Functional, frustrating and full of potential: learners’ experiences of a prototype for automated essay feedback

Bethany Alden Rivers1*, Denise Whitelock1, John T. E. Richardson1, Debora Field2, Stephen Pulman2

1Institute of Educational Technology, The Open University, Milton Keynes, UK
{b.alden, denise.whitelock, john.t.e.richardson}@open.ac.uk
2Computer Science Department, The University of Oxford, Oxford, UK
{debora.field, stephen.pulman}@cs.ox.ac.uk

Abstract. OpenEssayist is an automated feedback system designed to support university students as they write essays for assessment. A first generation prototype of this system was tested on a cohort of postgraduate distance learners at the UK Open University from September to December 2013. A case study approach was used to examine three participants’ experiences of the prototype. Findings from the case studies offered insight into how different users may perceive the usefulness, future potential and end-user of such a tool. This study has important implications for the next phase of development, when the role of OpenEssayist in supporting students’ learning will need to be more clearly understood.

Keywords: automated feedback • essay-writing • software evaluation • case study

1 Introduction

OpenEssayist is an automated feedback system designed to provide instantaneous support to university students as they draft essays for summative assessment. There are two components to the system: (1) the learning analytics engine (EssayAnalyser) and (2) the web application that provides feedback for students (OpenEssayist) [1, 2]. The rationale for developing an automated feedback tool, such as this one, rests largely on the knowledge that university students find essay-writing to be a challenging task [3]. A system that provides immediate feedback, or “advice for action” on students’ draft essays could be one way to support learners in the essay-writing process [4].

In September 2013, a first generation prototype of OpenEssayist was made available to students for testing. At this stage of software development, it was important for the research team to understand the answers to three questions: (1) How useful is OpenEssayist for helping students prepare drafts of their essays? (2) How could the prototype add more value to the essay-writing process? and (3) What type of student
would benefit from using an automated feedback tool for essay-writing? This paper reports on empirical research that was carried out to address these questions.

2 Computer based summative assessment

The bulk of work in the automated marking of free text has been concerned with essays. One of the earliest marking systems which was put into commercial use is E-rater [5]. E-rater uses various vector-space measures of semantic similarity to determine whether an essay contains the appropriate conceptual content. It also carries out some shallow grammatical processing, and looks for simple rhetorical features (e.g., a paragraph containing a phrase like ‘in conclusion’ ought to go at the end of the essay).

Other commercial essay marking systems include IntelliMetric and Pearson’s KAT engine, based on Landauer’s Intelligent Essay Assessor [6, 7]. Both of these systems use a vector-space technique for measuring semantic similarity to a gold standard essay, known as Latent Semantic Analysis. For the most part, these systems focus on assessment alone, rather than feedback. Some of the systems can be used to elicit generic feedback on a final version of a draft essay. However, this type of feedback is not tailored to the essay.

Thus while automated assessment of free text can be thought of as reasonably well understood (although of course current systems are relatively crude compared to a human marker) the process of constructing individualized feedback automatically is much less well established.

3 Evaluating prototypes for automated feedback

Thus far in the development of OpenEssayist, focus groups had been carried out with students to understand how they go about writing essays [8]. This research informed a model of students’ essay-writing processes that was used to develop the prototype (see Figure 1).
Additionally, some usability tests, employing a think-aloud protocol, were conducted with a few members of academic staff, and insights from accessibility experts were used to inform the early stages of the design process. However, up to the point of this study, students had not been involved in testing the prototype.

There is an abundance of literature dealing with software evaluation, usability trials and accessibility testing of prototype educational technology. Alden et al. [8] identified two research reports that were very relevant to the empirical evaluation of OpenEssayist. The first was a study by Chandrasegaran, Ellis and Poedjosoedarmo (2005), who carried out user-testing with 29 undergraduate students to evaluate their software called EssayAssist, a computer program that helps students make decisions during the essay writing process. Their study asked students to identify which features of the software were helpful, missing or problematic. Although their paper reported on a relatively early stage of development, the authors were hopeful that, eventually, this software would provide students with a much-needed level of essay-writing support. According to the authors, what set EssayAssist apart from other, similar writing
tools was the “in-process guidance” that helped students when they encountered a problem with their writing [9].

A second report by Roscoe et al. (in press) discussed various methods used to test the first version of the ‘The Writing Pal (W-Pal)’. W-Pal is an intelligent tutoring system that offers automated formative feedback to students as they prepare essays. Their system uses game based instruction and focuses on the development of writing strategies. This team of authors reported that W-Pal was unique to other ‘automated writing evaluation systems’ (AWEs) because it had been designed with a pedagogical focus. Other AWEs, according to Roscoe et al. had been designed to rate the quality of essays. Their team evaluated W-Pal using several phases and methods of testing. The first version of their prototype was deployed with different groups of users, each asked to complete usability and perception surveys after using the system [10].

As in these studies, the first generation prototype of OpenEssayist was deployed on a group of student-users. Insights from this testing will be useful in terms of developing the usability and accessibility of the system. However, it is also important for the next phase of development to consider whether students perceive such a system to add value to their learning experience and who, exactly, they believe would use such a system.

Fig. 2. Draft Overview [1, 2]
4 The OpenEssayist prototype

The core functionality of the OpenEssayist system can be grouped in two distinct parts: task and draft. The task side relates to the management of the system’s activities. These include tasks such as logging in and out, accessing specific essays, submitting new drafts and keeping a record of submissions. The draft side of the system relates to the activities around a specific draft that has been submitted for analysis. Outputs from these activities are provided to the user as external representations of different analyses—or “views”. Table 1 is a list of these views that were available to users of the first generation prototype of OpenEssayist and Figure 2 is an example of the ‘Draft overview’ view. See Van Labeke et al. (2013a, 2013b) for a more detailed account of the system’s functionality [1, 2].

Table 1. External representations of essay analyses in the OpenEssayist prototype

<table>
<thead>
<tr>
<th>View</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Draft overview</td>
<td>A page showing a structured version of the draft essay in which key words, phrases, and sentences can be highlighted</td>
</tr>
<tr>
<td>Key words and key phrases</td>
<td>A page showing the frequency distribution of the most used words and phrases in the draft essay, including their frequencies</td>
</tr>
<tr>
<td>Key sentences</td>
<td>A page showing the most important sentences in the draft essay</td>
</tr>
<tr>
<td>Key word dispersion view</td>
<td>A page showing how the key words and phrases are distributed across the entire draft essay</td>
</tr>
<tr>
<td>Word cloud view</td>
<td>A page showing a picture of a cluster of the key words and phrases presented in different colours and sizes to visualize their frequency distribution.</td>
</tr>
<tr>
<td>Word limit view</td>
<td>A bullet graph showing the number of words within each section of the draft essay and comparing the total words with the word limit of the assignment</td>
</tr>
<tr>
<td>Word count view</td>
<td>A pie chart showing the number of words within each section</td>
</tr>
<tr>
<td>Organize the key words view</td>
<td>A page that allows the user to group key words and phrases and to then see the groups highlighted in different colours in the draft overview view</td>
</tr>
</tbody>
</table>
5  A case study approach

A case study method allowed the team to explore an “exemplifying case”, or one that “provides a suitable context for certain research questions to be answered” [11]. To address our research questions of usefulness, future potential and perceived end-user, the research team employed a case study approach to gain a closer insight by way of three users’ ‘stories’ of OpenEssayist.

Students on a postgraduate module at the UK Open University were invited to engage with the OpenEssayist prototype between September 2013 and February 2014. Of this sample, two students agreed to participate in a more detailed discussion of their experiences of using OpenEssayist and one student in the cohort, who decided not to use the system, also agreed to participate in this case study.

Permission to carry out this research was approved by the University’s Student Research Project Panel and by the University’s Human Research Ethics Committee. Participants provided informed consent before engaging in recorded telephone interviews, which varied in length from 12 to 35 minutes. Transcriptions of the conversations were analyzed to investigate key themes and to summarize each participant’s own narrative.

6  The case of OpenEssayist

6.1  Maria’s story: “It encourages you to think but it’s too bewildering for a novice learner.”

At the time of sharing her experiences of OpenEssayist, Maria was in her late-fifties, working in part-time employment. According to university records, Maria was a White woman, held a postgraduate qualification and had a disability. She was reportedly enrolled in the present module for career and personal development reasons. During the interview it emerged that Maria had a very strong background in language and linguistics.

When asked to explain her approach to essay-writing, Maria was able to talk about how she took notes, created a plan (using ‘something between a mind map and a list’) and constructed a structure using headings. The planning, according to Maria, was the most important stage of essay-writing. She explained that an absence of a structure made her feel ‘less confident’ and that it was ‘a bit scary’ to forge ahead without clear direction.

Maria said that she prepares drafts of her essays, building on her own reflections after each draft. She has clear strategies for using feedback to improve her performance on subsequent essays, and was able to share a specific example of when she used the
tutor’s suggestion to demonstrate a deeper understanding of certain concepts within her writing.

There was a sense of excitement as Maria talked about the ways that OpenEssayist could assist students in these early stages of essay-writing. For example, Maria thought that the Key Sentences view could help learners during the planning process because it shows the most important sentences. She explained that if these were stripped out, then it should show the essay structure. When asked to talk about how OpenEssayist would help a student during the drafting stages, she commented that ‘it can throw up things that you haven’t been aware of.’ In particular, Maria thought there was potential for the system to help students with their writing style, word choice and essay structure.

Maria thought that the Key Word Dispersion view was a useful resource because, as she explained, students struggle to find different words or phrases to mean the same thing. She suggested that the functionality of OpenEssayist would add more value if it also worked as a thesaurus. Maria thought the system could highlight *variety* as a positive feature in one’s writing, by offering suggestions for alternative words and phrases.

When asked to consider the type of learner who might benefit from OpenEssayist, Maria was clear that it should not be aimed at a novice student. She explained that the present version is too ‘bewildering’ and that it would, therefore, be an inappropriate tool for a beginner or for someone who was ‘not so familiar with ICT’. (Later in the interview, however, Maria contradicted this notion by suggesting the tool be tested with students on a first year essay-writing module.)

Her earlier excitement endured throughout the interview and flowed into a few ideas of her own for the future of the system. She toyed with the idea that OpenEssayist could be a catalyst for peer support, explaining that ‘you are a bit more isolated as an Open University student’. Maria thought it would be a good idea for the system to help students help each other but, when pushed to explain, it was difficult for her to envision how this might work.

Furthermore, Maria saw the possibility for OpenEssayist to help students with other types of assignments, not just essays. She wondered, too, whether the system could help foreign language students with their assessed work.

6.2 Robert’s story: “It could be useful but mainly for students who are less confident.”

Robert had declined the invitation to participate in the earlier usability test and therefore, had not provided his demographic data. When asked in the interview to describe his professional and educational background, he explained that he had a background in criminal psychology, anthropology and religious studies. He said that the present
module was contributing towards his second Master’s degree and that he was also working towards a PhD.

Robert could not understand what OpenEssayist was supposed to do for him and he was unsure whether to use it for his own assignments. He already felt confident with his essay-writing skills. Later in the interview, he admitted that there are still a few areas of essay-writing where he was not so confident (word choice and structure). And, on ending the interview, he explained that, actually, he would have liked to have used OpenEssayist on his second assignment but could not find the link to the software.

Robert explained that he has different approaches to planning and writing depending on the length of the essay. For short essays, he just starts writing and applies a structure later. For longer essays, he creates a structure first and then ‘populates’. His activities prior to planning and writing include reading the task and taking notes based on what is required. He said he writes ‘many, many drafts’.

Robert does not use other people as sources of support, unless it is for proofreading. His described his approach for using feedback from the tutor to enhance his performance as threefold: (1) analyzing feedback by going through the comments, (2) understanding what they are saying and where they are coming from, and (3) trying to adapt what he has written to compensate for that. A lot of the feedback that Robert receives deals with organization or phrasing of ideas. He was able to describe an example of when he used feedback to improve his performance on a previous assignment, by including more theoretical discussion.

Although he didn’t use the prototype, he could see how such a system ‘would be very useful.’ He thought that students with learning disabilities would benefit from using such a system, such as students with dyslexia, as well as students who are not as confident writing essays. As a learning tool, he perceived OpenEssayist as one that could help a user focus on ‘what bits might be important’, like ‘structure or synthesis’.

Robert confessed that he still ‘struggles’ with word choice, not always knowing whether he is choosing words that are too colloquial or too academic. He also ‘struggles’ with structure—‘knowing how much to talk about, how much needs actioning’. While it did not seem relevant for Robert to use the prototype to support his own essay-writing, through conversation with him, it seemed that he was rethinking whether such a tool would be useful to him in the future.

6.3 Karina’s story: “Worrisome, confusing and fascinating: this system is for the younger generation, not for mature learners.”

Karina took early retirement after a ‘career in technology’. She had earned a Bachelor of Education and a Bachelor of Science when she was younger, but during retirement decided to enroll on the Master in Online and Distance Education. Even though she is
a ‘technologist at heart’, she was always interested in teaching and mentoring throughout her career. According to university records, Karina was a White woman in her early-fifties. She already held a higher education qualification and was enrolled in the present module for personal development reasons.

Karina admitted that she had really struggled with essay-writing during her higher education experience. The word count was always perceived as a constraint because she naturally had ‘way too much to say’. In her professional life, this had not been a problem because she just wrote to whatever length she needed to put her point across, and then attached an executive summary. ‘I always found that writing the essay was the hardest thing.’

Karina’s approach to essay-writing never felt methodical enough to her, and this used to frustrate her. She continued to feel that she ‘ought to be better organized in gathering the scenes and the supporting evidence, but it never seemed to work out that way.’ She explained that her pre-writing tasks include: understanding the theory, understanding what points she needed to make, understanding what academic material she needed to include to answer the question, structuring her thoughts around what approach she was going to take to answer the question, and trying to give some structure to the essay-writing process (the last which, she described as a ‘challenge’). Her process used to be to write a very long draft, usually spending a lot of time on the first two sections and then less time on the final section. She usually realized that she could have rewritten her first draft again.

Despite feeling that essay-writing was a challenging activity and that she lacked a rigorous method, Karina admitted that she feels fairly confident about writing essays despite experiencing ‘the real bad patch about two and a half weeks in’. She starts preparing and writing early. Her career has helped her become accustomed to working to deadlines and just ‘getting it done’.

Karina shared that she perceives online learners to be disadvantaged because they do not have opportunities to talk to their peers about how they are going to approach their essays. As an online learner, she found it strange how little dialogue there was around essay-writing. She said she was never told not to share this sort of information but rather she believes she was not allowed to do so because of the discourse around plagiarism ‘that features so heavily in everything’. Karina suggested that one way to offer peer support around essay-writing could be to talk about essay structures and the use of word count. This would not give away the actual essay, she explained, rather it would offer students the chance to talk about how the essay could be approached.

Karina’s first thoughts about OpenEssayist were of fascination and intrigue. She said that, as a technologist, she was interested in a tool that could deal with a variety of written essays. She felt surprised to learn that it was designed to deal specifically with each assignment, rather than as a general tool. She was not sure whether this perception was correct or not. After using the tool, she could see the potential of the system
based on what ‘was being got at’ but she was not sure that she was able to use it in the right way. She questioned whether her use of headings and formatting meant that the system could not recognize new sections or that it could not recognize that her essay was covering a particular point.

Reflecting on her experiences with OpenEssayist, she struggled to understand why, in the places where she was making 10 or 15 points, the system only picked up on her making two or three points. She again questioned whether this was due to the structure she imposed through stylized formatting. This mismatch between what she had understood her essay to achieve and what the system said her essay was achieving caused her to feel worried.

I think I thought I wasn’t answering the question (laughter), so so you know it slightly threw me off track and then I made my way back so it was a bit of a prompt to make sure that I went back and said ‘well am I making some faults here?’, umm as it was a bit of a checkpoint for me, but I ended up worrying that the tool was right and I was wrong.

Karina saw that the system could help students reflect on their essay by encouraging them to think about the essay in terms of the system’s output.

the messages you were trying to put through, you know the weight of the argument, the percentage of the introduction, conclusion, so yes, yes I do because those are important features as well as what you write in, how you write it and how you structure it.

Karina said she was used to incorporating feedback into her further work; doing so was important in her career (in building and designing systems). So, she felt okay about incorporating feedback from her tutor. ‘It is sort of part and parcel of the way I work, really.’

Karina suggested that a built-in narrative or preface to using OpenEssayist, would have helped her understand it better.

I think I would have liked, umm and I think this is true of any technology, is someone to talk me through it, so although it was very easy to use, it wasn’t easy to understand, if that makes sense?

She thought that this narrative would be best delivered as a tutorial prior to using the tool. Karina reflected on a previous experience of having an hour-long tutorial from a tutor for using an accessibility tool. She remembered that the tutorial ‘sort of transformed understanding’.

When asked to comment on what type of learner would be the target audience for OpenEssayist, Karina believed that a traditional-aged student would be best equipped to benefit from this tool. She stated that ‘younger graduates are quite a lot much more able [sic] to deal with these tools and so on.’

Like Maria, Karina believed that the functionality of OpenEssayist could be enhanced if it advised students on word choice, as a thesaurus might: ‘so here are the alternative
words that you could use’. Karina also thought that the system could encourage students’ reflection by offering examples, such as an exemplar introduction or conclusion. She felt that she struggles (sometimes) because she has only ever seen her own essays. She suggested that these examples could be specific to the module or general. One idea of Karina’s was to have examples of excellent essays that students can input into OpenEssayist and examine the results. The exercise then becomes understanding what output looks like from a good piece of writing.

7 Findings

7.1 Usefulness

Of the three stories, Maria’s was the most positive in terms of what the system could already do. She could see the potential of various views to prompt a user to think about their essay structure and the variety of words they had used in their writing. On the other hand, Karina’s experience of using OpenEssayist was frustrating and confusing. As a technologist, Karina brought certain expectations of what the system was going to do. When these hopes were not met, this caused disappointment. When the output from the system caused her to doubt its ability to pick up on her key points, she was left feeling puzzled at the root cause, rather than enabled by the system. Despite these worries, Karina was still able to see the possibilities of the system for supporting developing writers. Indeed, it was the participants’ ability to look past the system’s current functionality that illuminated its potential.

7.2 Potential for adding value

All of the participants were able to talk about the potential of OpenEssayist to add value to their essay-writing experience. There were two main themes—structure and word choice—that emerged from each of these narratives. Maria, in particular, was interested in how the system could enhance the essay-planning process. Both Robert and Karina saw this possibility too. All of them, in some way, mentioned that word choice was an issue for students as they write essays. Not only finding the right words to use but also finding a variety of words to use, is an area where they believed students struggle. A built-in thesaurus and mechanism for suggesting alternative words seemed to be a priority that they perceived OpenEssayist could address, with further development of the system.

The notion of peer support emerged as a theme, with Maria and Karina both wondering how or if students could share their ideas for essay-writing. Clear ideas for how OpenEssayist may be able to support this process were not explicated. However, they all saw the potential for peer support to benefit students, particularly distance learners. Karina’s suggestion that the system have a built-in narrative to support learners in understanding the various functions and outputs is one way to enhance the value that OpenEssayist may offer students.
7.3 Target user

Interestingly the findings of the case study indicated that, although these three users could talk about how this system might help students, none of them perceived the system as being targeted at them. Generally speaking, there was a sense of ‘it’s nice—but it’s not for me’. Maria and Karina, who actually used the system, agreed that the current prototype would not be appropriate for new students, nor for students who are new to ICTs. It was also suggested, by Karina, that such a tool would be more appropriate to younger learners, implying that mature learners would struggle with OpenEssayist. Robert, who did not use the prototype, perceived the system to be more suitable for students who are less confident. Using these parameters, it would seem that this system is most suitable for a traditional-aged university student in Year 2 or 3 of undergraduate study, who does not feel sure about his or her skills at essay-writing.

8 Discussion

These three stories offer insight into how students might perceive the usefulness, potential and intended audience of OpenEssayist. The current version of the system is somewhat useful in that participants, like Maria, can see the benefit of certain outputs, or views. These three students had several ideas for making the system more helpful, by using it to assist students with essay planning, word choice and by fostering peer support.

When considered alongside other themes in their stories, it seems that what these learners really want is for the processes of essay-writing to be more explicit. Maria prioritizes the planning stages of writing and wants clearer support with essay structure. Robert uses different strategies for approaching his writing but admitted that he struggles with structuring his writing and in choosing the right words. Karina called for a richer, more open discourse around essay-writing in general.

In contemplating the next phase of development, it is germane to question where a system like OpenEssayist fits among systems such as EssayAssist, W-Pal and other AWEs. Perhaps part of the answer will be: OpenEssayist is unique because it makes the essay-writing process obvious. Reflecting on earlier research, it is clear that Maria, Robert and Karina followed processes of essay-writing that were already proposed in Figure 1. Maria’s approach to note-taking and drafting showed that she employed “conversion strategies” from the outset. Robert’s approach included preparing a lot of drafts. Karina, despite her admission that she needed a more rigorous method, actually employed a series of strategies to move her through the process. All of them were able to talk about specific examples of how they used feedback to improve their performance on future pieces of writing.

Moving forward, there is scope for OpenEssayist to be more influential in providing automated assessment for learning. Drawing on Karina’s suggestion, perhaps, this will be possible by creating a narrative about essay-writing around the system that
encourages users to reflect and build on their own processes. In this way, the future of OpenEssayist may still involve scaffolding and skill development, while at the same time will prompt a metacognitive understanding of the development of one’s approaches and strategies.

9 Moving forward

Further evaluation studies are running with postgraduate students studying at the Open University, the University of Hertfordshire and the British University in Dubai. One of the interesting challenges that is being pursued by the team is the role of creating meaningful visualizations that promote “Advice for Action” [4].

Visualizations can promote thinking by helping individuals identify patterns in a set of data, and to promote the discovery of emergent properties that could not have been originally predicted. This is key to the team’s current empirical investigations where they are seeking to identify and refine a set of visualizations for the OpenEssayist system [12].

Another issue that needs attention is whether the user requires training in order to interpret pictures. Although it has been argued that people can interpret pictures without training, the question is still open. In this respect, it is important for the team to understand how diagrams are able to represent concepts unambiguously. The case studies presented in this paper suggest that the visualizations need to emphasize the personalization of the analysis.

Enabling higher education students to receive timely advice about their draft attempts at essay-writing can provide insights into the generic skills of essay writing. This type of feedback also opens to the possibility of not only self-reflection but also engaging in a productive discourse with peers and/or a tutor.

Acknowledgements

The SAFeSEA project team is particularly grateful to The Open University students who participated in this study and the Module Team on the MA in Open and Distance Education for endorsing the study. This work was supported by the Engineering and Physical Sciences Research Council (grant numbers EP/J005959/1 and EP/J005231/1).

References


