Transparency in AI-driven Grading Tools for Open-ended Questions in Higher Education

Jelmer Hofman
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands
j.hofman-2@student.utwente.nl

ABSTRACT

Automated grading systems have been developed to support the grading process of open-ended questions in exams. However, these systems lack transparency on the student side. This study aims to investigate how much and what information students would like to receive with regard to automated grading tools, specifically with regard to EasyGrader, an AI-supported grading tool that is being developed by the University of Twente. In order to achieve this goal, a mixed methods research was conducted, consisting of a literature review, a survey and interviews. The literature review explores current practices for making AI systems transparent. The survey and interviews with students are conducted to gather their perspectives and preferences on transparency in automated grading tools. The survey and interviews revealed that students prefer to be informed about these systems before they take an exam, with a reminder during or after the assessment. Students also expect different types of information, including how the system works and how teachers are still involved in grading. These findings can lead to a better understanding of how to design transparent AI-supported grading tools that meet students' needs.

KEYWORDS

Artificial Intelligence, Transparency, Grading, Open-Ended Questions, Higher Education, Students, Human-In-The-Loop

1 INTRODUCTION

Artificial Intelligence (AI) has become increasingly relevant in the field of education. One of the areas where AI can provide a significant benefit is the grading of open-ended questions in exams [8]. The University of Twente has been developing EasyGrader, an AI-supported grading tool that can automate the grading of such questions.

Automated grading systems, such as the EasyGrader tool, have the potential to reduce the time and effort required for grading open-ended questions in exams. To increase the trust in these systems, it has been proposed to use hybrid systems that combine the strengths of human graders and automated tools [2]. Additionally, previous research has highlighted the importance of transparency in the use of AI-supported grading tools [13].

From previous research, it emerged that transparency in AI is important in order to ensure trust [5, 7, 10, 11, 14]. In recent years, there has been growing interest in the development of Explainable AI, which aims to provide insights into the inner workings of AI algorithms and decision-making processes. It is used to make the AI algorithm more transparent by showing what is happening inside the so-called 'black box'. By increasing the transparency of AI, explainable AI can help to promote trust and understanding among stakeholders, including students, teachers, and educational institutions [6].

Although there has been research about automated grading systems [2, 13], it has not been researched what information students expect about such a system. This research aims to address this gap by investigating the requirements for transparent AI-supported grading of open-ended questions in exams from the student's perspective.

The study employs a mixed-methods approach, including a literature review, a survey, and interviews. The literature review synthesizes existing knowledge on transparency in AI, with a focus on educational contexts. This helps to design the survey and interviews. The survey collects data on students' perceptions and preferences regarding transparency in the EasyGrader tool, while the interviews allow for a more in-depth exploration of students' experiences and perspectives.

2 PROBLEM STATEMENT

The use of automated grading systems, such as EasyGrader, can simplify the grading of open-ended questions asked during exams. However, these systems raise concerns about a lack of transparency for students. While previous research has shown that transparency is important, little is known about how much and what information should be provided to students in the context of hybrid grading systems that combine human and AI-supported grading (human-in-the-loop). Furthermore, it remains unclear how to effectively achieve transparency in automated grading systems. Therefore, the problem addressed by this study is to determine how much and what information students would like to get with regard to the EasyGrader tool.

TScIT 39, July 7, 2023, Enschede, The Netherlands

From the problem statement, the main research question was formed:

What are the needs and preferences of students regarding the transparency of an AI-supported grading tool for open-ended questions in exams?

The following sub-questions were formulated to answer the main research question:

1. What are the factors that make an AI-supported grading tool transparent from a student perspective?
2. What are the expectations of students regarding the transparency of AI-supported grading tools?

3 METHODOLOGY AND APPROACH

This section describes the steps that are taken to answer the research question. This research utilizes a mixed-methods approach, combining a literature review with a survey and interviews.

The literature review focuses on research related to the use of AI in education, particularly the use of AI-supported grading tools, as well as research on transparency in AI. The aim of the literature review is to identify best practices and strategies for achieving transparency in AI-based grading tools, with a particular focus on educational contexts. The literature review will also be used to develop the survey and interview questions.

A survey is conducted to collect data on students' perspectives and preferences regarding the level of transparency in the EasyGrader tool. The survey includes questions about what information students want to know about the grading process, how they prefer to receive this information, and when they want to be informed. The survey helps to get a wide range of views in a limited amount of time.

In addition to the survey, interviews are conducted with a small sample of students to gain a more in-depth understanding of their perspectives on the use of the EasyGrader tool and their preferences for transparency. The interviews are semi-structured and allow for a more nuanced exploration of students' experiences and perspectives. The survey and interviews are followed by an analysis of the data collected.

4 LITERATURE REVIEW

This section discusses the literature review. A systematic search was carried out to get an initial idea of the research. The following tools have been used: Scopus, SemanticScholar, Google Scholar, and Inciteful.

Before searching in these tools some keywords and initial queries were considered, e.g., ‘Artificial Intelligence’, ‘Transparency’, and ‘Automated Grading’. Once some articles had been found and evaluated, Inciteful was used to find more articles on similar topics. These keywords and queries were adapted depending on the results they gave.

There are two main research streams, namely transparency and explainability of AI and AI in education. The first stream is about AI in all different areas, while the latter is specifically focused on education. These streams will be described in the subsections of this chapter.

4. 1 Transparency and Explainability

In a study by Haresamudram et al. [9], it was proposed that AI transparency exists on three levels, namely Algorithmic Transparency, Interaction Transparency, and Social Transparency. Algorithmic Transparency is the openness of the code of the algorithm (the so-called “black box”), Interaction Transparency is about the interaction between the human and the system. Social Transparency is more about the ethical and social side of transparency, for example, privacy [9]. All of these levels of transparency play a part in an automated grading system.

According to research by Ehsan et al. [5] (social) transparency could help people to trust AI more. Other research has found similar results [7, 10, 11, 14]. Some research mentions that explainability provides transparency and therefore also helps to build trust [11], which is important in automated grading systems, as both students and teachers should be able to trust the system.

Explainable AI (XAI) is a concept that aims to provide insights into the workings of AI algorithms. It is used to make the AI algorithm more transparent by showing what is happening inside the so-called ‘black box’ [6, 15]. According to research by Xu et al. [15], there are two different important factors in Explainable AI, namely, ‘transparency design’ and ‘post-hoc explanations’. The first explains how the AI model works. It shows what is happening inside the ‘black box’, this is helpful for developers. The latter explains how the AI came to a conclusion and why that conclusion is correct, this is helpful for users. From that research, they also mention that “Explainable AI is important to the users who utilize the AI system” and it “is important to the people who are affected by AI decision” [15].

In research from Balasubramaniam et al. [1] it was found that explainability is an important part of transparency. However, they mention that there have not been many studies about transparency and explainability requirements [1]. From other research, it was also found that transparency and explainability are both important requirements of AI systems [3, 4].

4.2 AI in Education

It emerged that it is important to keep the human-in-the-loop when using AI-based grading tools, in other words, a human should be involved in the grading process. Therefore, a hybrid approach was suggested. It was also found that transparency is an important factor for these systems, both on the grader's side and on the student's side. However, what was not explored was what transparency would look like on the student side. One concern that has been raised was that too much transparency for students could lead to students trying to please the AI system [2, 13]. If, for example, the system

would work based on keywords, students might cheat the system by just naming a lot of different keywords, instead of actually answering the question.

It is mentioned by Mirmotahari et al. [12] that automatic feedback and transparency in the grading criteria contribute to the performance of students, as it helps them to understand what is expected from them.

5 SURVEY AND INTERVIEW DESIGN

In this section, the survey and interview design procedure is discussed. The purpose of the survey and interviews is to gather data easily and efficiently from the target group. The target group is defined as students from the University of Twente, as that is the main place the EasyGrader system will be used. The goal of the survey and interviews is to answer the second sub-research question.

The questions are based on the research objectives and the information that is needed to answer them. The survey length was intentionally minimized to reduce the barrier to respond, questions can be found in Appendix A. The questions are mostly semi-closed ended, meaning that they have a list of predefined options for the respondents to choose from. This is done to make the survey as short as possible so that more people fill it in. The list of options is carefully selected to avoid bias and to cover the most expected possible answers. As a pre-test were these options tested with some people with knowledge about the subject. In addition, there is an extra option for respondents to give their own answer. For every question, there is also an 'explain why' question. This question is not mandatory to fill in, but when people fill in such a question, it can give a better understanding of their reasoning. The survey also included some demographic questions, such as age, gender, study program, and study year. The survey was created online using Google Forms because it allows easy access and analysis of the data.

The interviews are semi-structured interviews that aimed to validate and deepen the results from the survey. The questions are mostly the same as the survey questions. However, instead of giving a list of options, they were mostly open-ended, meaning that they allowed the respondents to express their answers in their own words. This was also done to validate the list of options in the survey and to get more information that might not have been captured by the survey.

The survey was distributed to the target group, consisting of students from the University of Twente. The survey link was sent to several WhatsApp and Discord groups. In addition, the researcher went around the University campus to ask students to fill in the survey. After filling in the survey, the respondents were asked to send it to other students. It is important that participants have experience with open-ended questions on exams, as the automated grading system will be used on this type of question.

Next, potential interviewees were contacted via WhatsApp or Discord and asked if they were willing to participate in an interview. If they agreed, a suitable time and place were arranged for the interview. These interviewees were also asked to send the survey around to fellow students, however, they did not need to fill it in themselves, as the survey and interview questions overlap.

A request was made to the Ethics Committee of the University of Twente before conducting the survey and interviews. Before doing an interview, the participant was given an informed consent form, accompanied by an information sheet. This contains all information needed by the participants about the research. For the survey, an opening statement was put into Google Forms, with participants needing to check a box in order to give consent.

6 RESULTS

In this section, the results of the research are described. The literature review answers sub-research question 1 and the survey and interviews answer sub-research question 2.

6.1 Literature Review

The literature review showed that transparency is closely related to explainability. Several studies have suggested that transparency can increase user trust and acceptance of artificial intelligence systems [5, 7, 10, 11, 14]. Some of them have also argued that explainability provides the necessary transparency for users to trust AI systems [11].

Another important finding from the literature review was that a hybrid approach, which combines human and AI input, is preferable for automatic grading systems. This is because human-in-the-loop systems can leverage the strengths of both humans and AI-tools. It also emerged that transparency is an important factor for both teachers and students in automated grading systems.

However, a gap was identified in the literature regarding how to achieve transparency in automatic grading systems and what are the best practices and guidelines for designing transparent grading systems on the student side.

6.2 Survey & Interviews

In this section, the results from the survey and interview questions are described. The survey questions can be found in Appendix A. The interview questions are similar to the survey questions, but instead of using semi-closed questions, the interview consists of open-ended questions. In this section, the responses from both are combined. First, the data was cleaned, meaning that the data from the interviews were merged into the data from the survey.

In the end, 25 students completed the survey and 9 interviews were conducted. Thus, a total of 34 responses were collected in this research. The results from the survey and the interviews are combined in this section as both groups answered the same questions. However, it should be noted that the interview questions were open-ended, while the survey questions were semi-closed, consisting of multiple choice with an additional open-ended question to provide an explanation.

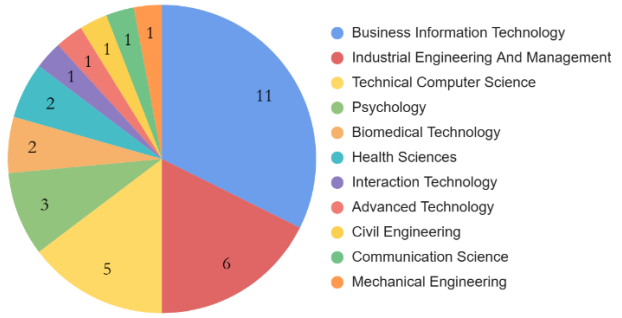

Figure 1. Studies

Students from a diverse range of studies have responded to the survey and interview, as can be seen in Figure 1. It can be observed that Business Information Technology, Industrial Engineering and Management, and Technical Computer Science have a share of 64% of the total responses. On the other hand, there are several studies where only one student each responded, namely Interaction Technology, Advanced Technology, Civil Engineering, Communication Science, and Mechanical Engineering.

Of the 34 participants, most are in their third bachelor year, namely 21 students, see Figure 7 in Appendix B. There are also responses from first- and second-year bachelor students, as well as one first-year master's student. The participants are in the age range of 18 to 26, although most respondents are in the range of 20 to 22. The other age groups all have one response. Other demographics can be found in Appendix B.

6.2.1 Do students want to be informed about automated grading?

The first question of the survey and interview is about whether students actually want information about the use of automatic grading systems on their exams. Most students responded agreeing to this question. This question uses a Likert scale ranging from 'Strongly disagree' to 'Strongly agree', the mean of the answers is 4.26 and the standard deviation is 0.710. This shows that the majority of students agree that they want to be informed when EasyGrader would be used on one of their exams (see Figure 2).

It can be observed that one student responded with 'Disagree' and two students were neutral on this question. Two of these students filled in the open explanation question accompanied

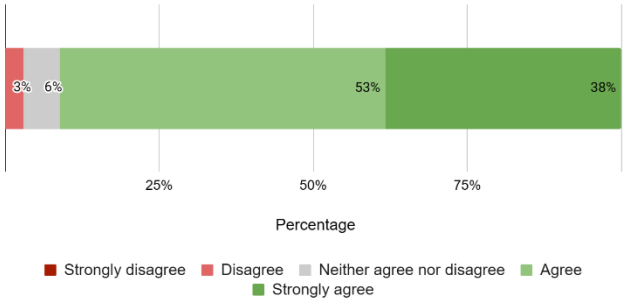

Figure 2. Result of whether students want to be informed about the use of EasyGrader

by the Likert question. One of them said: "*I don't care if the examiners use or do not use it, as long as my grade is correct*". The other said something similar.

As mentioned before, the majority wants to be informed. One student mentioned the following: "*I would like to know who is grading something so as to clear up possible miscommunications about the intended answers. This is especially relevant for open questions given that the answers can be wrongly interpreted*". Other students also mentioned that they find it important to know whether an AI is grading them. One student also mentioned that besides knowing whether an AI is grading them, this person thinks it is also important to know if a teaching assistant did the grading: "*(...) The same is true when TA's grade my exam; they'll probably do fine, but I want to ensure there wasn't an inconsistency between the graded exams*".

Some students gave another reason for being informed to know if EasyGrader would be used, they say that they would be more careful during the review moment. One student said: "*If I know AI is used, I will look at the grading in more detail since I don't fully trust that the AI can grade accurately*". Another student thinks that being graded by a teacher is important because it might give the teacher valuable insight into the students' abilities: "*It will definitely influence my grade, furthermore I believe it is good for teachers to get insight into what students do wrong and why their argumentations are not of good quality. I believe that if teachers grade the test by themselves they gain way more insight into the students' abilities*"

6.2.2 What information do students expect?

Now it was identified whether students want to be informed about the use of automatic grading systems on open-ended questions, but it is still unclear what information they want.

In the survey, the question about what information students expect was a semi-closed question. The reason for this was that it might be unclear what students could answer if there was no information given. By giving them a list of options and an extra open option, they could get an idea of what information is possible. Respondents could check multiple options. This list of options was not given during the interviews, as during an interview it is possible to ask for clarification if a question is unclear. From the interviews, a similar view occurred as from the survey. Students in interviews gave similar answers despite not having the options available. In Figure 3 the responses can be found.

It can be observed that most students want to know how the system would work (88%). The role of the teacher or teaching assistant (TA) and the calculation of the grade are also mentioned by 73% and 70% respectively. 61% of the participants expect to get information on what questions the automatic grading system was used. Only half of the respondents want to know what kind of feedback they could receive from the system, and 29% or 10 students want to know the reason(s) why the system is used at all.

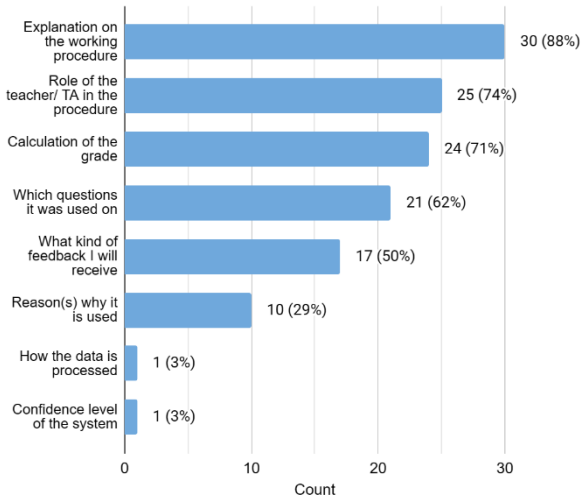

Figure 3. Results of what information students expect

On average 3.76 out of 6 given options were checked. There were seven (21%) students who only checked one or two options, and there were 11 (32%) students who checked five or six given options. Someone who checked all options gave as a reason: "*Transparency!*", another student wrote the following: "*I would want to know as much as possible about this tool, to be able to adapt to it in the best way possible*". A student that checked two options ('Explanation on the working procedure' and 'Role of the teacher/TA in the procedure') gave the following explanation: "*One general story about EasyGrader. But documentation with more information if you want to learn more. For example, in that general story there is a link to a website with documentation or more information*".

From the interviews, it was mentioned by one student that it might be useful to know how the data is processed. In this regard, it was meant that it would be explained what happens with the answers students give to questions on exams, for example, whether they are used to train the system. Another student said that he wants to know how confident the system is in the grading.

6.2.3 How to distribute information?

Information about automated grading systems can be provided to students on multiple occasions during a module or period. The first option is to give the information before an exam takes place, for example, at the start of a module, or when the subject is taught. Another option is to provide the information during an exam, this can be on the front page of the exam, or at the question. The last option is after an exam, this can, for example, be on the day the grade is received or during the review moment. This was a semi-closed multiple-choice question on the survey where multiple options could be selected. There was also an extra open option for respondents' own input. During the interviews, this question was asked as an open question.

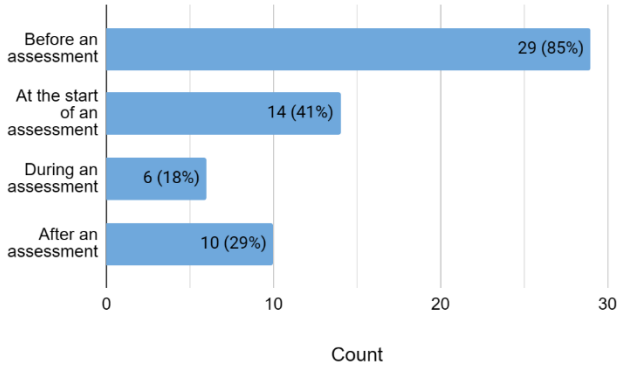

Figure 4. Results of when to be informed.

From the results of this question, see Figure 4, it can be observed that most students want to have the information before an assessment, namely 85%. Of the participants that checked this box, one gave the explanation: "*I don't like surprises, it should be clear beforehand how the tests will be graded*".

The option 'During an assessment (e.g. at the question in Remindo)' was checked the least amount of time, and it was never checked as the only option. Multiple participants gave an explanation as to why they do not want to be informed during an exam: they say that it might distract them. One student thinks that giving the information during the exam might lead to people altering their answers, and he also thinks it is distracting. However, he came up with a possible solution for people who do not find it distracting, he said: "*Maybe this is something to be able to turn on and off via a setting in Remindo by students. So that students can choose for themselves*". Only six students (18%) selected the 'during the assessment option.

Almost half of the students (47%, 16 people) checked only one box, of which 12 people selected 'Before an assessment'. Not one student checked all boxes, however, three students (24%) checked three boxes. Some of these students gave explanations on why they checked multiple boxes. One student said: "*Being informed beforehand would be sufficient on how you would be graded. However, it would be a nice reminder to have information at an open question in an exam*". Other students that checked multiple boxes also mentioned that it is good to have a reminder during and/or after the exam.

6.2.4 Risks

The last question in the survey was about the risks of being too transparent with an automated grading system. As mentioned in the literature review, a previous study found that teachers and teaching assistants think students might start pleasing the system if they know how the automated grading exactly works [13]. This question's aim was to find out whether students think the same thing. This question is also a Likert Scale question, with 'Strongly disagree' having a value of one and 'Strongly agree' having a value of five. The mean of this question is 3.29, and the standard deviation is 1.142. This shows that, on average, students neither agree nor disagree with the statement.

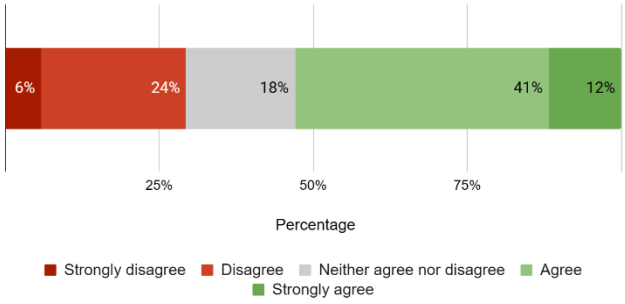

Figure 5. Results of whether students think there are risks about being too transparent

It can be seen in the chart in Figure 5 that students are quite divided on this statement. Some people see the risks, for example, one student said: "*Students can try to answer in such a way that it might covers the keywords in order to get the points*". Some other students agree that there are risks, but still think that the system should not be used if there are any risks, one of them answered the 'explain why' question: " *I understand that it could be tricky to be transparent because of possible misuse of the system. However, I feel like a system like EasyGrader should only be used in the first place if it is robust enough to withstand misuse*".

People who disagree with the statement also say that if the system would have any risks of misuse, it should not be used. One student who disagreed with the statement wrote the following: "*As long as the teacher is still in control of checking the answers and the suggested grading, I feel like the AI is tool to make the grading more efficient without lowering a teacher's ability to maintain fairness*". This answer mentions the hybrid approach that was also suggested by previous research [2, 13]. One student mentions that to misuse the system it is still required to have knowledge on the topic: "*That's the trade-off, the more info, the better someone can direct their answers to please the system. But you still need to have the knowledge to answer a question. If you have the capacity to answer a question, it shouldn't matter*".

Overall, it can be observed that students are divided on this topic, however, most students mention that if the system is not sound enough, it might be possible to misuse the system by giving answers that the algorithm is looking for.

6.2.5 Feedback

During the first few interviews, one theme came to the surface each time. The topic mentioned by the interviewees was feedback. They think that an automated grading system would be useful to give feedback to students on open-ended questions. After the first couple of interviews in which this topic came up, it was decided to ask an additional question in the other interviews, which was "What do you think such an automatic grading system would do with feedback?".

Most students reported that they consider feedback in open-ended questions important, especially to know what they did well and what they did not so well. Through the feedback, they know how to get it right the next time.

When asked if they think an automated system like EasyGrader could provide feedback, they responded positively. Although some students do not know if they would trust feedback from a system like EasyGrader right away, it could be a benefit. Currently, according to the interviewees, there is little to no feedback from tutors on exams, a system like this should be able to provide generic feedback. One interviewee said the following: "*Feedback is already very poor because teachers have little time. AI could possibly give general feedback that is useful*".

7 LIMITATIONS

One of the limitations of this study is the sample of students who participated in the survey and interviews. Most of them were from technical studies, such as Business Information Technology, Industrial Engineering and Management, and Technical Computer Science, while other studies had only one, two or three students who responded. These studies may have different expectations and preferences for automated grading tools than other disciplines. Therefore, the results may not be generalizable to the student population as a whole. It would be interesting to conduct a similar study with more students from other disciplines.

The second point is that the survey and interviews were conducted at one university, which may also affect the generalizability of the results. This university may have a particular culture or environment that influences how students perceive and interact with such tools. For example, the university may have a high level of trust or support for its teachers and staff, which may reduce the need or demand for transparency in automated grading tools. It would be interesting to conduct this research with more students from different universities and programs.

The study used a mixed-methods approach, combining a survey and interviews. The survey contained semi-closed ended questions that gave respondents a list of options to choose from. This may have introduced some bias into the data collection, as the options may not have reflected the full range of possible responses; to counter this, there was an extra open option in the list for respondents to fill in if they had other suggestions. However, the interviews were conducted with open-ended questions that allowed participants to express their thoughts freely. Analysis of the interview data showed that interview participants gave answers that were consistent with the options in the survey, suggesting that the survey was valid and reliable. Nevertheless, future studies could use more open-ended questions in the survey to capture more nuanced and diverse responses from the students.

8 CONCLUSION

This section presents the main findings and conclusions of this research, which aimed to investigate students' preferences and expectations regarding transparency in automated grading tools for open-ended questions. The research methods included a survey and interviews with students at a university in the Netherlands. In addition, this section suggests potential future research on the topic of automated grading systems and/or transparency in AI.

Students mainly want to be informed about automated grading systems before an assessment, rather than during or after an assessment. This suggests that students value transparency and want to know what to expect from the grading process. However, some students also expressed a desire to be reminded of the use of automated grading systems during or after the assessment, especially if they have doubts or questions about their grades. Some students mention that they will pay more attention during the examination if an automated system has graded them.

Furthermore, students expect different types of information when they are informed about the automated grading system. The most common types of information that students wanted to be informed about were how the system works, how it calculates the grade, and what the role of the teacher and/or teaching assistant is.

The main conclusion of this research is that transparency is important to students when it comes to automated grading systems for open-ended questions. Students want to be informed about the use of these systems and expect clear and comprehensive information about how they work and how they affect their grades. It is therefore recommended that educators and developers of automated grading systems take these preferences and expectations into account when designing and implementing such tools, and communicate effectively with students about them.

While these findings can lead to a better understanding of how to design transparent AI-supported grading tools that meet students' needs, there is still more research possible. One possible direction for future research is to investigate whether students perceive any risks or disadvantages associated with too much transparency in automated grading tools for open-ended questions. This was a small part of the survey and interviews, and students were quite divided on this statement. It would be interesting to explore if and how students would change the way they answer open-ended questions once automated grading systems are in use.

Another possible area for future research could be to design and evaluate different user interfaces for presenting transparent information in automated grading tools. For example, how can visualization help students understand how their answers will be graded? How can interactive features help students to ask questions or get feedback on their answers?

REFERENCES

[1] Balasubramaniam, N., Kauppinen, M., Rannisto, A., Hiekkanen, K. and Kujala, S. 2023. Transparency and explainability of AI systems: From ethical guidelines to requirements. *Information and Software Technology*. 159, (Jul. 2023), 107197. DOI:https://doi.org/10.1016/j.infsof.2023.107197.

[2] Braun, D., Rogetzer, P., Stoica, E. and Kurzhals, H. 2023. Students' Perspective on AI-Supported Assessment of Open-Ended Questions in Higher Education. *Proceedings of the 15th International Conference on Computer Supported Education* (2023), 73–79.

[3] Chazette, L., Brunotte, W. and Speith, T. 2021. Exploring Explainability: A Definition, a Model, and a Knowledge Catalogue. *2021 IEEE 29th International Requirements Engineering Conference (RE)* (Sep. 2021), 197–208.

[4] Cysneiros, L.M., Raffi, M. and Sampaio do Prado Leite, J.C. 2018. Software Transparency as a Key Requirement for Self-Driving Cars. *2018 IEEE 26th International Requirements Engineering Conference (RE)* (Aug. 2018), 382–387.

[5] Ehsan, U., Liao, Q.V., Muller, M., Riedl, M.O. and Weisz, J.D. 2021. Expanding Explainability: Towards Social Transparency in AI systems. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. (May 2021). DOI:https://doi.org/10.1145/3411764.3445188.

[6] Explainable AI (XAI) | IBM: *https://www.ibm.com/watson/explainable-ai*. Accessed: 2023-05-04.

[7] Galassi, A. and Vittorini, P. 2021. Automated feedback to students in data science assignments: Improved implementation and results. *ACM International Conference Proceeding Series*. (Jul. 2021). DOI:https://doi.org/10.1145/3464385.3464387.

[8] Haller, S., Aldea, A., Seifert, C. and Strisciuglio, N. 2022. Survey on Automated Short Answer Grading with Deep Learning: from Word Embeddings to Transformers. (Mar. 2022).

[9] Haresamudram, K., Larsson, S. and Heintz, F. 2023. Three Levels of AI Transparency. *Computer*. 56, 02 (Feb. 2023), 93–100. DOI:https://doi.org/10.1109/MC.2022.3213181.

[10] Hois, J., Theofanou-Fuelbier, D. and Junk, A.J. 2019. How to Achieve Explainability and Transparency in Human AI Interaction. (2019), 177–183.

[11] Miller, T. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*. 267, (Feb. 2019), 1–38. DOI:https://doi.org/10.1016/J.ARTINT.2018.07.007.

[12] Mirmotahari, O., Berg, Y., Gjessing, S., Fremstad, E. and Damsa, C. 2019. A Case-Study of Automated Feedback Assessment. *2019 IEEE Global Engineering Education Conference (EDUCON)* (Apr. 2019), 1190–1197.

[13] Stoica, E. 2022. A student's take on challenges of AI-driven grading in higher education. (2022).

[14] Tan, S.H.S., Thibault, G., Chew, A.C.Y. and Rajalingam, P. 2022. Enabling open-ended questions in team-based learning using automated marking: Impact on student achievement, learning and engagement. *Journal of Computer Assisted Learning*. 38, 5 (Oct. 2022), 1347–1359. DOI:https://doi.org/10.1111/jcal.12680.

[15] Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D. and Zhu, J. 2019. Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges. 563–574.

APPENDIX A: SURVEY QUESTIONS

1a. If EasyGrader is used for an exam, I want to be informed about that.

- Strongly disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly agree

1b. Please explain why.

2a. When you get information about the use of EasyGrader, what information do you expect?

- Calculation of the grade
- Explanation on the working procedure
- Reason(s) why it is used
- Role of the teacher/TA in the procedure
- What kind of feedback I will receive from EasyGrader
- Which questions it was used on

2b. Please explain why.

3a. How do you want to be informed about the use of EasyGrader?

- Before an assessment, in the module manual
- In the information sheet at the start of an assessment (e.g. front page of Remindo)
- During an assessment (e.g. at the question in Remindo)
- After an assessment (e.g. during the review moment)

3b. Please explain why.

4a. Do you agree with the statement that there are risks about being too transparent with such a tool?

- Strongly disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly agree

4b. Please explain why.

APPENDIX B: SURVEY & INTERVIEW DEMOGRAPHICS

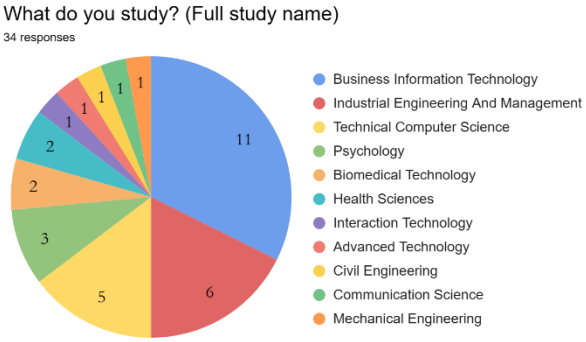

Figure 6. Studies

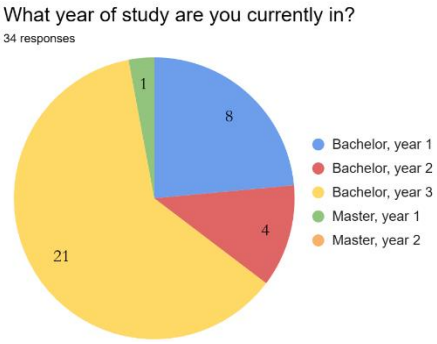

Figure 7. Study years

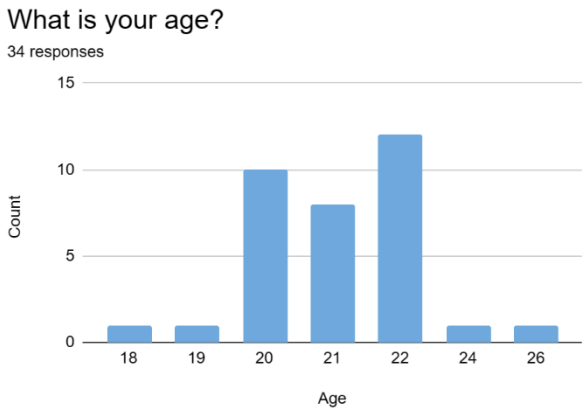

Figure 8. Ages of participants

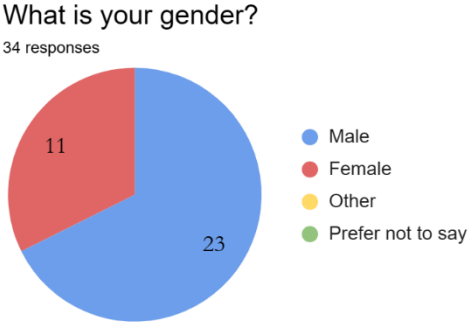

Figure 9. Gender of participants