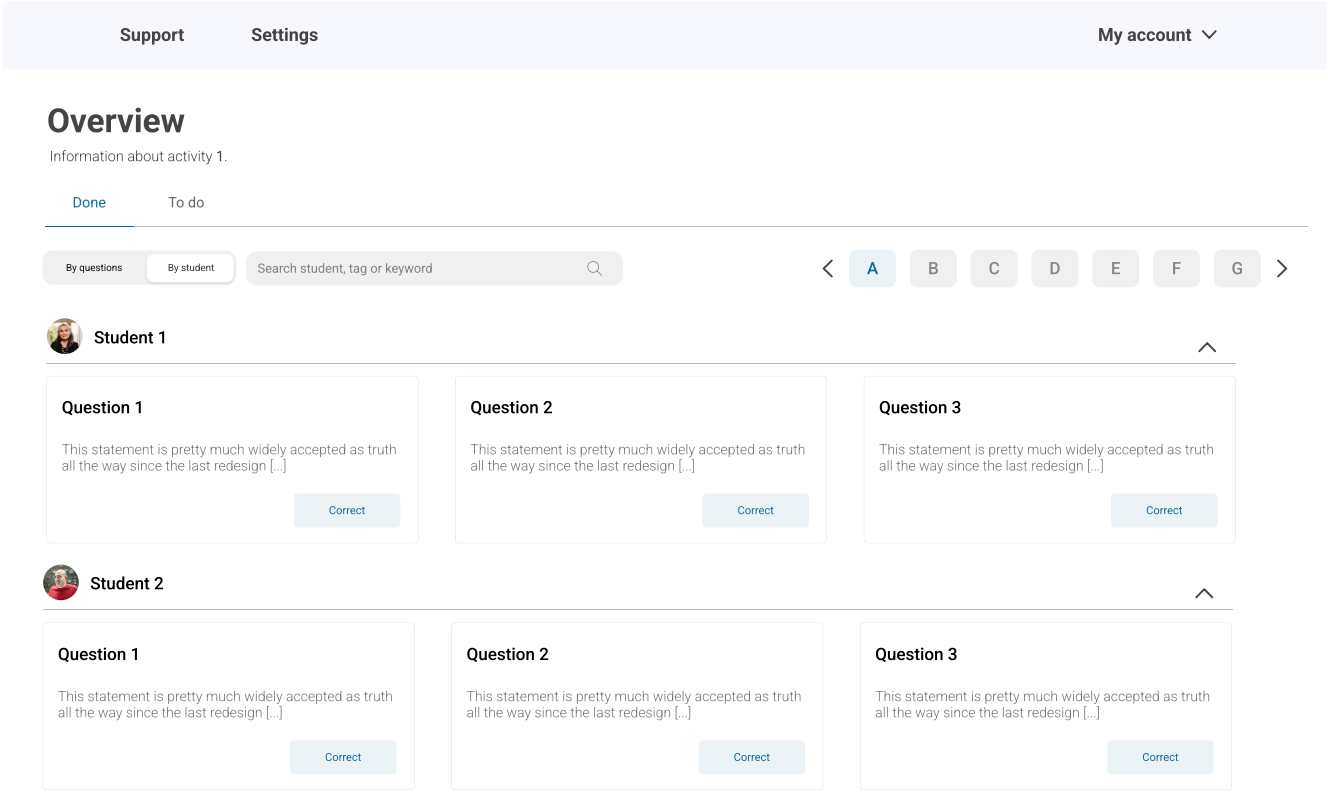

Figure 1. Home screen with students' assignments

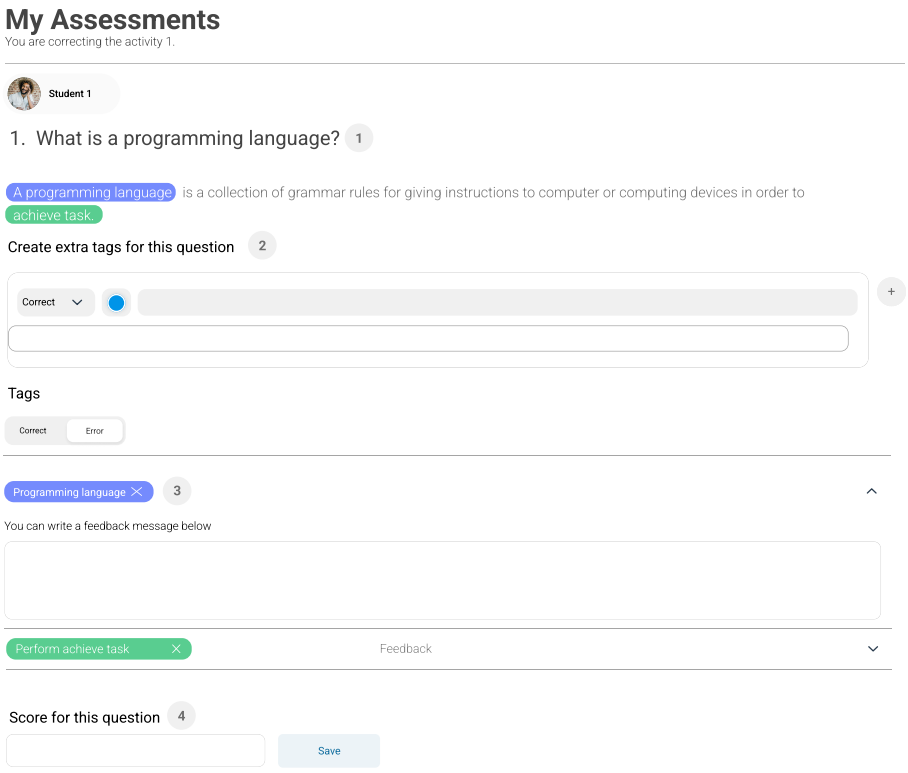

Figure 2. Assessment of an open-ended question.

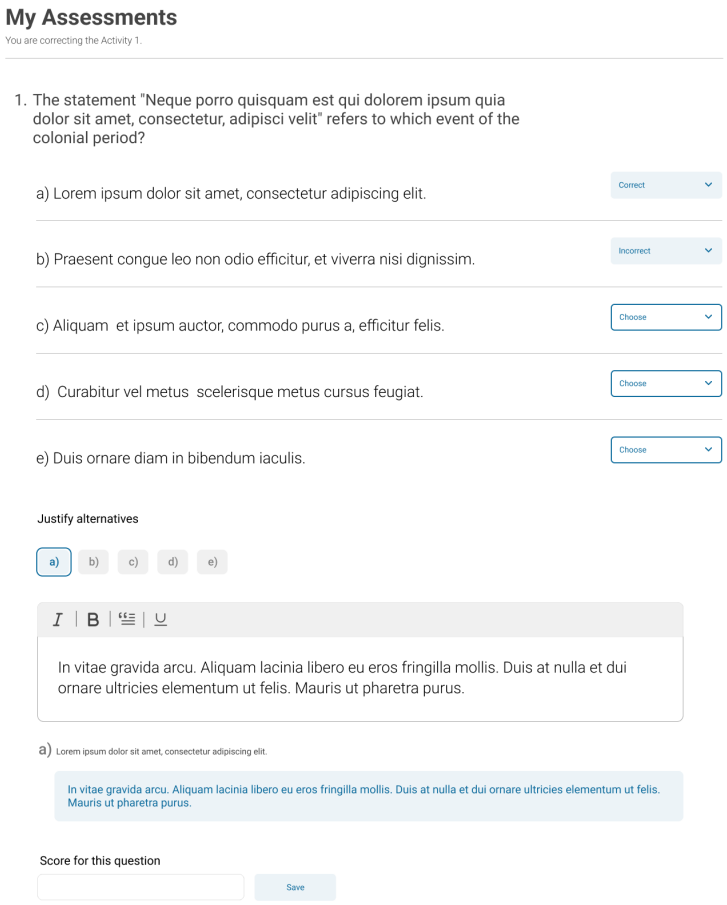

Figure 3. Assessment of a multiple-choice question.

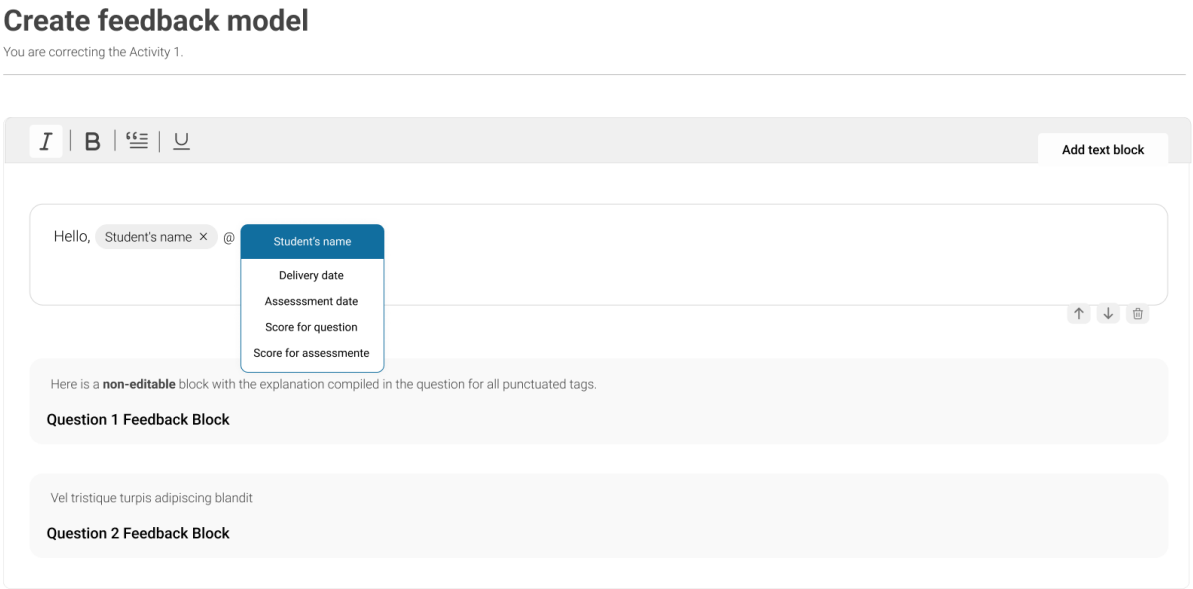

Figure 4. Construction of template for feedback message.

- And finally, the system was deployed in the Google cloud platform using the postgres database.

Tutoria can be accessed for free in the current format, but the code is unavailable. More information is provided on the website https://tutor-ia.com/.

6 User evaluation

We obtained a 73 SUS score from the usability test of our high-fidelity prototype, which indicates good usability. As in the SUS, odd questions are positive (ideally participants would agree with them), while even questions are negative (ideally, participants would disagree), the results shown in Figure 5 indicate an overall good evaluation of the platform. Most participants would like to use the system frequently (Q1) and agreed it is easy to use (Q3). There was a little less agreement about the good integration of the platform's different functions (Q5); quick learnability (Q7); and user confidence (Q9). With regard to the negative questions, participants mostly disagreed that they had to learn a lot of things to use the system (Q10); and that there is a lot of inconsistency (Q6). However some thought the platform was unnecessarily complex (Q2), that technical help might be needed (Q4), and that the use was somewhat cumbersome (Q8).

After completing the questionnaire, we had an informal discussion with the group of instructors, who orally reported specific problems and spontaneously gave suggestions for improvements. Regarding the correction process, the main problem was that none of the instructors understood the use of the extra tags, which meant providing a space where they could make general comments without marking specific excerpts in the answers. Interestingly, they asked for this very same functionality in their suggestions, which demonstrates that this is a need, but also that the interface is not communicating the functionality in a clear way. Another improvement needed is in how the existing tags are shown to the instructors during the correction (through a filtered list in a pop-over window) - they found it confusing and did not easily understand what the interface was showing. Instructors also suggested having a third category of tag, besides right and wrong, which would be something along the lines of "partially right". Moreover, regarding the progress of correction shown dynamically, the instructors gave suggestions for showing information in a way that reflects every little progress, thus being more motivational for the completion of the task. In other words, the progress should be updated as each question is corrected, as well as when the whole assignment from one student is completed; or when all answers for a certain question are corrected.

As to formatting the template for the feedback messages, improvements are needed in the flow between editing, saving and sending, and in how to access the screen to build the template. Regarding the plagiarism, instructors said that the possibility of configuring the threshold needs to be clearly communicated. At present, the instructor must access the Configurations menu, and instructors said they would not think this was an option they would look for or expect to find. General comments included the need for a fixed lateral menu for global navigation (which currently is shown only in certain screens); and clear feedback for all actions performed (e.g. saving changes and adjusting configurations).

7 Conclusion

With the Covid-19 pandemic, instructors from higher education were forced to work from home and teach remotely. Interaction with students was reduced drastically and opportunities to give feedback became more limited, mainly reduced to writing. Instructors were overwhelmed by the demands of a totally new way of teaching and emotional stress. Workload augmented significantly, and time for feedback, which was already little, disappeared. This situation aggravated a problem that is not new: although the importance of feedback seems consensual, instructors do not have time to produce it in good time and quality.

In order to develop a software solution to support the process of giving educational feedback, we performed interviews with instructors and students in higher education. For the instructors interviewed, quality feedback is, above all, personalized. It explains the errors, indicates what is missing and what can be improved, and highlights positive aspects to motivate students, using adequate language and tone. Some instructors added that evaluation criteria are important for students to understand their performance more clearly. However, feedback does not scale: more students means less feedback. Often, instructors are only able to give students a grade, although they agree that grades are not helpful for students to learn from their mistakes. In online courses, feedback is seen as more challenging and time-consuming, as the possibilities of interaction became more limited: opportunities for perceiving students facial expressions during class and spontaneously giving oral feedback in real-time and class discussions became rare. Thus, most feedback in online courses is being given in writing.

Interviews with students confirmed that the feedback they receive from instructors is often grades-only, leading to a lack of clarity as to evaluation criteria and missed opportunities to learn from mistakes. In addition, it is often delayed: at times, students received all their grades at the end of the course when there is nothing left to be done to improve achievement in that particular course. All this leads to high levels of frustration from students when it comes to feedback.

Although we must acknowledge the limitation of these results as the qualitative analysis was performed by one researcher only, which can introduce bias, it is clear from the findings that giving timely and quality feedback to individual students manually is unfeasible for instructors in the present context of Brazilian higher education, as much as they believe in its importance. We hypothesize that software solutions could make a difference in assisting instructors and making this task possible.

In this sense, online teaching broadened the opportunities of using digital technologies in formal education, as all activities and assignments were migrated to the virtual world. However, in the instructors' opinions, the tools they use do not give proper support for feedback (e.g. the learning management system Moodle, available in most Brazilian public universities). Software tools specifically developed for giving

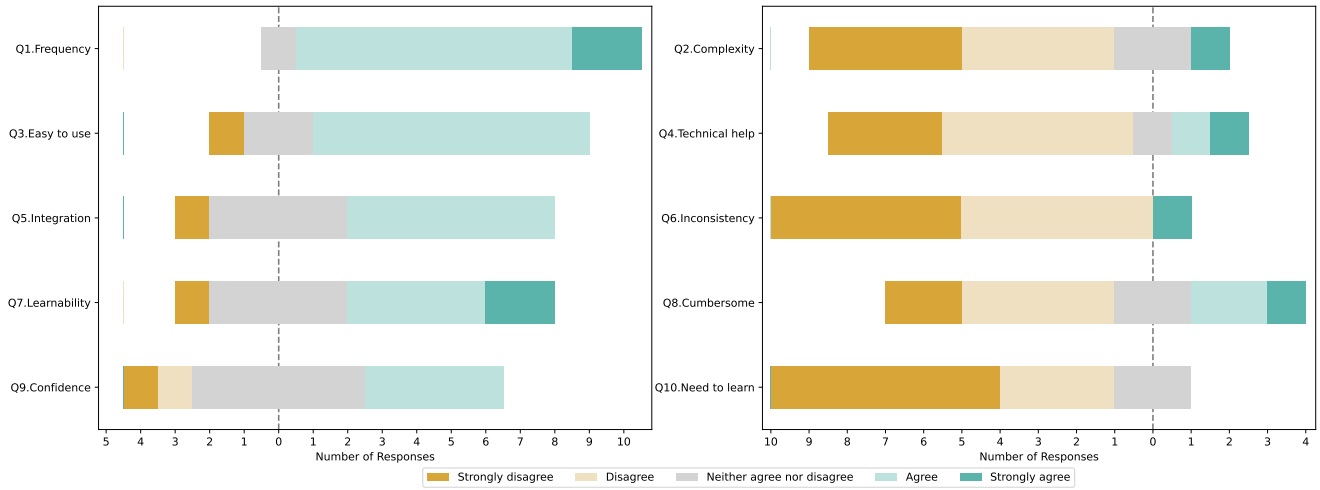

Figure 5. Results from System Usability Scale (SUS)

feedback can enable instructors to perform this activity more efficiently. Although several tools already exist on the market[3], they do not guide instructors towards composing quality and informative feedback messages. Indeed, how teachers should frame feedback comments, the discourse they should use, the quantity of comments among other aspects are under-researched topics in the area (Nicol and Macfarlane-Dick, 2006; Higgins et al., 2001). We aim to provide this support with Tutoria, a software platform which, through the process of correction of assignments, helps instructors compose quality feedback messages for students. By using AI algorithms, Tutoria facilitates and accelerates the correction of open-ended questions and provision of written feedback. The platform was developed through a UCD process, collecting target users' opinions and improving the interface and functionalities iteratively. User tests with instructors indicated that Tutoria has good usability and relevant functionalities. Future work includes evaluation with students about the type of feedback received by instructors using Tutoria.

The practical implications of this study include demonstrating the potential of using AI (i.e., the tag recommendation system) to assist the instructor in the activity of assessing open-ended responses effectively. Differently from other studies (Ferreira-Mello et al., 2019; Cavalcanti et al., 2021), the approach is the based process proposed by Pardo et al. (2018) that allows reducing the workload of instructors by allowing them to reuse previously defined correct statements and errors, which increases the reliability and consistency in grading students' activities (Ragupathi and Lee, 2020) and potentially reduces bias in assessment (Erickson and Botelho, 2021). Language processing could be used to adjust the communication process, identifying inadequate tone or complex language (Higgins et al., 2001), and suggesting replacements. Evaluating the length of the feedback message (i.e. quantity of comments) could also be done, as research has indicated that too many comments become unproductive as they overwhelm students (Freeman and Lewis, 2016). Finally, Tutoria could have a module dedicated for programming courses, with functionalities specific to the process of correcting codes.

We hope Tutoria can eventually be adopted at a large scale, contributing to give feedback the role to which it is entitled in the learning process.

[3]Avalia (https://siteavalia.grupoa.com.br/), OnTask (https://www.ontasklearning.org/), Gradescope (https://www.gradescope.com/)

Acknowledgements

This paper is an extended version of the paper "What did I get Wrong? Supporting the Feedback Process in Computer Science Education", published in the proceedings of the 30o Workshop about Education in Computing (Workshop sobre Educação em Computação - WEI). This project was funded by RNP - Rede Nacional de Ensino e Pesquisa. All participants were volunteers who agreed to participate in the research and had their anonymity guaranteed, as well as the right to withdraw at anytime. According to the CRediT Taxonomy, authors' contributor roles were: Taciana Pontual Falcão - Conceptualization, Investigation, Methodology, Supervision, Validation, Visualization, Writing original draft and Writing review and editing; Verenna Arêdes - Conceptualization, Software, Validation, Visualization; Samuel Souza - Conceptualization, Software, Visualization; Giuseppe Fiorentino and José Rodrigues Neto - Conceptualization, Software, Validation; Gabriel Alves - Conceptualization, Project administration, Software, Supervision; Rafael Ferreira de Mello - Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Writing review and editing.

References

Barbosa, S. and Silva, B. (2010). *Interação humano-computador*. Elsevier Brasil.

Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in continuing education*, 22(2):151–167.

Boud, D. and Molloy, E. (2013). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in higher education*, 38(6):698–712.

Carless, D., Salter, D., Yang, M., and Lam, J. (2011). De-

veloping sustainable feedback practices. *Studies in higher education*, 36(4):395–407.

Cavalcanti, A. P., Barbosa, A., Carvalho, R., Freitas, F., Tsai, Y.-S., Gašević, D., and Mello, R. F. (2021). Automatic feedback in online learning environments: A systematic literature review. *Computers and Education: Artificial Intelligence*, 2:100027.

Cavalcanti, A. P., Diego, A., Mello, R. F., Mangaroska, K., Nascimento, A., Freitas, F., and Gašević, D. (2020). How good is my feedback? a content analysis of written feedback. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, pages 428–437.

Erickson, J. A. and Botelho, A. (2021). Is it fair? automated open response grading. In *International Conference on Educational Data Mining*.

Falcao, T. P., Ferreira, R., Rodrigues, R. L., Diniz, J., and Gasevic, D. (2019). Students' perceptions about learning analytics in a brazilian higher education institution. In *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)*, volume 2161, pages 204–206. IEEE.

Falcão, T. P., Mello, R. F., Rodrigues, R. L., Diniz, J. R. B., Tsai, Y.-S., and Gašević, D. (2020). Perceptions and expectations about learning analytics from a brazilian higher education institution. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, pages 240–249.

Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., and Romero, C. (2019). Text mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(6):e1332.

Freeman, R. and Lewis, R. (2016). *Planning and implementing assessment*. Routledge.

Garcia, S., Marques, E., Mello, R. F., Gašević, D., and Falcão, T. P. (2021). Aligning expectations about the adoption oflearning analytics in a brazilian highereducation institution. In *Proceedings of the Conference of Artificial Intelligence in Education*, pages 1–6.

Gulwani, S., Radiček, I., and Zuleger, F. (2014). Feedback generation for performance problems in introductory programming assignments. In *Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering*, pages 41–51. ACM.

Hattie, J. and Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1):81–112.

Higgins, R., Hartley, P., and Skelton, A. (2001). Getting the message across: the problem of communicating assessment feedback. *Teaching in higher education*, 6(2):269–274.

Hounsell, D. (2004). Reinventing feedback for the contemporary scottish university. In *Quality Enhancement Conference on Assessment, University of Glasgow*, volume 4.

Ivanic, R., Clark, R., and Rimmershaw, R. (2000). What am i supposed to make of this?: the messages conveyed to students by tutors' written comments.

Krusche, S. and Seitz, A. (2018). Artemis: An automatic assessment management system for interactive learning. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, pages 284–289. ACM.

Marin, V. J., Pereira, T., Sridharan, S., and Rivero, C. R. (2017). Automated personalized feedback in introductory java programming moocs. In *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, pages 1259–1270. IEEE.

Mello, R. F., Neto, R., Fiorentino, G., Alves, G., Arêdes, V., Silva, J. V. G. F., Falcão, T. P., and Gašević, D. (2022). Enhancing instructors' capability to assess open-response using natural language processing and learning analytics. In *Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption: 17th European Conference on Technology Enhanced Learning, EC-TEL 2022, Toulouse, France, September 12–16, 2022, Proceedings*, pages 102–115. Springer.

Nicol, D. J. and Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education*, 31(2):199–218.

Pardo, A., Bartimote, K., Shum, S. B., Dawson, S., Gao, J., Gašević, D., Leichtweis, S., Liu, D., Martínez-Maldonado, R., Mirriahi, N., et al. (2018). Ontask: Delivering data-informed, personalized learning support actions. *Journal of Learning Analytics*, 5(3):235–249.

Pardo, A., Dawson, S., Gašević, D., and Siemens, G. (2019). Supporting feedback processes at scale with ontask a hands-on tutorial. In *International Learning Analytics & Knowledge Conference 2019*, pages 285–288. Association for Computing Machinery (ACM).

Pereira, F. D., Oliveira, E. H., Oliveira, D. B., Cristea, A. I., Carvalho, L. S., Fonseca, S. C., Toda, A., and Isotani, S. (2020). Using learning analytics in the amazonas: understanding students' behaviour in introductory programming. *British journal of educational technology*, 51(4):955–972.

Ragupathi, K. and Lee, A. (2020). Beyond fairness and consistency in grading: The role of rubrics in higher education. In *Diversity and inclusion in global higher education*, pages 73–95. Palgrave Macmillan, Singapore.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science*, 18(2):119–144.

Santos, J. C. and Ribeiro, A. R. (2012). Jonline: proposta preliminar de um juiz online didático para o ensino de programação. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 1.

Tsai, Y.-S., Mello, R. F., Jovanović, J., and Gašević, D. (2021). Student appreciation of data-driven feedback: A pilot study on ontask. In *LAK21: 11th International Learning Analytics and Knowledge Conference*, pages 511–517.

Wiggins, G. (1998). *Educative Assessment. Designing Assessments To Inform and Improve Student Performance.* ERIC.

Winstone, N. E. and Carless, D. (2021). Who is feedback for? the influence of accountability and quality assurance agendas on the enactment of feedback processes. *Assessment in Education: Principles, Policy & Practice*, 0(0):1–18.

Yujian, L. and Bo, L. (2007). A normalized levenshtein distance metric. *IEEE transactions on pattern analysis and machine intelligence*, 29(6):1091–1095.