

Was our pilot study successful? An approach when data are sparse

Chris Edwards¹[0000-0002-0585-2697] and Denise Whitelock²[0000-0002-9897-4975]

¹ and ² The Open University, Milton Keynes, UK, <https://www.open.ac.uk>

Abstract. Developing new ways to support distance learning students using online technologies can be a challenge at every stage, and the transition from pilot study to embedded offering, available to all students is one of the hardest to complete. Universities are often very supportive of experimentation with new tools but rightly very cautious when approaching the decision to move from exciting project to what is often called business as usual. This paper considers some of the issues that are being experienced on one such development to illustrate some of the issues involved, particularly around evidence of impact where data are insufficient.

We consider a development that has succeeded in its technical aims to make an automatic feedback tool to support students in developing their academic writing but for which a pilot study resulted in fewer data than needed for a confident decision. By employing the Technology Acceptance Model, and the concept of plausibility as an alternative view of uncertainty, we show how it is possible to create indicators of impact. Whilst this does not lead to certainty, it does provide a framework that improves confidence in decision making.

The paper concludes that this approach can enable decisions to be made about progressing with a development in educational technology where the data are insufficient for a statistically robust conclusion to be made.

Keywords: Assessment, Feedback, Automated feedback, Educational Technology.

1 Introduction

The OpenEssayist (OE) was developed over several years, as a collaboration between three UK institutions. Initial trials with real students on a masters module [1] were very encouraging and the developers found evidence that this would add valuable feedback to students as they developed their academic writing. However, the university context at the time meant that there was insufficient institutional interest to support the next stage of development that would make the tool user friendly and more accessible to use.

Five years later, the institutional context had changed to the point that the University was actively supporting the development of a range of technologies that could improve outcomes for students, and the outstanding development of OE, was endorsed with resource to enable this to be completed. As a result, the Open Essay Optimiser (OEO) was created and ready to trial with students at the beginning of 2021.

The planned benefits of using OEO are that it provides consistent, automated feedback on aspects of a student's academic writing, including the coherence of ideas, and the key words and phrases. Together, these will provide students with individual feedback and prompts for reflection on their writing prior to submitting their work for assessment. Therefore, the tool introduces an additional learning opportunity that holds no risk for students.

2 Development of the tool: summary of changes

In the development of OE, functionality was the priority and the user interface was necessarily basic. It was not surprising that feedback highlighted that this could be improved. This was therefore, a major focus for the production of OEO. As a result, the user interface of OEO was carefully designed and taken through a user testing phase prior to release for the pilot study.

Feedback and reflection on OE also led to some improvements within the functionality of each of the elements within the OEO, which was tuned with a further 200 scripts. These central elements are: an argument coherence visualization diagram, which has been shown to be useful [2]; key sentences; key words and phrases; a dispersion chart showing the position of key words and phrases.

Finally, it was possible to add some new features. These are: the ability to save and edit different versions of paper within tool; provide an indication of which module references are used; provide suggestions of potentially relevant text in module reference texts.

3 Piloting OEO

The ideal module for piloting the new tool was a Masters module with innovating elearning as its subject. This module, Openness and innovation in elearning (H817), has been used previously as a test bed for new developments. An intended benefit to students is that they are interacting with the latest developments in educational technology. All the students in the module cohort were invited to participate in the trial of the tool. There was an initial questionnaire, then students were asked to use the tool on the first two module assignments, Ass1 and Ass2. After these, there was a post-questionnaire. The tool remained available for use throughout the module, enabling students to use it for Assessment 4 and the Final Assessment (FA). In addition, students were encouraged to make and send in a screencast of their use of the tool. This was the closest we could practically get to observing the tool in use. Although, requiring additional effort this is also a relevant practical skill and aligned with the ethos of the module.

4 Results and analysis of trial data

We do not intend this section to be an exhaustive setting out of all the data from the pilot study. It includes the data required for the ensuing discussion. The level of engagement of participants with the tool is represented in Table 1.

Table 1, Numbers of students for each pilot activity

Number of students:	Continuing	New	Total
At University start milestone	22	9	31
Completing module	21	9	30
Completing pre-questionnaire	13	5	18
Logging into OEO	12	5	17
Uploading a file to OEO	11	4	15
Uploading a version of Ass 1 to OEO	7	3	10
Uploading a version of Ass 2 to OEO	2	2	4
Completing post-questionnaire	1	0	1
Providing screencasts of use	2	0	2
Reporting technical issues:	3	0	3

1 was properly fixed. All were login issues

4.1 Demographics

Age was the first demographic variable to be compared. The spread of the ages of the two groups are plotted in Fig. 2a below. The spread of biological sex (this is the only value held within University records) are also plotted in Fig. 2b. For both these variables, there is no significant difference between the group of students engaging with OEO and those who do not. In terms of ethnicity, employment, Special Educational Needs (SEN), and career motivation, the two groups, engaged and not engaged, also essentially look the same.

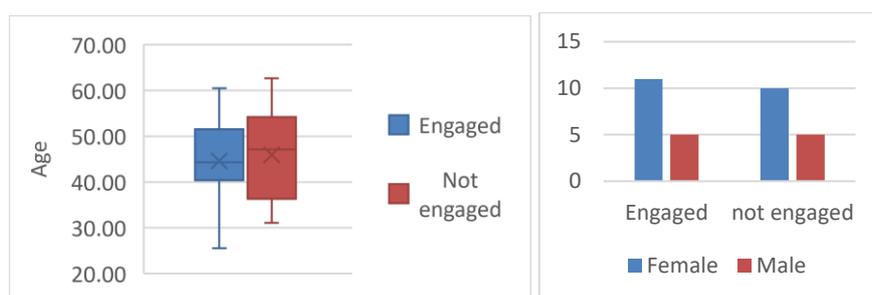


Fig. 2a, Box plots for the spread of ages of the two main groups considered: engaged and not engaged with OEO; **2b**, Sex as recorded in the student record. Plots reproduced here for convenience [3]

The level of engagement of each student is summarised in Table 2. It shows a decreasing level of engagement through the module. One student used OEO for all possible assessments. Three of the sixteen participants used the tool beyond the requested engagement.

Table 2, The points in the module where participants interacted with the OEO. 'N/C' records a New or Continuing student; S=Save a file; E=Edit a file; A=Open the argument coherence diagram. Assessments 1-4 form the continuous assessment element of the module and lead to a combined single score, the Continuous Assessment Score (CAS).

Note: Where students saved and edited a file, and viewed the coherence diagram, there is a mid-grey shading, where they then engage in further editing, this is recorded in white on a dark-grey background. The bold rows indicate a high level of engagement over an extended period. Table reproduced here for convenience [3]

user	student	N/C	Continuous assessment				FA
			Ass1	Ass2	Ass3	Ass4	
7	010	N	SEAE				
8	009	N	SEAE	SEA			SEA
9	029	C	SE				
10	026	C	S				
11	006	C	SEA				
12	027	C	SEAE	S			
13	028	N	SEAE	SEA		SEA	SEA
14	005	C	SEAE				
15	007	C	S				
16	020	C	SEAE				
17	001	C	SEA				SEA
18	011	C	SEAE	SEA			
19	013	C	SE	ES			
21	025	C		S			
23	019	N		SEA			
24	012	C		SEA			

5 Does engaging with the OEO tool impact on results for continuing students?

5.1 Part 1 – mean changes in rank

As these students previously studied a range of modules where there was no formal alignment of assessment between modules, it is likely to be more informative to compare a student's position within the cohort (rank) rather than their actual marks awarded. These values are brought together in Table 3 below. With this small set of values, there is a suggestion that there is a link between using the OEO tool and improved results although nothing that can be demonstrated with any degree of statistical certainty.

Table 3, Comparison of continuous assessment scores (CAS) between previous module and H817: only continuing students. The mean difference in rank shows those engaging with the tool (top section) on average gained 1.2 rank positions, whilst those not engaging with the tool (next section) dropped an average of 3.2 rank positions. The four most engaged users (shaded) averaged 3.1 (in italics) rank places higher for their H817 CAS. User 20 did not engage with OEO. Table reproduced here and simplified for convenience [4]

user	N/C	student	CAS Prev (X)	CAS H817 (Y)	Rank X	Rank Y	D	Mean D
9	C	029	80	80.19	18	9	-9	
10	C	026	71.2	85.92	7	14	7	
11	C	006	71	77.26	6	5	-1	
12	C	027	62	79.77	2	6	4	
14	C	005	77	86.46	13	15	2	
15	C	007	74	85.71	9.5	13	3.5	
16	C	020	70	80.13	4.5	8	3.5	
17	C	001	79	61.12	17	4	-13	
18	C	011	78	88.29	15	18	3	<i>3.1</i>
19	C	013	40	79.78	1	7	6	
21	C	025	77	88.89	13	19	6	
24	C	012	72	82.99	8	10	2	1.2
20	C	022	74	92.22	9.5	21	11.5	
	C	002	82	86.87	20	16	-4	
	C	003	81	54.93	19	3	-16	
	C	004	69	50.5	3	1	-2	
	C	008	78	87.13	16	17	1	
	C	016	77	84.65	13	11	-2	

C	021	85	85.61	21	12	-9
C	023	76	54.28	11	2	-9
C	024	70	91.73	4.5	20	15.5 -3.2

5.2 Part 2 – Using Spearman’s rank coefficient as a measure of impact

We can calculate the non-parametric Spearman rank coefficient for each of the two groups: those that did engage with the tool (Spearman Rank coefficient = 0.35) and those that didn’t (Spearman Rank coefficient = 0.05). This is a measure of the correlation of the pairs of ranks. Whilst the values of the Spearman coefficient are both low, the value for those engaging with the tool is higher than for those that do not. This may suggest an impact associated with the tool. The sample is too small for this to be seen as significant.

6 Discussion

6.1 Considering plausibility

Halpern introduces plausibility as an alternative way to view uncertainty, arguing that this allows belief to be considered alongside statistical predictive measures [5]. For the purposes of this paper, we consider the relevant belief is based on experience and expert knowledge. Therefore, while the data we have available are meagre, there is considerable expertise of students’ learning that can be called upon. We use this idea of plausibility to bring this knowledge into the following discussion.

6.2 Considering the Technology Acceptance Model

The Technology Acceptance Model (TAM) states that in order for a new technology to be accepted, it needs to be perceived as trusted, useful, and accessible to the user [6]. The low level of responses to the questionnaires does not provide sufficient evidence of trust, usefulness, or accessibility. However, we can make some broad assessment of these. We would expect there to be a range of responses for each of these three factors. Given the subject matter of the module is to do with innovating online learning, we might further expect the responses for trust and usefulness to be somewhat higher than in the student body as a whole. However, because there had not been time to fully incorporate OEO within the module, it was clear to students that this was an add on. We had also explained, of course, that there was no obligation to use the tool or take part in the study, and that there would be no impact on their module result if they did not engage with the tool. This, along with the fact that students were studying whilst dealing with lockdowns and the other impacts of COVID-19, may have caused a reduction in participation, rather than a deficit of trust, usefulness and accessibility. On the pilot website and through communications with students, we had stat-

ed that any text students submitted was only accessible to and evaluated by the tool, and not by any person. Also, that the results generated from their use of the tool, were not available to their tutors or anyone else. The exception is that those evaluating the tool did have access to the data but only for the purpose of evaluating the tool. During the module, there had been no comment made by any of the students that they did not trust the tool, or its implementation. So, this is unlikely to feature strongly in reasons not to use it.

Whether OEO is perceived as useful may have been a factor for some not using it. Due to the logistics of teaching out a module and qualification, in this presentation of H817 the students in the pilot study were generally studying their final module towards their Masters qualification, rather than being in the middle of this study. This may mean more of the cohort felt they had already developed their writing sufficiently. However, as the screencasts that students made whilst using the tool made some enthusiastic comments, it was clear that at least some students appreciated the potential use of OEO, suggesting a proportion of undergraduate students would also find it useful. We should note here that one student reported they believed their mark on one assessment was reduced because they used the tool. If this is the case, it would suggest there is still work to do to improve usefulness. However, it is unclear how the student knew what mark they would have received without using OEO. This would be considered in any further testing of OEO.

In terms of accessibility, OEO appeared to give this group of students no challenges in its use. This was a significant improvement on OE. It is clear however, there is still scope to improve accessibility, particularly for some groups with special educational needs. Therefore, some further development would need to be required before a full release of the tool.

When these points are considered together, it would seem there is unlikely to be any issue with introducing OEO to students from the perspective of the Technology Acceptance Model

6.3 Discussion of results

An unexpectedly small cohort for the final presentation of H817 resulted in a lower number of participants than expected. One consequence of this low number is that we cannot claim statistical confidence in any measure of impact in using the tool. We do, however see some encouraging indicators:

1. Three of the sixteen students who used the OEO, did so beyond the assessments they had been asked to (Assessments 1 and 2).
2. When the ranks of continuing students' positions within the presentation cohort of their previous module are compared with those for *Openness and innovation in elearning*, there is some correlation between increase in rank with level of engagement with OEO.
3. The Spearman coefficient of rank also suggests a possible positive impact of OEO.
4. Two students made screencasts of their use of the tool and in the audio commentary were clearly finding they were being prompted to reflect on

how they were linking the ideas and statements within their work, as well as whether they should use further references.

As noted, due to the low number of participants, these statements are not made with any degree of certainty. However, taken together, they appear to further suggest it is plausible that some students could benefit from the automated OEO feedback, and the reflection it promotes. This therefore, supports a further study, with a larger number of participants.

7 Conclusion

By using the concept of plausibility and the Technology Acceptance Model, we find that whilst data on the effectiveness of OEO in improving student grades from this pilot study are sparse, it is plausible to expect the tool would be of use to and benefit a proportion of students. As an automated optional tool that would essentially be on call, whenever a student wanted to reflect on their academic writing, this may be a valuable addition to the student experience. Even if ultimately no strong link between using the tool and overall grade were found, it may well prove to be one element that encourages students to persist with their studies, rather than give up.

Therefore, whilst obtaining sufficient data is always a goal in any trial, this approach can be used to approach situations where the volume of data is less than required for a confident conclusion. On the basis of the analysis in this paper, we would recommend a further trial of OEO, on a much larger cohort.

References

1. Whitelock, Denise; Twiner, Alison; Richardson, John T. E.; Field, Debora and Pulman, Stephen (2014). OpenEssayist: real-life testing of an automated feedback system for draft essay writing. In: #design4learning: from blended learning to learning analytics in HE, 26-27 Nov 2014, Milton Keynes, UK.
2. Foster, Stephen; Whitelock, Denise; Cross, Simon and Kear, Karen (2021). To what extent can graphical feedback from a 'rainbow diagram' help students develop coherence in their academic writing? *PRISM* (Early access).
3. Edwards, C; Whitelock, D; Cross, S and Wild, F (2022a) Developing automated ways to give essay writing feedback to students. In: INTED2022 Proceedings, INTED. DOI: <https://doi.org/10.21125/inted.2022>
4. Edwards, C; Whitelock, D; Cross, S and Wild, F (2022b) Towards autonomous, confident academic writing: supporting students in developing their writing through a stand-alone online tool. In: AHE Proceedings, *In progress*
5. Halpern, J.Y. (2017) Reasoning about uncertainty. Second edition. Cambridge, Massachusetts: The MIT Press.
6. Mortenson, M. J. and Vidgen, R. (2016) 'A Computational Literature Review of the Technology Acceptance Model', *International Journal of Information Management*. Elsevier Ltd, 36(6), pp. 1248–1259. doi: 10.1016/j.ijinfomgt.2016.07.007.