Tutoria: Supporting Good Practices for Providing Written Educational Feedback

Taciana Pontual Falcão[1], Verenna Oliveira[1], Samuel Souza[2], Giuseppe Fiorentino[1], José Rodrigues Neto[1], João Victor Galdino[1], Gabriel Alves[1], Rafael Ferreira Mello[1]

[1] Universidade Federal Rural de Pernambuco (UFRPE)

[2]Universidade Federal de Pernambuco (UFPE)

{taciana.pontual, rafael.mello, gabriel.alves}@ufrpe.br, sbjs@cin.ufpe.br,
{verenna.aredes, fiorentinogiuseppebcc, victor.jgfsilva}@gmail.com

Abstract. *In learning processes, feedback given by instructors is essential to guide students and help them improve from their mistakes. However, in higher education, instructors feel unable to give quality and timely feedback due to work overload and lack of time. As online classes became dominant due to the Covid 19 pandemic, and with increasing numbers of students per class, giving feedback beyond grades is even less realistic. Software tools to support feedback processes typically focus on automatic messages, which is not ideal as they lack personalization. Aligned with more recent research which suggests a broader perspective on the feedback process, we propose a software tool to help instructors construct quality written feedback efficiently. Through iterative user-centered design and applying artificial intelligence techniques, we developed functionalities that support correction of activities and allow building personalized written feedback, thus allowing instructors to give quality feedback to large groups, within realistic time frames.*

1. Introduction

Quality feedback in education is often overlooked. Yet, it is a fundamental element, as it can support self-regulated learning, by making students perceive clearly the gaps between their achievement and the competences they should develop [Wiggins 1998, Sadler 1989]. In the way the educational system is organized, instructors have a hierarchical and intellectual power which makes them the most obvious and reliable source of feedback in the eyes of students. Being a process of communication, feedback includes issues of discourse, identity, power, control and social relationship that must be taken into account by instructors [Higgins et al. 2001]. However, 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].

[Hattie and Timperley 2007] proposed that educational feedback is mainly related to: learning tasks, learning process, student self-regulation, and student motivation. [Nicol and Macfarlane-Dick 2006] 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; and encouraging positive motivational beliefs and self-esteem. Another crucial aspect, which often is beyond instructors' control due to work overload, is to give feedback shortly after the delivery of

DOI: 10.5753/sbie.2022.225074

the assignment. Ideally, students should have a second chance to submit their assignment, thus taking the advice that emerges from feedback to improve their work and their learning [Boud 2000, Hounsell 2004]. What typically happens in higher education, however, is that students move to the next assignment as soon as 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.

Instructors may agree with the importance of educational feedback, but the ideal feedback seems out of reach for two main reasons which are deeply connected: work overload and lack of time. In this sense, software platforms can support and facilitate the feedback process, especially since the use of virtual learning environments has expanded into in-person courses, and hybrid models are in expansion [Gulwani et al. 2014, Marin et al. 2017, Krusche and Seitz 2018]. Nevertheless, so far, most platforms that help giving feedback have focused on the automatic correction of multiple-choice assessments or on sending automatic messages from the instructor's correction, which lack personalization and thus 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].

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 being 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 structured as follows. In Section 2, we present the research method. In Section 3, we present the results of interviews with instructors and students. In Section 4, we present our platform Tutoria, followed by the results of the user evaluation. Section 5 presents the conclusions.

2. Method

In order to propose a software platform to assist instructors in giving educational feedback, we are following 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 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 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 interviews 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 a form with the 5-point Likert System Usability Scale (SUS). The results of all phases of design are discussed in the next sections.

3. Results from user research

3.1. Interviews

In the first phase of the design process, the aim was to better understand the process of assessment and feedback in higher education, and identify the main opportunities for developing a software solution that could meet the needs of instructors and students. Five categories emerged from the qualitative analysis of interviews: type of assessment; feedback format; feedback contents; feedback characteristics; and barriers for 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.

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

Table 1. Categories of analysis

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	

The analysis shows that assessment is done mostly through writing (with written assignments being the most popular form, and written exams the third). Accordingly, 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. So, as the instructors had to give written feedback mostly, they often write their comments directly on each students' assignments (typically a PDF file). However, many said that doing this to all students was infeasible. Workload was the most cited barrier for giving feedback, closely followed by the number of students per class; and lack of time. Of course, these three aspects are closely related.

As a result, the most common feedback content is grades. However, instructors try to add explanations to their feedback, when possible, pointing 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. Students were mostly dissatisfied with 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 multiple-choice 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.

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 were divided as to the educational 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 group or discuss results of exams, and found it useful for students to learn (also) from their peers' errors. Others worry that such situations might expose and make students uncomfortable.

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. Nevertheless, instructors feel they are unable to achieve it. Other characteristics of feedback less cited were: it should be two-way (from instructors to students and vice-versa), iterative and contextual. From their part, students expect to be continuously evaluated, receiving feedback iteratively and frequently throughout the course. Beyond grades, they would very much like to receive explanations, in particular about their errors or aspects to improve. They also mentioned the benefits of personalized feedback, although this seemed like a very distant scenario for them.

Given this scenario, where students desire better feedback, and instructors have a clear opinion about the importance of quality feedback, but are admittedly unable to attain it, both groups were divided about automatic feedback using software tools. Although instructors fear the lack of the human touch, and that feedback will not be personalized enough, they also admit the impossibility to deliver personalized 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. Although instructors currently use various software environments and communication tools (such as Google classroom, Moodle, Google colab, Trello, Excel, Telegram and Whatsapp), none of them are specific for giving educational feedback.

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.

3.2. 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 with instructors, by showing them the platform and asking for their opinions. Overall, instructors thought the platform to be an 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 numbers 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 to 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.

4. Development of the Tutoria prototype

4.1. Overview of the Tutoria platform

Tutoria is a software platform whose main goal is to help instructors compose written feedback for assignments. Assignments can be imported from Google classroom (and in the future, from 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.

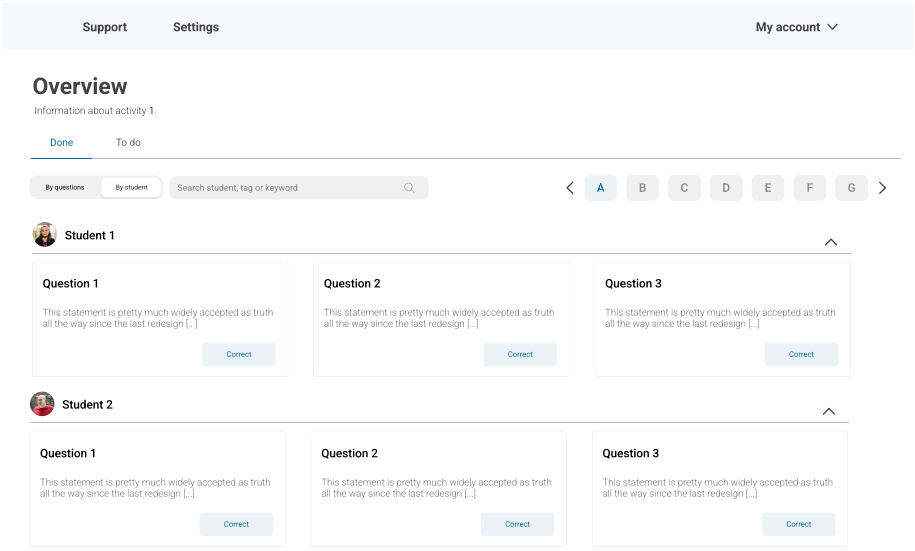

Figure 1. Home screen with students' assignments

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 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, such as semantic similarity, and accuracy of recommendations improve as the number of tagged excerpts increases. Another use of Artificial Intelligence (AI) techniques in Tutoria is plagiarism detection among students' answers, flagging to the instructor every case with

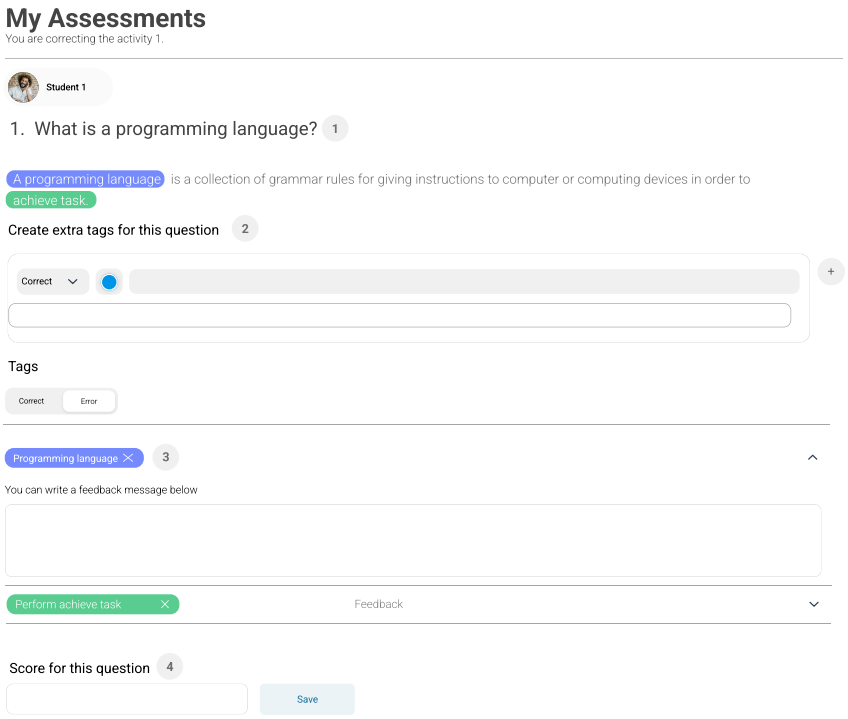

Figure 2. Assessment of an open-ended question.

similarity above a certain threshold (the default is 85%, but this can be configured by the instructor). 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 instruc-

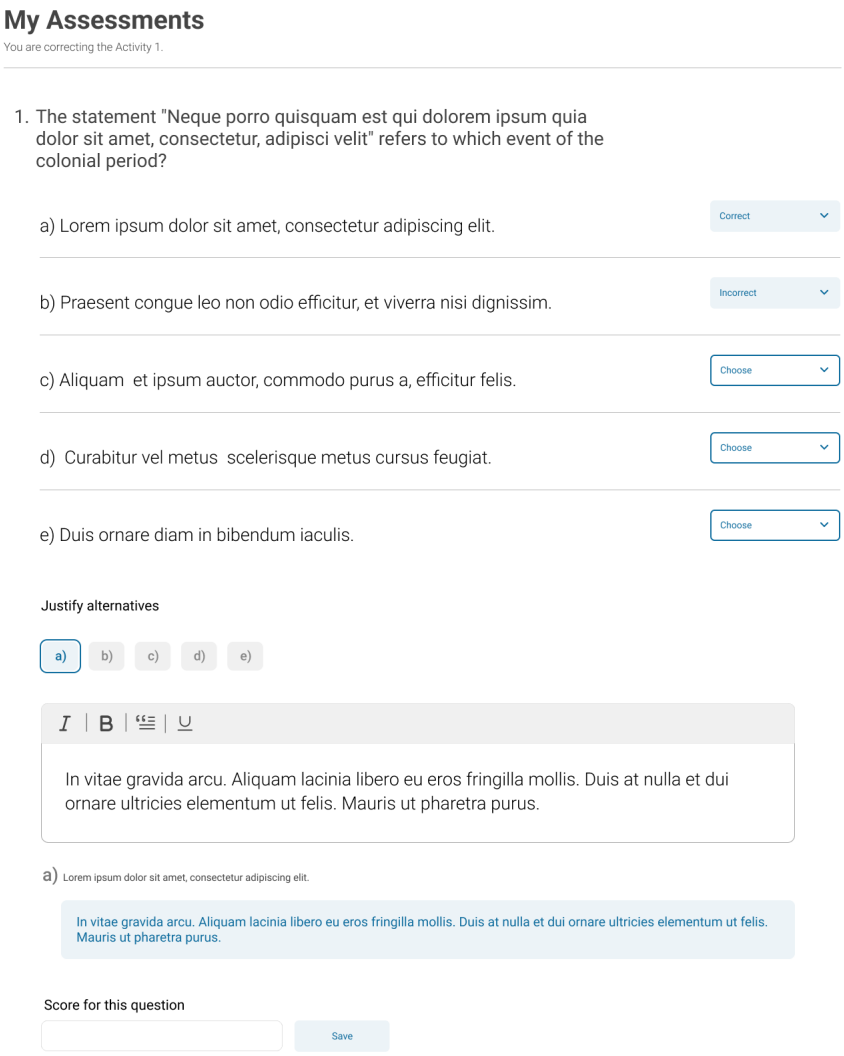

Figure 3. Assessment of a multiple-choice question.

tor can send the feedback messages to all students in a batch or send the message for a specific student only. Again, this gives the instructor autonomy in notifying students.

4.2. Results of the usability test

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

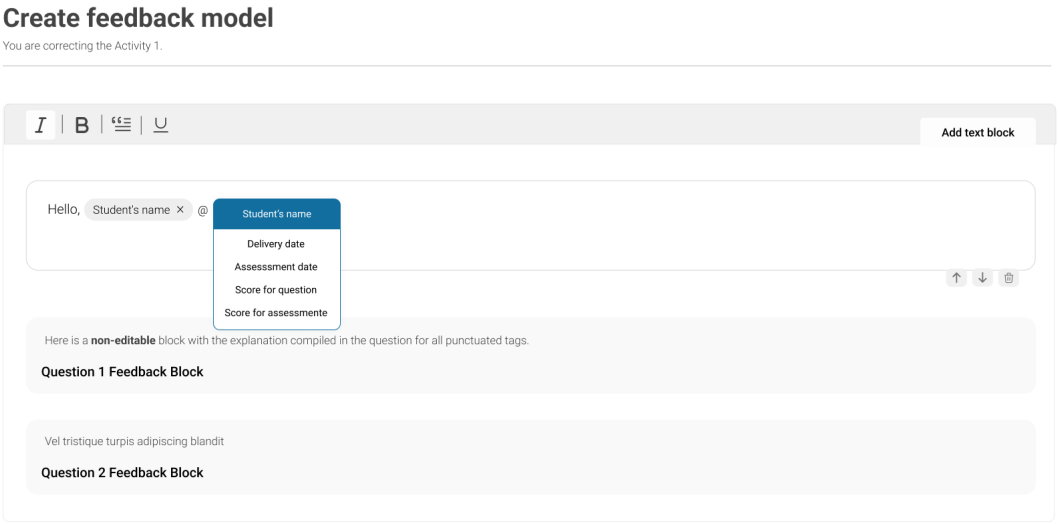

Figure 4. Construction of template for feedback message.

some thought the platform was unnecessarily complex (Q2), that technical help might be needed (Q4), and that the use was somewhat cumbersome (Q8).

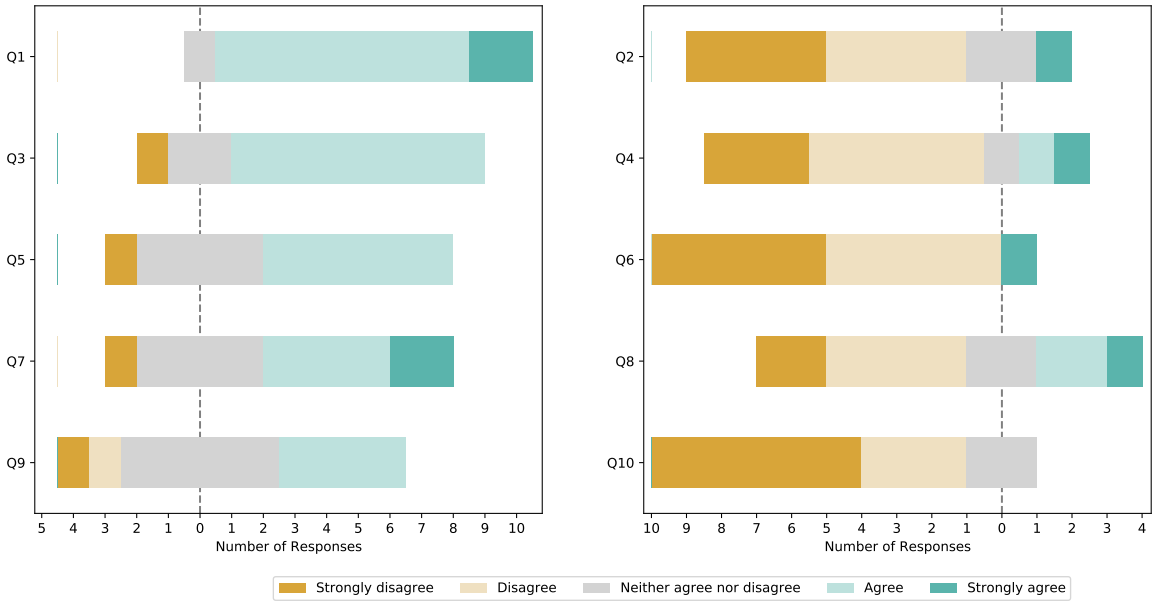

Figure 5. Results from System Usability Scale (SUS)

In addition to completing the questionnaire, instructors reported specific problems and gave suggestions for improvements. Regarding the correction process, the main problem was that none of the instructors understood the use of the extra tags, meant to provide 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 correction (through a filtered list in a pop-over window) - they found it confusing and did not easily understood what the interface was showing. Instructors also suggested having a third category of tag, besides right and wrong, which would be something in the lines of "partially right". Moreover, about 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 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).

5. Conclusions

Written assignments are the most popular ways used by instructors to assess students' learning. However, giving written feedback beyond grades is a very time-consuming task for instructors, often beyond their working capacity. The impossibility of providing quality and timely feedback, despite recognizing its importance for learning, causes frustration for the instructors while leaving students dissatisfied with the feedback they receive or its absence. In this paper, we presented 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 is being developed through a UCD process, collecting target users' opinions and improving the interface and functionalities iteratively.

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 proposed approach is based on the learning analytics process [Freitas et al. 2020] 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].

Future work includes refining the prototype based on the results of the usability test, doing another round of user testing, and collecting students' opinions about the feedback messages generated through Tutoria. Next, we will make Tutoria available for a group of instructors to use it with a real class and collect their feedback. 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.

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). Developing 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*.

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.

Freitas, E., Falcão, T. P., and Mello, R. F. (2020). Desmistificando a adoção de learning analytics: um guia conciso sobre ferramentas e instrumentos. *Sociedade Brasileira de Computação*.

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.

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.

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., 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).

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.

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.