

A Triple Challenge: Students' Identification, Interpretation, and Use of Individualized Automated Feedback in Learning to Write English as a Foreign Language.

Ingeborg Krange¹, Meerita Segaran¹, Siv Gamlem², Synnøve Moltudal²
Irina Engeness¹

¹ Department of Pedagogy, ICT and Learning, University College Østfold, P.O. Box 700, NO-1757 Halden, Norway

² Department of Humanities and Teacher Education, University College Volda, P.O. Box 500, NO-6101 Volda, Norway
{ingeborg.h.krange}@hiof.no

Abstract. The aim of this study was to investigate eighth-grade students' assessment literacy and writing skills in English as a foreign language using an AI-based automated essay assessment tool (EAT). Data were gathered from a design-based research initiative where the EAT was designed, developed, and tested in naturalistic school settings. Fifty-six eighth-grade students wrote individual essays, for which they received automatic feedback. The feedback was discussed with their teachers and peers. Both the writing process and teacher and peer interactions were video recorded. The video data were analyzed using an interaction analysis. The improvements made on the essay based on the feedback logs registered by the EAT for each student's writing trajectory and the different versions of the essay were examined using frequency analyses. The findings demonstrate that automated essay assessment might be useful for fostering students' writing skills if teachers help students get started, identify errors, and share interpretations.

Keywords: Assessment for learning, sociocultural interpretation of learning and teaching, design-based research, interaction analysis, frequency analysis, artificial intelligence (AI), automatic essay assessment, junior high school

1 Introduction

Research has documented that formative assessments seem to enhance students' learning processes [1-4]. Formative assessment is a broad form of assessment that is integral to everyday learning and has been conceptualized as embedded in pedagogy to develop a deeper understanding of how assessment is used as a pedagogical tool [1]. This assessment is comprehensive, as it involves fluid transitions between assessment, learning, and guidance while supporting students through tailored feedback. This feedback initially focuses on areas where students require more

guidance and calls for a systematic and consistent learning process based on dialogues between teacher and student, with continual assessment as an integral element [3-6]. This requires a consistent approach to testing or tight collaboration between teachers and students during problem solving, with teacher-initiated follow-up procedures and a considerable degree of student involvement from the outset. Teachers' mindful pedagogical use of formative assessment can foster students' learning processes and beliefs in their own capabilities to succeed [1, 5, 7].

However, studies have indicated that formative assessment practices are not easily implemented in real-life teacher-student feedback [8, 9]. These practices are demanding for teachers to complete because of the high number of students in each class and limited time and particularly challenging in classes where the topic is process-oriented writing [10]. This challenge brought to attention the question of whether individually tailored artificial intelligence (AI) technology could be a productive automatic essay assessment tool (EAT) to supplement teachers' feedback to students during their writing processes [10, 12]. Thus, we investigated how such technology could help foster students' assessment literacy [1, 13] and writing skills in English as a foreign language (EFL) [14]. Previous comparative studies have documented promising results and reported that students using such technology may produce essays richer in content than those who do not [15].

This article contributes to the understanding of how an AI-based technology designed and developed to foster formative automated essay assessment can be used in a naturalistic school setting where students' use this technology to improve their writing skills in EFL. We arranged for this design, development and use as part of a design-based research (DBR) project [16, 17] where our interdisciplinary group of researchers designed and developed EAT that not only promotes the correction of students' language errors but also stimulates their examination of the rationale behind the errors and prompts a deeper understanding of the writing process. Thereafter, we introduced this technology as an automatic formative assessment tool to eighth-grade students and their teacher during a writing trajectory in an EFL setting in three junior high schools. More specifically, we investigated how this technology supported the students' writing trajectories, assessment literacy and writing skills in EFL in light of the triple challenge presented to them: having the ability to identify relevant feedback, the competence to interpret this feedback, and the capability to use the feedback to improve the quality of their essays. For this purpose, we analyzed video recordings of how the students, together with their teachers, identified and interpreted individually tailored automated feedback to improve the quality of their essays throughout their personal writing trajectories [18], from the first draft to the submission of the final version of the essay. Moreover, we also analyzed log data gathered from the EAT to determine the frequencies of each feedback the students' received, and we examined how the students used this feedback or whether they used it at all to improve the essays they submitted at three separate times during their writing processes. To our knowledge, no similar studies have followed junior high school students' writing trajectories from the beginning of writing to the submission of the final version of the essay while using AI-based feedback technology. Both the teachers and students used the first prototype of the EAT that offered students individually tailored automated syntactic feedback. In this paper, we narrowed our analysis to the writing mechanics, focusing specifically on punctuation and capitalization. Limiting the analysis to these

two aspects allowed for a more focused and in-depth examination of the students' uptake of these elements.

The following research questions were addressed in this study:

- How did the students identify and interpret the individualized automated feedback provided by the EAT throughout their writing trajectories?
- How did the teachers facilitate the students' writing processes throughout their writing trajectories, and how did they build on or supplement the EAT?
- How did the students use individualized automated feedback to improve their essays throughout their writing trajectories?

To address the above-mentioned research questions, we used an interaction analysis [19] to study the interactions between the students and their teachers as they collaboratively participated in identifying and mutually interpreted the individualized automated feedback. We also used the log data of the EAT-generated automated feedback to evaluate how each student responded to the feedback and used the students' individual essays to calculate the frequency of the feedback each student used to improve their essays. Based on these analyses, we discuss the relationship between the EAT and how the teacher exercised formative assessment and the students' assessment literacy [13] and writing skills in EFL [14, 15]. The study findings will be used to improve the current understanding of formative assessment and the use of individually tailored AI technology to inform the future versions of the EAT and the design of upcoming writing trajectories.

2 Assessment and Feedback as Support for Learning

Numerous studies have indicated the importance of involving students more actively in the feedback and assessment processes of learning trajectories, which is aligned with the core principle of assessment for learning (AfL) [1, 20, 21]. The aim of including students in assessment activities is to enhance their learning and develop their understanding and assessment literacy. To better understand freshman students' assessment literacy, Smith et al. [20] conceptualized this kind of literacy by delineating it into three essential dimensions. First, students must understand the purpose of assessment and its intricate connection to their writing trajectories. Second, students should possess a heightened awareness of the assessment processes and their potential implications for their ability to successfully submit assignments. Finally, students must develop the ability to critically evaluate their own work and discern strategies for improvement. In our study, we investigated whether these assessment literacy principles might also be useful for understanding and discussing students' assessment literacy at the junior high school level to identify their greatest need.

Involving students actively in assessment activities is perceived to foster learning when learners are placed at the center of their educational journey, promoting their empowerment, and nurturing their sense of ownership and responsibility for their own learning [23]. However, recognizing the need to include students' perspectives and

voices in the assessment process, also challenges the traditional perspectives of teachers as the sole assessor and provider of feedback. Many researchers have documented the importance of feedback practices in promoting students' assessment literacy [5, 13, 24]. One of the most significant benefits of feedback is its ability to enhance assessment literacy among learners. Through constructive feedback, students can gain a deeper understanding of the evaluation criteria and standards. This knowledge empowers them to evaluate their own work critically, identify areas of weakness, and take measures toward improvement [1, 3, 20, 25]. As their assessment literacy increases, students become active participants in their own learning processes, elevating their overall academic performance. In the present study, we examined how students participated in the identification and interpretation of the individually tailored automated assessment and feedback offered by the EAT.

Providing instructional feedback has proven effective for addressing the quality of students' writing [9, 26-28]. Feedback that identifies concrete areas for improvement fosters students to further develop their written outputs [3, 5]. This kind of feedback affords close collaboration between the teacher and the students to identify and develop a shared understanding of what needs to be improved. However, offering individually tailored feedback can be challenging in classroom settings. Teachers often struggle with a lack of time and the ability to cater to each individual due to the huge number of students in the classroom [10, 15]. Research on educational technologies aimed to personalize the learning trajectories of students has consequently emerged during the past decades, and studies exploring real-life implementation of e.g., adaptive learning technologies has demonstrated that both students and teachers vary in their ability to make use of both integrated feedback loops and additional activity data in everyday contexts [29-30]. In general, research on educational use of digital technologies seem to provide little support for teachers' facilitation for learning and is often reduced to pointing out that specific technology has potential for leaning rather than showing how the potential plays out in practice [31]. Deepening and broadening the knowledge of how to integrate the use of such automated feedback loops and activity data in holistic learning processes are therefore much needed both regarding how students and teachers partake in these learning and teaching settings. Studies have reported that feedback on students' writing tends to consist mainly of praise; consequently, there is an absence of feedback on how to proceed with learning [25, 32-33]. This uninformative practice could be a product of lack of time but can also be considered a larger cultural challenge in schools. With this as a backdrop, it is intriguing that promising results have been documented from the adoption of individually tailored AI technology aimed at optimizing teachers' efficiency and giving them a better starting point in conversations with their students [34]. We aimed to examine how both teachers and students collaboratively establish a mutual comprehension of areas for improvement in students' written work. In addition, we investigated whether teachers could engage with individual students while utilizing a personalized automated assessment tool. We also assessed whether teachers offered formative assessments that extend beyond mere praise instead of providing constructive guidance for students in refining their essays.

In summary, the field of formative assessment has a well-established presence, encompassing insights into this type of assessment, which is aimed at enhancing students' writing skills. Nevertheless, our assertion is that despite the promising

outcomes of the implementation of AI-driven technology tailored to aid formative assessment and foster students' writing skills in EFL, there exists a significant gap in essential knowledge within this research domain. Thus, this study was conducted to comprehensively examine the following: first, the complete writing trajectories, from the beginning to the end, of a small cohort of students in collaboration with their teachers; second, the role of EAT in this process, including how the teacher and the students as interlocutors jointly identified and interpreted the individualized automated feedback and explored the issues discussed; and finally, the automated feedback they received were subsequently incorporated into their essays. In this way, we aimed to contribute to the emerging understanding of how students' triple challenge in identifying, interpreting, and using individually tailored automated feedback can be supported.

3 Research Design

3.1 A Design-Based Research Initiative

This study was a design-based research initiative [17] organized as a design experiment [16, 35]. Design experiments aim “to engineer innovative educational environments and simultaneously conduct experimental studies of these innovations” [16, p. 141]. In our case, a team of researchers created an innovative learning tool known as the “essay assessment tool, EAT.” This tool was designed and developed with the primary aim of enhancing students' assessment literacy and English writing skills as they learn EFL. Our focus extended to the practical application of this technology in naturalistic educational settings, where we closely examined how students' writing trajectories, guided by their teachers, unfolded while they identified and interpreted feedback provided by the EAT, which they could use to improve the quality and content of their essays. This design, development, and testing were the first iteration of the first prototype of EAT.

3.2 EAT Technology

The EAT is an individually tailored automated tool designed and developed to enhance students' writing processes. It has different functionalities, of which the most relevant to the present study is writing mechanics focused on errors such as spelling, punctuation, missing commas, and capitalization. The feedback does not give the right answer but provides feedback that the text needs to be improved and how it should be revised. Thus, the feedback indicates for the students that a specific word or

sentence has “a problem” that should be addressed. In this way, the EAT differs from many other AI tools such as Grammarly. The latter technology predominantly focuses on suggesting changes to text and necessitates user acceptance or rejection of those changes. By contrast, the EAT goes beyond an approach of static corrections and rather introduces developmental feedback aiming to foster reflection and learning about the rationale behind the errors made, promoting a deeper understanding of learning to write.

As shown in Figure 1, the interface is divided into two parts. On the left side is one of the student's texts, and on the right side is a feedback list. By providing this feature, students can view and analyze their feedback alongside their original essays. The possible errors are marked in red.

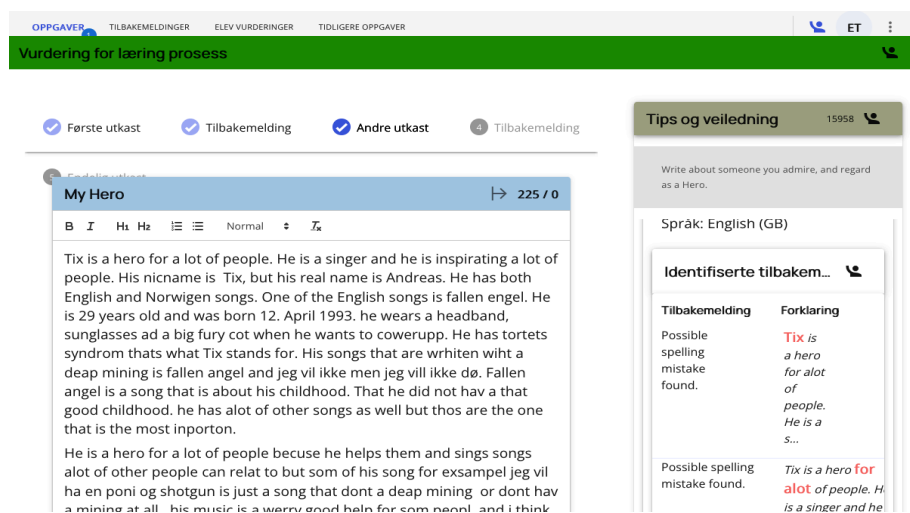


Fig. 1. User interface of the EAT. The interface is currently only available in Norwegian. We will therefore explain the different categories starting at the top. Oppgaver = Tasks. Tilbakemeldinger = Feedback. Elev vurdering = Peer Feedback. Tidligere oppgave = Previous task. To navigate through the tool (on the left), select 'Første utkast' for the initial draft, 'Andre utkast' for subsequent versions, and 'Tilbakemelding' to view feedback. The 'Tips og veiledning' feature (right) on this platform provides users with structured guidance and suggestions to enhance their writing. It is an automated tool designed to identify areas for improvement and offer constructive feedback, helping users to refine their work.

The EAT that was used to develop the first prototype of a technology was leveraged on an open-source language tool called the “Language Tool.” This tool makes it possible for the EAT to analyze a text and highlight errors in grammar, punctuation, and spelling. Open-source language tools are often used for natural language processing (NLP) tasks such as text classification, part-of-speech tagging, and machine translation [33-34].

3.3 Context and Participants

In this study, we scrutinized 56 eighth-grade students' writing trajectories and teacher-student interactions in three classes, each selected from various junior high schools in a municipality in Norway. These schools confirmed their voluntary participation in our 4-year DBR project, which was aimed at improving the students' writing skills in EFL by testing out different versions of the AI-based technology EAT. The eighth grade is the first level at junior high school, and the writing trajectory reported in this study is the very first trajectory of this kind that the students participated in. This means that not only the EAT but also the way of working was new for the students. Recognizing the need to introduce EAT to the teachers, the present study detailed a professional development course to facilitate teachers in exploring the prototype and explaining the nature of the tool to the teachers. The teachers who attended the workshop had time to navigate through the system, gain firsthand experience of its functionality, and discuss how to apply it to the classroom.

3.4 The task

The task was designed in collaboration between the teachers and the researchers (in line with the design principles of DBR) during a workshop weeks before the intervention was initiated in autumn 2022. The students were asked to:

Write a story about someone you consider to be a hero. It can be a fictional or a [sic] real person doing something heroic. The length of your essay should be approximately 400 words. The program will only focus on spelling, grammar, and punctuation. However, you still need to remember to include the title, introduction, main part, and ending.

In addition, a checklist was included as a practical guide for students, covering essential aspects such as the use of capital letters, subject-verb agreement, tenses, spelling, and punctuation. To stimulate the students' ideas, the task was illustrated with pictures of potential heroes such as the civil rights activist Rosa Park and the soccer player Erling Haaland. The task was introduced to the class at the beginning of the writing session and handed out on paper.

3.5 Data

Design experiments often consist of different types of data and is also the case in our experiment. Our data included interviews with students and teachers, video recordings of students' and teachers' interactions during the whole writing trajectory in both the control and target classes, and three submitted versions of the students' essays. In the target classes, these were followed by EAT-generated feedback logs given on the first and second drafts. For this study, we scrutinized a selection of the video recordings and different versions of the students' essays. The project was approved by the Norwegian Agency for Shared Services for Education and Research, and informed

consent was acquired from the participants before participation. All names used in the text are pseudonyms.

3.6 Video Data Selection, Analysis, and Transcription Methods

The intervention was applied in November 2022. We collected video recordings from respectively three target classes and three comparison classes at three different schools. In each class we filmed two groups, and in addition, we had one camera that was used to record the teachers' interactions with their students in the different classrooms. In total, we spent one day at each school where the students had midterm in EFL, and the recordings in the target class and the comparison class took place in parallel.

In this study, we focused on the target classes because our objective was to investigate how this individually tailored automated essay assessment technology could foster students' assessment literacy and writing skills in EFL. Moreover, we achieved this objective by examining the students' writing trajectories through their writing processes in creating several drafts, with a particular emphasis on how they identified and interpreted feedback from the EAT together with their teacher. We recorded 14 hours 8 minutes (848 minutes) of interactions between the selected groups of students and teachers in the target classes. After reviewing these video recordings, we decided to further delimit the data selection, concentrating on the writing trajectories of one group of four students and their teachers' identification and interpretation of the feedback from the EAT. To study students' writing trajectories are space-consuming and the word limitations in journal papers makes it impossible to include more groups when a whole writing trajectory is the scope of the analysis and interaction analysis combined with frequency analysis are used as analytical lenses. When we, in this paper, wanted to develop an understanding of the students' writing trajectories by the use of AI (EAT), their progress throughout a writing process with three drafts and feedback utilization, such a smaller sample size allowed for an in-depth examination of the interactions and learning experiences, ensuring a thorough data analysis. In total, the video data constituted more than 3 hours (198 minutes) of recording.

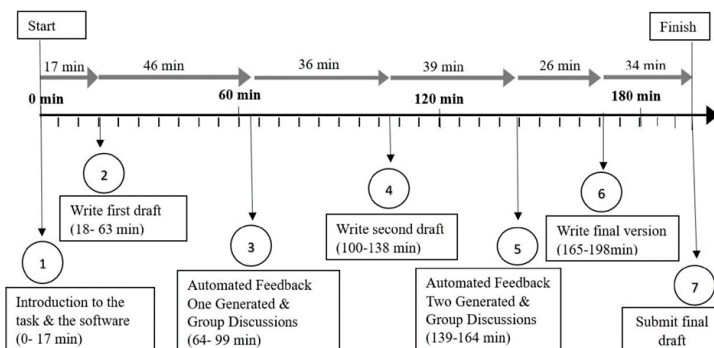


Fig. 2. Timeline of the group's writing trajectory, from beginning to end, divided into different activities and minutes.

We reviewed the video data several times to ensure that all aspects of the students' and their teacher's identification and interpretations of the automatic feedback were transcribed complete and accurate. Gradually, two episodes were selected, one from the first group session, where the students and their teachers identified and interpreted the automatic feedback (see number 3 in Figure 2), and the second session, where they did similar kinds of work (see number 5 in Figure 2). The first episode lasted 36 seconds, and the second episode continued for 66 seconds. Each episode was then transcribed using Mercer's [38] strategy to present these in a nontechnical way and to make the episodes accessible to audiences outside our research community. This means that we wrote down the students' and teacher's talk verbatim, noting gestures and clarifications in brackets, and later translated it into English. As the conversation transpired in an EFL class, the talk was partly in Norwegian and English. To make this shift visible in the transcriptions, we bolded the transcription of the talk when the teacher or students originally talked English to distinguish it from that of the talk in Norwegian. In addition, pauses were marked as number of seconds placed in brackets to indicate the tempo of the talk.

We used interaction analysis as an analytical approach [19], which is well suited for scrutinizing how interactions evolve from moment to moment and how the students and their teacher as interlocutors found ways to identify and interpret the automatic feedback. The analysis was divided into two parts: the analysis itself and its discussion. In the first part, we started by commenting on each utterance to explicitly convey how we understand it and how each utterance builds on the previous utterances. We then looked for different themes introduced in the teachers' and students' interactions. In the second part, we discussed the findings in light of the research background outlined earlier in the text and used it to deduce our research contribution.

The second data source used in this study was the automatic feedback log data from the EAT for each student. These data were collected to analyze how the students followed up on the given feedback and group discussion after the first round of generated automated feedback (see numbers 2 and 3 in Figure 2) and similarly after the second round of automated feedback and group discussion (see numbers 4 and 5 in Figure 2). Finally, we studied the third version of the students' essays (see number 6 in Figure 2) to document the individual students' follow-up of the second feedback loop.

In the analysis, we focused on evaluating the students' responses to the feedback, and we calculated the frequency of the feedback each student received on the first and the second drafts of their essays and how they followed up this feedback correspondingly in the second draft and, finally, in the latest version. The feedback frequencies calculated from the EAT log data were categorized according to the topics identified from the interaction analysis: punctuating with an apostrophe, missing commas and hyphens, and capitalization. When analyzing the second draft and latest version of the essays and how the students used the feedback they received to improve the texts, we divided the use of the feedback into three categories: effective implementation, partially incorporated, or not utilized. Effective implementation entails that the students revised their text into correct versions, while partly incorporated refers to settings where they followed the feedback but, for example, placed a comma after *but* when it should have been before it. This enabled

the identification of trends to investigate areas of improvement or where students underutilized feedback. For this analysis, we followed the same four students as during the interaction analysis and examined their utilization of the feedback they identified and interpreted in collaboration.

4 Analysis

The analysis comprised of two primary segments. Initially, our focus involved documenting the interactions between students and their teachers during the identification and interpretation of feedback on the *first* version of the students' essays. Subsequently, we conducted a frequency analysis to assess how each individual student incorporated the feedback from the EAT in the *second* draft of their essays. Following this, we replicated the same procedure for interactions between the students and their teacher during their identification and interpretation of feedback on the subsequent version on the students' essays. This was accompanied by a following frequency analysis to evaluate how each student addressed the feedback from the EAT in the third and final draft of their essays.

4.1 Students' Identification, Interpretation, and Use of Feedback after Submitting the First Version of Their Essays

Students' Identification and Interpretation of Feedback after Submitting the First Version of Their Essays. The students submitted the first version of their essays and received automatic feedback from the EAT. Four female students sat in a group. They were expected to share the kind of feedback they received, interpret this, and discuss how they could use the feedback to improve their essays. The teacher walked around the classroom to guide the students. Initially, the students appeared to have little trouble getting started. The teacher addressed the group and narrowed in on how the students could approach the task. The students agreed to identify three feedback topics that they would like to share with the group and discuss what these implies. This indicates that the students initiated this collaborative effort and were in the process of sharing their received feedback and engaging in the discussion.

Episode 1: Students talk about the individualized automatic EAT feedback they received on the first draft of their essays.

- 1 Astrid: I will choose *spelling mistakes*. (She points to the EAT feedback on the screen.)
- 2 Mette: Me too.
- 3 Astrid: What is *capitulation*? (She utters *capitalization* wrong and looks at Sarah.)
- 4 Sarah: You are missing a capital letter.
- 5 Astrid: Oh yeah. Okay.
- 6 Kari: I'll choose that too (Referring to the topic that Astrid introduced, emphasizing capitalization feedback.) and *missing comma*.
- 7 Sarah: I am going to choose *missing comma, spelling mistakes, possible*

- 8 Mette: *grammar error*. (She points to the EAT feedback list.)
Okay. I'll choose that *punctuation* and *spelling mistakes*.
9 Astrid: I take *capitalation* (Uttering *capitalization* wrong.), *spelling mistakes* and *missing commas before but*. (She looks at Sarah.)

The teacher encourages Astrid to start, and Astrid says that she will choose “spelling mistakes” while pointing toward the EAT (utterance 1). Mette follows up and says that she will do the same (utterance 2). Astrid then introduces a new feedback topic, “capitalization,” or “capitulation” as she said, asking what it implies (utterance 3). Sarah replies and explains that it means that Astrid is missing a capital letter (utterance 4). Astrid confirms that she understands by saying, “Oh yeah. Okay” (utterance 5). Kari continues by saying that she will choose that too, referring to “capitalization” and adds “missing comma” (utterance 6). Sarah follows up and says that she will choose “missing comma, spelling mistakes, possible grammar error” while pointing at the EAT (utterance 7). Mette signs up and says that she will choose “punctuation and spelling mistakes” while looking at Sarah, who was the last one to speak (utterance 8). Meanwhile, Astrid sums up that she will focus on capitalization, spelling mistakes, and missing commas before *but* (utterance 9). Kari rounds off the talk by taking out a pen and piece of paper, where she notes down what she will emphasize when improving her text (utterance 10).

This short episode contains several aspects. First, the students do not get started before the teacher narrowed in on how the students could approach the task. Further, all students managed to identify the feedback in the EAT and all shared partly overlapping feedback topics and issues for which they had received feedback. In total, six types of feedback appeared in their talks in the following order: spelling mistakes, capitalization, missing comma, possible grammar error, punctuation, and missing comma before *but*. Moreover, it is evident that the students managed to build on each other's utterances by saying, “Me too” or “I'll choose that too,” referring to the previous utterance. Furthermore, the students demonstrated that they contributed to each other's interpretation of the feedback when it was not entirely clear what the feedback entailed. For example, Sarah explained to Astrid that capitalization means that Astrid is missing a capital letter. Astrid confirmed that she understood the explanation, and Kari followed up by saying that she would also highlight this as one of her feedback topics. In addition, Astrid's repetitive use of “capitulation” rather than “capitalization” indicates that she had not yet really made this concept hers, although she understood that it is about capital letters. In this context, the students did not go into any details about what sort of capital letter error Astrid made. Was it a capital letter as the first letter in a sentence or, for example, the capital letter of the nationality Norwegian? The students talked about the feedback in general terms and did not use examples to enrich and specify the discussion. They noted a typographical error but did not investigate it, share knowledge about why it was an error, or refer to any writing rule.

In addition, this group of students used feedback on different specification levels such as punctuation, which is a general grammar rule; missing comma, which refers to all sorts of problems with missing commas as a type of punctuation error; and “missing comma before *but*,” which is a specific comma error. However, these differences were not discussed. This could be due to the fact that the students were

not aware of the difference at this stage in their writing processes, simply read aloud the feedback listed in the EAT, and/or did what they were asked to by sharing three feedback issues each from the EAT without necessarily elaborating them. Some feedback concepts are more absolute, such as spelling mistakes, leaving no doubt that a word is written wrong, whereas other feedback concepts are more open for interpretation, such as a possible grammar error.

Students' Use of Feedback in Preparing the Second Version of Their Essays. We followed up on the feedback and the students' talk to investigate how they addressed the feedback from the EAT in the second draft of their essays. We found several categories of errors made by the students, such as spelling, punctuation, capitalization, and subject-verb agreement errors; repetitions; missing prepositions; incorrect capitalization of proper nouns; incorrect use of adverbs and articles; and typographical errors. To provide a comprehensive analysis, we decided to delimit the focus to two feedback topics that the students identified and partly interpreted during the first episode, namely punctuation focused on apostrophes, missing commas, and hyphens, and capitalization. This choice was made because we wanted to address the most common errors and punctuation and capitalization have significant impacts on sentence structure, clarity, and overall readability. Table 1 presents an overview of the numbers of errors made in the first draft and how the feedback was utilized in improving the essays.

Table 1. Utilization of feedback in the second draft.

	Use of feedback in the second draft of the essay															
	Number of feedback items from the EAT based on the first draft				Effective implement ation in the second draft				Partially incorporate d into the second draft				Not utilized in the second draft			
Participant	A	M	S	K	A	M	S	K	A	M	S	K	A	M	S	K
Punctuation																
Apostrophes	0	2	0	6	0	0	0	0	0	0	0	4	0	2	0	2
Missing commas	4	14	5	8	2	2	2	5	1	0	0	0	1	12	3	3
Hyphens	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Capitalization	4	13	3	4	4	6	3	3	0	1	0	0	0	6	0	1

Note: The participants pseudonyms: A, Astrid; M, Mette; S, Sarah; and K, Kari. The "0" in the first column indicates that the students did not receive this type of feedback from the EAT. The "0" in the second, third, and fourth columns indicates that the students did not follow up on the feedback they received knowing that they had gotten 1 or more feedback in the first column.

We observed that the students made efforts to implement the feedback provided by the EAT. One area of focus was capitalization. Capitalization errors occur when students fail to capitalize proper nouns, sentence beginnings, or specific titles. Acknowledging the significance of proper capitalization, the students demonstrated commendable efforts to correct these errors. A comprehensive analysis of their written work revealed that most students were conscious of the rules governing capitalization and made deliberate attempts to adhere to them. Only Mette received feedback on many more errors than the other three. We noticed that Sarah and Astrid revised all capitalization errors, whereas Mette and Kari addressed only some of them.

While the students demonstrated a willingness to rectify this issue, there were instances in which the suggested correction was not fully incorporated. The EAT suggested inserting a comma before the word *but* in certain instances, as this would enhance the clarity and coherence of their writing. However, instead of placing the comma before *but*, the students partially incorporated the feedback by placing the comma after it. It is possible that the students misunderstood the suggested correction or the reason behind it. In the case of placing the comma, the student might have interpreted the feedback as indicating a need for a comma after *but* rather than before it or simply just misplaced it. This misinterpretation led to the partial incorporation of feedback.

In addition to examining the students' implementation of feedback, our study also shed light on situations where students chose not to utilize the feedback provided. Through our observations, we noticed that when the student wrote “dont,” with a missing apostrophe, the EAT suggested the addition of an apostrophe to the word to make it “*don't*,” but the students attempted to address the error by changing the form to “*do not*,” which is probably even more correct in a formal school setting. Finally, this analysis also revealed that some students removed entire sentences identified as errors in their initial drafts. Consequently, the feedback provided by the EAT was not utilized in their subsequent drafts. This pattern raises questions regarding the students' decision-making process and suggests that the role of teachers in the implementation of this technology in EFL classes might be to encourage students to address as much feedback as possible, among others.

4.2 Students' Identification, Interpretation, and Use of Feedback after Submitting the Second Version of Their Essays

Students' Identification and Interpretation of Feedback after Submitting the Second Version of Their Essays. The students now further developed their essays, submitted their second drafts, and again received automatic feedback from the EAT. They talked about how many words they had written, how much feedback they had received compared with the first draft, and what feedback the others received. One student said that she dreaded seeing how many errors there were, and all four students were worried about whether they would have time to do anything with all the feedback. Again, the teacher encouraged the students to choose three types of feedback to share with the other students in the group, investigate how the feedback was explained in the EAT, and discuss how these could be corrected. However, the students never got started. First, when the teacher bent down next to one student, the

feedback interpretation started. In the following paragraphs, we examine an episode based on feedback that one student received on her essay. The teacher was present and actively participated in the dialogue, and all students paid attention. The teacher and students had just had a similar discussion about formal and informal language based on the feedback that one of the other students received. The teacher moved toward a new student but remained in the same group.

Episode 2: The students' and researcher's talk about the individualized automatic EAT feedback that Mette received on the second draft of her essay, which was on the use of comma when two complete sentences are part of one sentence.

1. Teacher: You might scroll (Pointing to the EAT feedback), and you might choose what to discuss.
2. Mette: Should I take that one? (She points to the EAT feedback.)
3. Teacher: Yes, you can take that one.
4. Mette: Should I read that one?
5. Teacher: Yes.
6. Mette: *She is on the way to the school and she will meet with her friends.*
(She reads aloud from the EAT feedback.)
7. Teacher: Yes, and then it writes that there must be a comma. (She points to the feedback explanations.) Comma, where should it be comma in that sentence? (She turns the laptop so that the other students can see too.) Now we show it to the group again. (2)
8. Sarah: Before *and*?
9. Teacher: Why? (She looks at the other students as well.)
10. Kari: Because it is, when it is two different. (The teacher confirms by nodding.)
11. Sarah: When it is, or...
12. Astrid: So, you going to take...
13. Teacher: Yeah, right. Two independent sentences. For example, *she is on the way to the school, and she is going to meet her friends. She is on the way to school.* Okay. Right. Could we place a full stop here? (She points to the feedback explanations.) *She is going to meet her friends.* Could we place a full stop here too? So, there are two equivalent sentences. There is one that, for example, yes—that is why it should be a comma.
14. Mette: Yes.

The teacher stands next to Mette and encourages her to identify feedback that they can discuss. Mette points to a particular spot at the EAT and asks, “Should I take that one?” (utterance 2), and the teacher confirms her suggestion by saying, “Yes, you can take that one” (utterance 3). Mette then follows up and asks, “Should I read that one?” (utterance 4). Again, the teacher confirms, “Yes” (utterance 5). Mette then reads aloud, “She is on the way to the school and she will meet with her friends” (utterance 6). The teacher follows up by pointing to the feedback explanation and says, “Yes, and then it writes that it must be a comma” (utterance 7). She turns toward the whole group, asking them, “Where should it be comma in that sentence?” when she turns to Mette's laptop so that the other students in the group can see the screen too, stating explicitly, “Now we show it to the group again” (utterance 7). Sarah

answers questioningly, “Before and?” (utterance 8). The teacher follows up and encourages the students to explain why this placement of the comma could be a good idea (utterance 9). Kari starts explaining, “Because it is, when it is two different” (utterance 10), and the teacher nods to confirm that Kari is on the right track. Both Sarah and Astrid demonstrate engagement by saying, “When it is, or” (utterance 11) and “So you are going to take” (utterance 12), respectively, although they do not have time to complete their thoughts before the teacher confirms Mette’s interpretation by saying, “Yes, right. Two independent sentences” (utterance 13). The teacher then gives an example of how to read the different parts of the sentence. First, she reads, “She is on her way to school. Okay. Right. Could we place a dot here?” (utterance 13) while pointing to the feedback in the EAT. She then follows up with the same strategy for the second part of the sentence and concludes, “So, there are two equivalent sentences (...) that is why it should be comma” (utterance 13). Mette confirms that she follows the teacher’s argument by saying, “Yes” (utterance 14).

It is evident that the students managed to identify and share feedback when explicitly asked by the teacher. However, it was difficult and challenging for them to get started. There was some sort of role division between the student who read the sentence that was identified for improvement and the teacher who read out loud what sort of error the student should be looking for, which was incorrect comma placement. The teacher also played an essential role in mobilizing all students to participate by addressing all students when asking where the comma should be, nodding to confirm that the students were on the right track, looking at them alternately, and making eye contact, and by turning Mette’s laptop around, making sure that all students could see the screen. Moreover, when the teacher was sure that she had all students’ attention and had made them come up with suggestions for where and interpretations of why there should be a comma in the sentence, she helped the students become aware of the general grammatical rule that explains that they should divide two sentences with a comma if these are equivalent. She concretized this grammatical rule by using the sentence in Mette’s feedback stream as an example. The teacher guides the students’ learning processes. When using the AI-based technology as an assessment tool in EFL classes, the teacher’s role does not at all become less important. Rather, it becomes evident that the role might be to supervise students’ writing trajectories by using the EAT as a point of departure for their talk with the students. However, it is also undeniable that the students actively participated in both identifying and interpreting relevant topics for discussion although this participation was teacher initiated and guided.

Students’ Use of Feedback after Submitting the Second Draft of Their Essays.

Like the feedback analysis on the first draft, this part of the analysis focused on how students used the second draft feedback from the EAT to improve the final version of their essays. In this feedback round, we observed that Sarah, one group member, did not submit the second version of her essay, preventing us from assessing the extent to which the feedback from this version was incorporated into her final draft. Therefore, the values in her column were marked with an asterisk (*). Table 2 presents an overview of the number of errors Astrid, Mette, and Kari made in their second drafts and how these were utilized in their essays or not.

Table 2. Utilization of feedback on the third version of the draft

Participant	Use of feedback in the third version of the essay															
	Number of feedback items from the EAT based on the second draft				Effective implementation in the third version				Partially incorporated into the third version				Not utilized in the third version			
	A	M	S	K	A	M	S	K	A	M	S	K	A	M	S	K
Punctuation																
Apostrophes	0	0	*	7	0	0	*	0	0	0	*	0	0	0	*	7
Missing commas	12	8	*	13	0	8	*	12	0	0	*	1	12	0	*	0
Hyphens	1	0	*	0	0	0	*	0	0	0	*	0	1	0	*	0
Capitalization																
	4	3	*	4	0	3	*	4	0	0	*	0	4	0	*	0

Note: The participants' pseudonyms: A, Astrid; M, Mette; S, Sarah; and K, Kari. The asterisk means no data available/feedback generated from the EAT because the student did not submit a third version. The "0" in the first column indicates that the students did not receive this type of feedback from the EAT. The "0" in the second, third, and fourth columns indicates that the students did not follow up on the feedback they received, knowing that they had gotten 1 or more feedback items in the first column.

Table 2 demonstrates that Sarah did not submit her second draft and that Astrid did not address any EAT-generated feedback in writing the final version of her essay. Regarding Mette and Kari, we noticed that when we were addressing their errors, they displayed proficiency in implementing corrections pertaining to capital letters, indicating a reasonable likelihood that they had abilities to identify and rectify such errors. However, punctuations, particularly apostrophes, were not adequately addressed by the students, although in several instances, they demonstrated an awareness of these errors. For instance, one student recognized the absence of an apostrophe in the contraction "isn't" and attempted to address it but altered the sentence structure and used "is" instead. This suggests that while students acknowledged the presence of errors, the student did not directly implement the feedback by, for example, adding the missing apostrophe.

A similar pattern was observed with commas. In one case, the EAT suggested adding a comma after "right now," but instead of following this suggestion, the student opted to modify the sentence to "At this moment." This highlights the students' independent decision-making process, indicating that they may have changed their mind and decided to change the structure of their sentence instead. These findings indicate that although the students demonstrated awareness of the errors and made attempts to address them, their decision-making process influenced the utilization of the feedback given by the EAT. It is not always the case that identifying and interpreting something on one hand, and using it on the other hand,

are directly connected in a cause-and-effect manner. We also found that owing to the multiple tasks that had to be completed within a given timeframe, the students might have prioritized addressing some errors over others. They probably perceived them to be more significant or crucial to the overall quality of their work. However, we cannot make a conclusive claim because we did not confirm this prioritization with the students by asking questions.

5 Findings and Discussion

The overall objective of this study was to contribute to the understanding of how an AI-based technology can be used to support students' writing processes in EFL through a formative assessment approach. We investigated students' assessment literacy and writing skills in EFL considering the triple challenge: identifying relevant feedback, interpreting this feedback, and using it to improve the quality of the essay. As part of this, we also examined the teacher's role and how she performed her role throughout the students' writing trajectory. We conducted the investigation by studying interactional data from the writing trajectories of a group of eighth-grade students, where they, together with their teacher, identified and collaboratively interpreted EAT-generated feedback. We employed a combination of EAT feedback given on different versions of each essay and the students' submitted essays to examine how the individual students used the automated feedback to improve the drafts and final versions of their essays.

These findings contribute to the literature regarding students' assessment literacy and writing skills in EFL using AI-based automated feedback and strengthen research work on AfL. Moreover, it also informs our DBR project by enlightening further designs of the EAT as an automated essay assessment technology, and other similar AI-based essay assessment tools, and the designs of the writing trajectory as such. In this article, we demonstrate that automated feedback from an AI-based assessment tool can foster students' assessment literacy and writing skills in EFL. However, we also prove that this fostering needs to be followed up by the teacher contributing to initiate the task, moderate the students' talk throughout the trajectory, and support the students to get deeper into the interpretations of the disciplinary aspects of the feedback.

5.1 Students' Assessment Literacy and the Teacher's Role Performance

In line with the factor emphasized in previous studies as particularly important for classroom work with AfL [1, 20-21], the writing trajectory was designed to involve the students in the assessment process to foster their assessment literacy. We documented that the students participated in identifying feedback, which they shared with their peers in their group. However, the teacher's presence and adeptness in using the formative assessment approach played a crucial role in motivating the students to initiate sharing in both the initial and subsequent group discussions. Except for a single occasion where the concept of capitalization was interpreted during the first group discussion, the students' talk was only about what feedback they wanted to

focus on. The students then omitted potential interpretations of what the feedback meant. This finding is in line with a previous study that reported that changes in assessment practices are challenging [e.g., 5, 13, 24]. However, we argue that the students' assessment literacy was in development in that all students identified feedback to improve their essays and took measures to improve the text [3, 25]. Moreover, we also witnessed that the students' choice of feedback to share and the interpretations of the feedback became richer during the second group discussion. This was likely because the teacher participated and led the discussions between the students and herself and among the students. When the teacher did not give direct guidance or did not participate in the discussions, the students' assessment literacy was too short. Together with the teacher, the students managed to assess the EAT feedback, although they also obviously demonstrated a need for further development of their individual and collaborative assessment literacy. Therefore, it is important to remember that this was the students' first experience participating in this kind of AfL activity while using an EAT. Still, we argue that the students' assessment literacy was enhanced, although they obviously demonstrated a need for further development of this literacy.

In a previous study, Smith et al. [22] identified dimensions for assessment literacy that can further nuance the picture of the students' assessment literacy. First, there is no doubt that the students understood the purpose of the assessment and its connection to the quality of their essays. Furthermore, we argue that although this type of AfL was new for the students, they developed a heightened awareness of the assessment process. At the same time, the students struggled to make improvements in their texts. They understood that it would be wise to follow the suggestions given as part of the automated feedback, but it became apparent that it was challenging for all students to rectify the extensive number of errors. However, the students did not always follow up on suggestions for improvement. Our findings revealed that the students opted to employ an avoidance strategy, deciding to tweak the feedback rather than to follow up on the suggestion for changes and to reformulate the text without this being necessarily successful in terms of text quality. Finally, the analysis of the students' talk and discussions revealed that they could participate in the feedback and critically evaluate and discern strategies for improvement, but only when this talk was initiated and organized by the teacher. On their own, the students appeared somehow passive without an assessment literacy repertoire, but together with the teacher, it was easy for them to identify traces of such literacy. This finding contributes to the understanding of how teachers could actively make use of automated and data-driven feedback loops to assist assessment for learning processes beyond students and teachers' individual interaction with digital technology dashboards and hence support learning in a wider classroom context [29-31].

In summary, the AI-based EAT can be productive in supporting students' development of assessment literacy. However, this kind of literacy is not a quick fix. Although we identified all three dimensions proposed by Smith et al. [20] for assessment literacy, the students still have a long way to go to become self-driven and fully assessment literate, considering that they are eighth graders and that both the type of writing session and EAT were new to them. They are still dependent on their teachers' contribution to organize the activity, identify feedback, and make interpretations of the feedback. The teacher's contribution to the talk, identification

and interpretation, is of vital importance for the student to succeed with their writing trajectory also when AI-based essay assessment tools are used.

5.2 Students' Writing Skills in EFL and the Teacher's Role Performance

Research has provided evidence of the importance of instructional feedback as effective in addressing the quality of student writing [27]. If this feedback is concrete about where to make the improvements, it will foster students to increase the quality of their texts [3]. The EAT offers such concrete instructional syntactical feedback to students, through which the quality of their written outputs increases throughout their writing trajectories. However, our focus was on the triple challenge, which can enrich the understanding of this kind of feedback that is given to improve students' writing skills in EFL. The relationships between identification, interpretation, and use are not trivial. Although the feedback has been identified and made explicit by the students who shared it with their collaborating peers, this does not mean that the mistakes would be interpreted further. If it is interpreted further, this is neither a guarantee for an actual improvement of the written text. This also counts for feedback that the students have identified and believed they do not need to interpret and discuss further, such as commas, because the students tend to correct only some of the errors. It is important that teachers become aware of how students handle various syntactical errors differently. As part of our design-based initiative, we will design and develop systems that take these kinds of variations into account as part of both the EAT and teachers' practice.

After the teacher initiated the feedback, all students managed to identify concrete feedback in the EAT and shared it with the group. However, the students seemed to have limited skills to make further interpretations of this feedback, unless the teacher participated in their talk. The EAT has sources such as links to websites that explain syntactical issues, but the students and their teachers must employ these actively. For teachers facilitating students' writing processes, it becomes crucial to demonstrate the potential tools built into the EAT to illustrate how these sources can be used to improve students' writing skills and the quality of their essays. Another issue was that the students used the feedback to improve their essays to a varying degree. Again, the reason might be that this was a new situation for the students and that all the feedback might be overwhelming for them to deal with. However, two of four students did not revise the second draft of their essays at all, submitting the same essay twice as the second and third versions. Was the design of the writing session too extensive for some eighth graders? Was the amount of feedback too overwhelming? Moreover, we found a difference in the kind of syntactic feedback the students decided to revise. While most students managed to correct the use of capital letters, many struggled sorting out punctuation errors such as how to use an apostrophe correctly or where to place a missing comma.

In summary, the EAT is necessary but is inadequate for ensuring that students interpret the feedback and/or use it to improve the quality of their writing skills in EFL. For this reason, students' assessment literacy is still immature. The teacher's fostering of students' assessment literacy is also decisive when the EAT is used. To a large degree, the students are dependent on their teachers' feedback on how to proceed

with learning, which in our case entailed organizing, identifying, and interpreting feedback to be used to improve each student essay. Thus, it is important to keep in mind that we are talking about eighth graders, who are only halfway through the obligatory educational program in Norway and still about to develop not only their writing skills but also their knowledge in EFL.

6 Concluding Remarks

In line with assessment literature, the EAT was designed and developed to provide automated feedback on areas where students require more guidance [5-6]. In the first prototype used in the first iteration of our design-based research initiative, the students received feedback on syntax errors. However, to make the automated feedback successful in fostering students to improve the quality of their essays, the feedback should be followed by productive talk among teachers and students where the assessment elements are an integral part. In our case, this implies that this talk must address the individually tailored automated feedback that the EAT offers simply to make the feedback relevant to students' further works. Previous research has provided evidence that formative assessment proves to be challenging to implement in real-life classroom settings [8-9, 21]. This is also the case when EAT, as an automated assessment tool, is used.

In line with previous research [3], our study documents that the teacher plays a decisive role in fostering both students' writing skills in EFL and assessment literacy. We claim that the EAT worked as a shared object for the teacher and the students, and as a point of departure for their talk, making the automated feedback individually tailored and collectively relevant. Ranalli [34] reported that teachers had free time while using an AI-based automated EAT. Our data did not indicate a similar tendency. Rather, we argue that the teacher's role was slightly altered and that time was therefore spent differently: First, the EAT contributed to teachers' linking of their talk with their students to the individually tailored automated feedback. The EAT allowed the teachers to become very specific in their feedback, although many students were in the EFL class. Furthermore, the process-oriented writing session that the students participated in for the first time and the use of the EAT, which they had also never used before, was not surprisingly challenging for the students. Therefore, the teacher's presence fostering the organization of the activity and contributing to the interpretational work throughout the students' writing trajectories was decisive for the students to improve and complete their essays. The teacher's presence in the students' work with EAT was crucial. It might be worth considering if the teachers should be encouraged to pay particular attention to those writing mechanics that the students particularly struggled to improve, such as punctuation errors. We argue that the students' and their teachers' use of the EAT as a formative automated assessment tool is promising, but this technology is not a general quick fix for implementing this assessment practice in classrooms. The teacher as facilitator is as always equally important, although a new and effective tool has been introduced. As this study demonstrates, the teacher plays an important role as facilitator for students' development of assessment literacy and students' ability to identify, interpret and make use of the automated feedback, although a new and rich tool has been

introduced. However, we also know from previous research that teachers' real-life practices, strategies, and abilities to make sense of automated feedback data may vary greatly [29]. The findings from our study therefore underline that automated AI technology such as EAT should not be perceived to replace the need for teachers to engage in feedback and assessment during a writing process. Rather, such technology supports new ways for teachers to engage in feedback practices as a facilitator for and participant in- shared assessment for learning rather than being the sole assessor.

Through an examination of the students' writing process, it is evident that all students recognized the feedback generated by the EAT, but the interpretation and utilization of the feedback were not at all necessarily easy. The limitations of the students' current assessment literacy were apparent, and we argue that it is still in its infancy. When the teacher's involvement was absent, the students' assessment literacy was poor. At best, the students managed to follow up the teacher-initiated procedure of, for example, identifying feedback to share, without further interpretations of these errors. Regarding the students' use of individualized automated feedback to improve their essays throughout their writing trajectories, our findings indicate that the students' work and effort were somehow differently weighted throughout their trajectories. While all students improved the first draft of their essays, only half did the same for the second draft. In our DBR project, we designed and developed the EAT, which is equally important as the design of the students' writing trajectories (Figure 2), when and for how long the students are going to write, when they would interpret and discuss feedback, and when they are going to improve their essays. Therefore, it will also be important to carefully consider the timeline of the writing session. We must rethink how to encourage students to continue improving their essays all the way to the end of the writing session. We might stimulate this by altering the timeline and adding more time at the end to give the students time to finish the final version of their essays. As part of this, we must rethink how teachers might be encouraged to inspire students to use all opportunities of improving their essays. Figure 2, which presents the chronological development of the group's writing trajectory, from beginning to the end, categorized according to tasks and durations, clearly shows that as the writing session progressed, the students allocated progressively less time for both the act of writing and the ensuing discussions. This pattern indicates a gradual reduction in the available time for these crucial activities as the writing session unfolds. Given this observed trend, it is advisable to consider strategies that maintain a balanced distribution of time for writing and discussion throughout the session. This approach will likely contribute to a more productive and comprehensive collaborative writing process.

Finally, it is noteworthy that prevailing AI-driven automated feedback technologies, such as Criterion (39, 40) and Pigai (41, 42), predominantly revolve around personalized learning and are tailored to furnish feedback to university or college students, irrespective of teacher involvement. Nonetheless, our research findings assert that the efficiency of such automated feedback mechanisms is notably inadequate at a junior high school level, where the proficiency of students in identifying and interpreting feedback is markedly contingent on teacher guidance. The EAT technology transcends the mere application of AI-driven automated feedback for grammatical accuracy; it equally underscores the importance of providing students with feedback conducive to enhancing their comprehension of syntactical rules.

However, in alignment with the argument presented by Hopfenbeck et al [11], the research community is currently at an early stage of comprehending the ramifications linked to the incorporation of AI technologies to facilitate feedback support. Consequently, there exists a pressing need for further insights into the interactional dynamics involving AI-driven automated feedback, students' responses to such feedback, and the consequential influence on teachers in formulation differentiated instructional strategies.

Acknowledgments. For their comments and helpful feedback at various stages of manuscript preparation, we thank Professor Halla Holmarsdottir, Oslo Metropolitan University; Associate Professor Sima Caspari-Sadeghi, Østfold University College; and our research group RIDE (Research in Digital Education). For their participation in this study, we thank the teachers and students. For the design and development of the essay assessment tool, we thank our interdisciplinary research team from Østfold University College, Volda University College, Hypatia Learning, and Halden Municipality. For providing grant No. 326607, we thank the Norwegian Research Council.

CRedit author statement. **Ingeborg Krange:** Data collection, Conceptualization, Methodology, Analyses, Writing- Original draft presentation. **Meerita Segeran:** Data collection, Writing- Methodology, analyses, and development of figures. **Siv Gamlem:** Data collection, Writing- Review. **Synnøve Moltudal:** Writing- Methodology, analyses, and presentation of overall argument. **Irina Engeness:** Data collection, Writing- Review.

References

1. Black, P. Wiliam, D.: Classroom assessment and pedagogy, *Assessment in Education: Principles, Policy & Practice*, 25(6), pp. 551–575 (2018). <https://doi.org/10.1080/0969594X.2018.1441807>
2. Black, P., Wiliam, D.: Assessment and classroom learning. *Assessment in Education*, 5, pp. 7–74 (1998). <https://doi.org/10.1080/0969595980050102>
3. Hattie, J., Timperley, H.: The power of feedback, *Review of Educational Research*, 77(1), pp. 81–112 (2007). <https://doi.org/10.3102/003465430298487>
4. Hattie, J. A.: *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. London and New York: Routledge (2009). <https://doi.org/10.4324/9780203887332>
5. Gamlem, S.M., Smith, K.: Student perceptions of classroom feedback, *Assessment in Education: Principles, Policy & Practice*, 20(2), pp. 150–169 (2013). <https://doi.org/10.1080/0969594X.2012.749212>
6. Schildkamp, K., van der Kleij, F.M., Heitink, M.C., Kippers, W.B., Veldkamp, B.P.: Formative assessment: a systematic review of critical teacher prerequisites for classroom practice, *International Journal of Educational Research*, 103, p. 101602 (2020). <https://doi.org/https://doi.org/10.1016/j.ijer.2020.101602>
7. Gamlem, S. M., & Vattøy, K.-D.: “Feedback and classroom practice” in *International encyclopedia of education*. eds. R. J. Tierney, F. Rizvi, and K. Ercikan, vol. 13. 4th ed (Elsevier), 89–95 (2023). <https://doi.org/10.1016/j.tate.2019.06.024>

8. Black, P., Wiliam, D.: Developing the theory of formative assessment, *Educational Assessment, Evaluation and Accountability*, 21(1), pp. 5–31 (2009). <https://doi.org/10.1007/s11092-008-9068-5>
9. Gamlem, S.M.: Feedback to support learning: changes in teachers' practice and beliefs, *Teacher Development*, 19(4), pp. 461–482 (2015). <https://doi.org/10.1080/13664530.2015.1060254>
10. Wilson, J., Olinghouse, N.G., Andrada, G.N.: Does automated feedback improve writing quality?, *Learning Disabilities: A Contemporary Journal*, 12, pp. 93–118 (2014)
11. Hopfenbeck, T. N., Zhang, Z., Sun, S. Z., Robertson, P., & McGrane, J. A.: Challenges and opportunities for classroom-based formative assessment and AI: a perspective article [Perspective]. *Frontiers in Education*, 8. (2023). <https://doi.org/10.3389/educ.2023.1270700>
12. Hegelheimer, V., Dursun, A., Li, Z.: Automated writing evaluation in language teaching: theory, development, and application, *Computer Assisted Language Instruction Consortium*, 33(1), pp. i–v (2016). <https://doi.org/10.1558/cj.v33i1.29251>
13. Hannigan, C., Alonzo, D., Oo, C.Z.: Student assessment literacy: indicators and domains from the literature, *Assessment in Education: Principles, Policy & Practice*, 29(4), pp. 482–504 (2022). <https://doi.org/10.1080/0969594X.2022.2121911>
14. Engeness, I.: What teachers do: facilitating the writing process with feedback from EssayCritic and collaborating peers, *Technology, Pedagogy and Education*, 27(3), pp. 297–311 (2018). <https://doi.org/10.1080/1475939X.2017.1421259>
15. Engeness, I., Mørch, A.: Developing writing skills in English using content-specific computer-generated feedback with EssayCritic, *Nordic Journal of Digital Literacy*, 10(02), pp. 118–135 (2016). <https://doi.org/10.18261/issn.1891-943x-2016-02->
16. Brown, A.L.: Design experiments: theoretical and methodological challenges in creating complex interventions in classroom settings, *Journal of the Learning Sciences*, 2, pp. 141–178 (1992). [brown-1992.pdf \(uio.no\)](https://doi.org/10.1207/s15327809jls0401_2)
17. The Design-Based Research Collective: Design-based research: an emerging paradigm for educational inquiry, *Educational Researcher*, 32(1), pp. 5–8 (2003). <https://doi.org/10.3102/0013189X032001005>
18. Ludvigsen, S., Rasmussen, I., Kränge, I., Moen, A., Middleton, D.: Intersecting trajectories of participation; temporality and learning, in *Learning Across Sites; New Tools, Infrastructures and Practices*, pp. 105–121, Pergamon, (2010). <https://doi.org/10.4324/9780203847817-13>
19. Jordan, B., Henderson, A.: Interaction analysis: foundations and practice. *Journal of the Learning Sciences*, 4, pp. 39–103 (1995). https://doi.org/10.1207/s15327809jls0401_2
20. Black, P., Wiliam, D.: *Assessment and Learning* (2 edition), SAGE Publications Ltd., (2012). <https://doi.org/10.4135/9781446250808>
21. Vattøy, K.-D., Gamlem, S.M., Kopperstad, L.R., Rogne, W.M.: Students' experiences of assessment and feedback engagement in digital contexts: a mixed-methods case study in upper secondary school, *Education Inquiry*, pp. 1-22 (2022). <https://doi.org/10.1080/20004508.2022.2122202>
22. Smith, C.D., Worsfold, K., Davies, L., Fisher, R., McPhail, R. Assessment literacy and student learning: the case for explicitly developing students 'assessment literacy', *Assessment and Evaluation in Higher Education*, 38(1), pp. 44–60 (2013). <https://doi.org/10.1080/02602938.2011.598636>
23. Price, M., Rust, C., O'Donovan, B., Handley, K., Bryant, R. *Assessment literacy: the foundation for improving student learning*, ASKe, Oxford Centre for Staff and Learning Development. (2012).

24. Torkildsen, L.G., Erickson, G.: 'If they'd written more...'—On students' perceptions of assessment and assessment practices, *Education Inquiry*, 7(2), p. 27416 (2016). <https://doi.org/10.3402/edui.v7.27416>
25. Gamlem, S.M., Munthe, E.: Mapping the quality of feedback to support students' learning in lower secondary classrooms, *Cambridge Journal of Education*, 44(1), pp. 75–92 (2014). <https://doi.org/10.1080/0305764X.2013.855171>
26. Graham, S.: Changing how writing is taught. *Review of Educational Research*, 43, pp. 277–303 (2019). <https://doi.org/10.3102/0091732X18821125>
27. Graham, S.: Instructional feedback in writing, in *The Cambridge Handbook of Instructional Feedback*, pp. 145–168, Cambridge University Press, (2018). <https://doi.org/10.1017/9781316832134.009>
28. Graham, S., McKeown, D., Kiuahara, S., Harris, K.: A meta-analysis of writing instruction for students in the elementary grades, *Journal of Educational Psychology*, 104(4), pp. 879–896 (2012). <https://doi.org/10.1037/a0029185>
29. Moltudal, S., Høydal, K., & Krumsvik, R. J. Glimpses Into Real-Life Introduction of Adaptive Learning Technology: A Mixed Methods Research Approach to Personalised Pupil Learning, *Designs for Learning*, 12(1), 13–28. <https://doi.org/10.16993/dfl.138>
30. Moltudal, S. H., Krumsvik, R. J., & Høydal, K. L. Adaptive Learning Technology in Primary Education: Implications for Professional Teacher Knowledge and Classroom Management. *Frontiers in Education*. <https://doi.org/10.3389/educ.2022.830536>
31. Munthe, E., Erstad, O., Njå, M.B., Forsström, S., Gilje, Ø., Amdam, S., Moltudal, S., Hagen, S.B. (2022). *Digitalisering i grunnopplæring: kunnskap, trender og framtidig forskningsbehov*. Kunnskapssenter for utdanning: Universitetet i Stavanger. https://www.uis.no/sites/default/files/2022-12/13767200%20Rapport%20GrunDig_0.pdf
32. Danielsen, A.G., Samdal, O., Hetland, J., Wold, B.: School-related social support and students' perceived life satisfaction. *The Journal of Educational Research*, 102(4), pp. 303–318 (2009). <https://doi.org/10.3200/JOER.102.4.303-320>
33. Vattøy, K.-D., Gamlem, S. Teacher-student interactions and feedback in English as a foreign language classroom. *Cambridge Journal of Education*, 50(3), pp. 371–389 (2020). <https://doi.org/10.1080/0305764X.2019.1707512>
34. Ranalli, J., Link, S., Chukharev-Hudilainen, E.: Automated writing evaluation for formative assessment of second language writing: investigating the accuracy and usefulness of feedback as part of argument-based validation, *Educational Psychology (Dorchester-on-Thames)*, 37(1), pp. 8–25 (2017). <https://doi.org/10.1080/01443410.2015.1136407>
35. Collins, A., Joseph, D., Bielaczyc, K.: Design research: theoretical and methodological issues, *Journal of the Learning Sciences*, 13, pp. 15–42 (2004). https://doi.org/10.1207/s15327809jls1301_2
36. Dasgupta, J.: Imparting hands-on industry 4.0 education at low cost using open-source tools and python eco-system, in *New Paradigm of Industry 4.0: Internet of Things, Big Data & Cyber Physical Systems*, pp. 37–47, Springer, (2019)
37. Qi, P., Zhang, Y., Zhang, Y., Bolton, J., Manning, C.D.: Stanza: a Python natural language processing toolkit for many human languages, arXiv preprint arXiv:2003.07082 (2020). <https://doi.org/10.48550/arXiv.2003.07082>
38. Mercer, N.: The guided construction of knowledge. *Talk Amongst Teachers and Learners*, Multilingual Matters Ltd., (1995)
39. Han, T., & Sari, E. An investigation on the use of automated feedback in Turkish EFL students' writing classes. *Computer assisted language learning, ahead-of-print(ahead-of-print)*, 1-24. (2022). <https://doi.org/10.1080/09588221.2022.2067179>

40. Link, S., Mehrzad, M., & Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer assisted language learning*, 35(4), 605-634. <https://doi.org/10.1080/09588221.2020.1743323>
41. Gao, J. Exploring the Feedback Quality of an Automated Writing Evaluation System Pigai. *International journal of emerging technologies in learning*. (2021). <https://doi.org/10.3991/ijet.v16i11.19657>
42. Geng, J., & Razali, A. B. (2020). Tapping the Potential of Pigai Automated Writing Evaluation (AWE) Program to Give Feedback on EFL Writing. *Universal Journal of Educational Research*, 8, 8334-8343. <https://doi.org/10.13189/ujer.2020.082638>