

## **OpenEssayist: Real-life testing of an automated feedback system for draft essay writing**

Whitelock, D.\* (denise.whitelock@open.ac.uk), Twiner, A.\* (alison.twiner@open.ac.uk), Richardson, J.T.E.\* (john.t.e.richardson@open.ac.uk), Field, D.† (debora.field.work@gmail.com) & Pulman, S.† (stephen.pulman@cs.ox.ac.uk)

\*The Open University, Walton Hall, Milton Keynes, MK7 6AA.

† University of Oxford, University Offices, Wellington Square, Oxford, OX1 2JD.

### **Introduction**

Educational institutions continuously strive to offer the best learning experience for their students, and to support them toward achieving their academic potential. In an era where university students are commonly enrolling onto courses with very large numbers, often studying, researching and submitting at least parts of their courses online, the methods by which university tutors can offer support to their students are constantly evolving. This is particularly the case in universities such as the Open University (OU), where the current study took place, as online and distance learning is the norm for most OU students. This paper will focus on OU students' use of a computerised system (OpenEssayist) which provides them with automated feedback on draft essays, developed as part of the SAFeSEA project (Supportive Automated Feedback for Short Essay Answers). We will outline the system itself, and present analysis of observed patterns of activity as a course of students engaged with and explored the system for their module assignments. From this we are able to report a significant positive correlation between the number of drafts submitted to the system, and the marks awarded for the first assignment.

Thus technologies offer great potential to support learning and academic performance, if learning environments are structured in ways that make learning requirements clear to students, and if students understand and are motivated to follow the 'rules' of academic writing. This onus on students is particularly important within the distance and online learning scenario. To return to the context of the present study, a quotation from the Open University website is relevant here:

There is no typical OU student. People of all ages and backgrounds study with us, for all sorts of reasons – to update their skills, get a qualification, boost their career, change direction, prove themselves, keep mentally active. The OU is open to them all. (The Open University's Vice-Chancellor, Martin Bean)

Supporting such a diverse range of students places potentially high demands on OU tutors. Alongside this, UK university fees have been increasing for over 15 years, and so students want to feel they are getting the support they need when they need it. Taras' (2003) research supported this assertion, based on 34 students in a British university doing translation assignments, concluding that students' grade expectations were related to the effort and time invested in doing the submitted work, regardless of students' ability and their perceptions of their ability. Whilst fees for students are increasing however, and students may expect more from their tutors in return, university budgets are being squeezed which places higher demands on teachers and tutors, meaning they often cannot give individual students individualised support. Having an accessible, always-ready online system that can offer students feedback on their work in progress, would be an invaluable

means of starting to bridge this gap. By this argument OpenEssayist appears to offer a relatively simple, but immensely significant, technologically-supported case of demand and supply.

In considering the potential of OpenEssayist to offer such support as outlined above, in this paper we address the following research questions:

1. How did students use OpenEssayist, and which system features proved popular with students in terms of frequency of use?
2. What relationships are evident between system use and essay performance?
3. What rationale do students put forward for their use and non-use of certain system features?

Having outlined the broad educational context, we will now describe the technical system developed for the study, followed by a review of the literature on feedback that informed the design of the system interface.

### ***OpenEssayist***

Many students come to the OU to return to study after some time in the workforce, and so it is common that a significant period of time has passed since their last experience of writing academic essays. It is not surprising therefore that many find this task difficult, and without adequate support may decide to leave their course (Simpson, 2003). This is one crucial reason why a system that can intervene and offer support between students' draft and final submitted essays could be so valuable for students and tutors alike.

OpenEssayist is a computerised system that offers automated feedback for students drafting written academic essays (Whitelock, et al., 2013). It does this through the combination of a linguistic analysis engine, which processes the text in the essay, and a web application that uses the output of the linguistic analysis engine to generate the feedback. The system presents users with feedback on their writing in a number of different ways, including identification of the essay's most prominent words, with graphical illustrations of their use across the essay; identification of the essay's most representative sentences, with hints encouraging the user to reflect on whether these key sentences express, in their view, the central ideas of the essay; and graphical illustrations of the essay's internal structure. Users can move between the different representations, drawing on the features they find most informative and helpful in drafting their work.

Following this outline of the forms of feedback offered by the system, it is appropriate to review the literature on feedback provision in general.

### ***Feedback***

Feedback has been a popular topic of educational research for some time, and it is largely agreed that feedback is central to learning (Black & Wiliam, 1998). Some researchers have however argued that the positive effects of feedback are not guaranteed (Kluger & DeNisi, 1996), and so it is important that research continues to investigate how feedback can be offered in ways that support improvements in students' learning (understanding of topics) as well as performance (grades achieved). Chickering and Gamson (1987) listed 'gives prompt feedback', and 'encourages active learning' as two of their seven principles of good practice for undergraduate education. Therefore by this argument facilitating students to take ownership of and reflect on their work, through provision

of feedback at the point when they are engaging with the topic and task, should have the most impact on students' final submissions and understanding of the topic.

In a similar vein, Butler and Winne (1995) concluded that feedback is most likely to influence performance positively when given during a task, rather than waiting until students submit what they believe to be a finished piece. They also found that students who are better able to make use of feedback can more easily bridge the gap between their expectations or goals and their performance. In terms of making use of feedback, Nelson and Schunn (2009) proposed that understanding problems raised within feedback about one's own work is a critical factor in implementing suggestions. They continued to explore this potential, in stating that understanding is more likely where those giving feedback provide solutions, specify where identified problems occur, and summarise performance. Nelson and Schunn's findings were based on correlation analysis of 1073 segments of undergraduate peer review feedback on writing tasks.

Within their work Nelson and Schunn (2009) identified feedback as involving: motivation, reinforcement and information. They addressed five features of feedback: summarisation; specificity; explanations; scope (i.e. local or global); and affective language (praise, inflammatory or mitigating language). The first four features are classed as cognitive factors, whereas the fifth is an affective factor. It is these five features of feedback that we have drawn on in the present study, in determining the types of feedback to offer on students' draft essays. Referring to the first feature, summarisation, Nelson and Schunn claimed that 'Receiving summaries has previously been found to benefit performance: when college students received summaries about their writing, they made more substantial revisions (Ferris 1997). Therefore, receiving summaries in feedback is expected to promote more feedback implementations' (p. 378). It is on this basis that summarisation is a key feature of OpenEssayist.

Nelson and Schunn proposed that there are 'mediators' that operate between the provision of feedback features, and implementation of suggestions. The authors addressed the mediators 'understanding feedback' and 'agreement with feedback'. They suggested cognitive feedback factors are most likely to influence understanding, and affective factors are more likely to influence agreement. These are then said to influence implementation. Nelson and Schunn's results therefore showed a focus on how understanding feedback is critical to implementing suggestions from feedback. Thus it is important in designing course resources that we consider how to increase the likelihood that feedback is understood, if we want students to make use of it in current and future work – to *learn* from it (and improve performance) by *understanding* it, rather than just *improving one-off performance by blind implementation*.

Just as in Nelson and Schunn's work, peer review feedback is a common strategy used in many universities, which can reduce the workload on tutors whilst also offering an extra opportunity for students to engage with topic material. On the flipside however tutors and course designers need to consider whether students need training to give effective feedback, that feedback given may need monitoring, and that students' motivation to review others' work may vary. Equally, consideration should be given to whether there are concerns of students copying others' work and ideas when doing peer review, whether 'bright' students might hold some points back in their initial drafts so that others cannot 'steal' them, and whether some 'less bright' students might do better in revised versions by using ideas from the students' work they reviewed. In some instances tutors may conclude that none of these issues are concerns, if it means all students have the potential to

improve their submitted work on the basis of their engagement with the topic through peer review. The potential for plagiarism, not directly from peer review however, is one reason why the OpenEssayist system offers visual representations of 'good' essays, rather than showing students 'good' written model essays from past students.

Alden Rivers, et al. (2014) reviewed some of the existing technical systems that provide automated feedback on essays for summative assessment, including E-rater, Intellimetric, and Pearson's KAT. As Alden et al. identify however, most systems such as these three focus on assessment rather than feedback, which is where OpenEssayist offers something unique.

We will now outline our method, including details of the course on which students were enrolled, the student sample and tasks, and will respond to the three research questions within the conclusions that follow.

## **Method**

### ***Course of study***

H817 is an optional postgraduate module contributing to three Masters-level qualifications at the Open University, entitled 'Openness and innovation in elearning'. It was designed to introduce students to the latest educational communication and open education developments. The use of OpenEssayist was therefore highly appropriate with this group of students, interested in the cutting edge of educational technology, and who are required to submit assignments as part of their progression through the course.

### ***Participants***

During their assignment work in the 2013-14 academic year, 41 users who were enrolled on H817 accessed OpenEssayist at least once, using the login details they had been given. The system was opened to the students on 27<sup>th</sup> January 2014, with the final assignment deadline that made use of this resource being 5<sup>th</sup> May 2014 (although access was still available after this point).

In the following results section we highlight patterns and trends in how H817 students used OpenEssayist over time, and across the various system features. We also include data from an interview with a student about their system use.

## **Results**

### ***Draft essays submitted***

Of the users for whom we have both data on when they accessed the system and how many drafts they submitted per session (27 of the 41 users), most (23 users) did not submit a draft on their first visit. The majority of users accessed the system on two, three or four different sessions (11 users, 8 users and 9 users respectively). The access times clustered around submission dates for the assignments, which is what we expected, and shows that OpenEssayist was used to support the essay drafting process.

### ***Duration of use***

For 30 users, we have data on the duration of each session in which they were logged onto the system. From this we can calculate the total time they spent on OpenEssayist (although it is not

possible to ascertain whether they were always engaged and ‘active’ in the system – they may have also been on other websites, or moved away from their computer for a period of time whilst still logged on), and also the mean time spent per session (see table 1 below). For most students having three or more sessions, one session was substantially longer than the others.

Table 1 –Mean session length

| Mean session length  | Number of users |
|----------------------|-----------------|
| Less than 10 minutes | 18              |
| 10-30 minutes        | 6               |
| 31-60 minutes        | 1               |
| 61-2 hours           | 3               |
| Over 2 hours         | 2               |

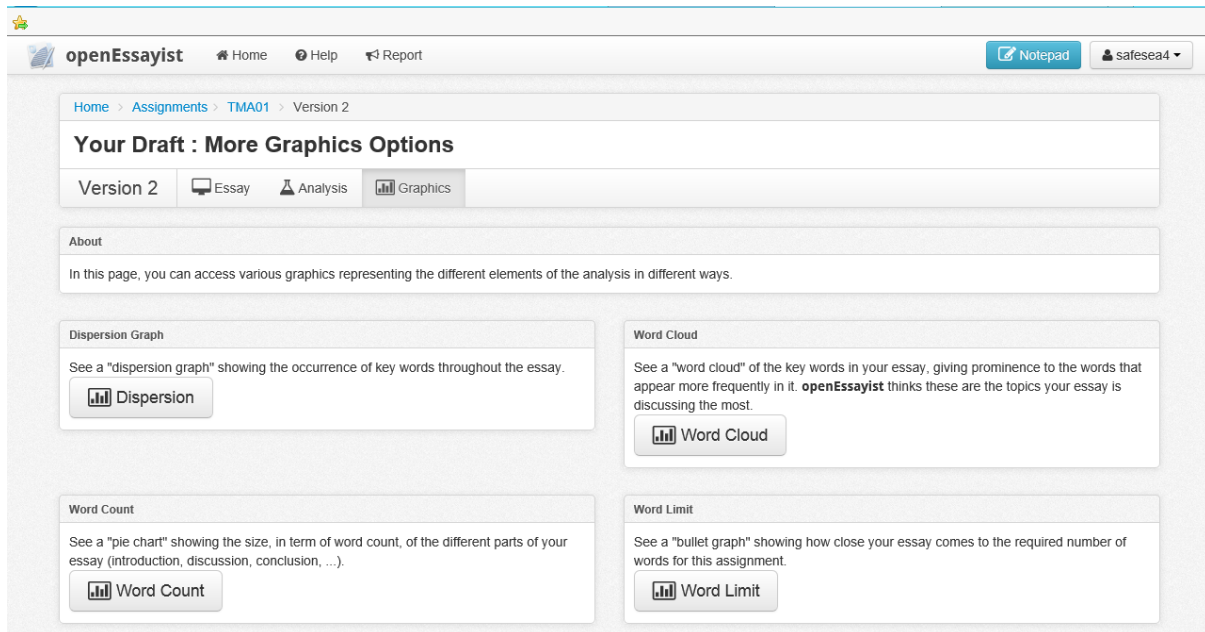
Here is an example of one student’s use. Linking the data across analyses, one user accessed the system on 10 sessions, for a total time period of 31 minutes – giving a mean session time of three minutes. Within these 10 sessions however they submitted seven drafts. Two drafts were submitted in a session lasting 10 minutes, one was submitted in a session lasting eight minutes, and the remaining four drafts were submitted during sessions lasting no more than four minutes. This user was therefore very active in the system during these short periods – but were they using the time well to draft, submit and revise lots of essays (with revision work carried out outside of OpenEssayist), or were they submitting lots of drafts without making much use of the system features to support them in reflecting on and revising their work? Clearly we can only speculate on this issue.

Although this is an example, there is a complicated picture of usage. Some students used OpenEssayist evenly and sparsely, with many short sessions spread over a period of weeks or months. Others used the system in fewer but more concentrated chunks. Others used OpenEssayist very little overall, whilst some used the system – or at least were logged onto it – for long stretches of time. Having considered duration of use and drafts submitted, it is now appropriate to address in more detail which system features students accessed.

### ***System features***

Data was recorded from 35 users on which system features they accessed, and how many times. OpenEssayist was made up of three major components, or ‘landing screens’ – ‘essay’, ‘analysis’ and ‘graphics’ (see figure 1 below). Users are initially brought to these landing screens when they log in, from which they can choose different options.

Figure 1 – The OpenEssayist ‘graphics’ landing screen, with ‘essay’ and ‘analysis’ tabs visible



The first main component, ‘essay’, offered representations of students’ texts concerning the essential elements of an essay: key words and sentences. The screens presented to students the introduction, discussion and conclusion sections of their essay, and identified where key words and sentences occurred throughout the text. The second component, ‘analysis’, pulled the key words and sentences out of the essay text, and offered students the opportunity to organise key words. These options were aimed to encourage students to consider whether their essay contained the key concepts, and development of argument they had intended. The third main component, ‘graphics’, included two different visual representations of key words: as a word cloud, a very common technique for visual analysis of documents which gives prominence to frequent words (Paulovich, et al., 2012), and a dispersion plot, showing the distribution of key words through the essay. The graphics landing screen also offered two ways to view word count: in terms of how the word count was divided across the essay sections, and how a student’s essay word count aligned with the target of the essay.

In terms of students’ use of the system features, all users submitting a draft accessed the ‘essay’ landing screen as this is where a user is automatically routed to once they have uploaded a draft. All but one user accessed at least the landing screen for the analysis and graphics features (this was a different user for the analysis and graphics screens).

Users ranged from accessing one feature (one user), to accessing them all (seven users accessed all 10 features). The majority of users accessed seven or more features (23 users).

Table 2 – Access to system features

|                                 | Essay               |                         |                                   | Analysis                 |                              |                    | Graphics                 |                               |                                 |                            |
|---------------------------------|---------------------|-------------------------|-----------------------------------|--------------------------|------------------------------|--------------------|--------------------------|-------------------------------|---------------------------------|----------------------------|
|                                 | Highlight key words | Highlight key sentences | Highlight key words and sentences | Show extracted key words | Show extracted key sentences | Organise key words | Show key word word cloud | Show key word dispersion plot | Show word count structure chart | Show word limit comparison |
| Users accessing this feature    | 28                  | 29                      | 23                                | 30                       | 28                           | 15                 | 24                       | 28                            | 19                              | 21                         |
| Total accesses for this feature | 115                 | 99                      | 49                                | 105                      | 66                           | 26                 | 103                      | 96                            | 47                              | 59                         |

Of the 35 users and 10 features, this gives 350 potential opportunities of each student using each feature. Of these 350, 105 were zero values (a user did not access a feature), and a further 113 were 1 values, where a user accessed a feature but did not return to it. This leaves 132 occurrences in which a user accessed a feature on two or more occasions. Indeed there were 21 instances of a user returning to a feature five or more times, up to a maximum of 10 visits to any feature (for the ‘highlight key words’ and ‘show word limit comparison’ features).

Of the users who accessed five or fewer features (11 users fell into this category), most accessed the features only once (this made up 81% of access for this category), with a small number returning to any feature 2-4 times (19%). Of the users who accessed more than five features, most accessed them 2-4 times (50% of access for this category), a further 39% of features were accessed just once by these users, and the final 11% of features were accessed five or more times. Thus we can see that users who accessed more features also tended to return to more features. It is important to note however, that a third of all access was for features that were not returned to, and a substantial proportion of features were not accessed at all (as identified by the zero value cells above), even for the higher use students.

Of the essay features, users preferred to show key words highlighted on their text (115 accesses), followed by highlighting embedded key sentences (99), rather than showing both of these aspects highlighted simultaneously on their text (49). Of the analysis features, users mostly requested a list of key words (105 accesses). Some also looked at the list of key sentences (66 accesses), which

offered them a summary of their essay in text or importance order for them to reflect on. For the graphics features, the most popular was viewing the word cloud (103 accesses), closely followed by the dispersion graph (96 accesses).

From this we can see that students largely took advantage of the key word and sentence options. This is significant, as these features highlight key concepts within an essay, and the representations allow students to consider if their essay structure and content presents a coherent argument. Students are given information on the spread of ideas throughout the essay, and of the connectedness and development of concepts across their introduction, discussion and conclusion, enabling them to reflect on what may be missing and how they can make best use of the word count. These are important aspects in supporting students with the content and structuring elements of essay writing, and in helping to prevent students from over-writing, for which they could lose marks. Substantially fewer students used the option to re-group their key words, but it was apparent that those who used more system features then started to explore more of what they could do through the system.

We now present data analysis on how use of OpenEssayist related to students' marks on their assignments.

### ***Comparison of marks and use of OpenEssayist***

There are 30 participants for whom we have marks for essay 1, and 14 participants with marks for essay 2. The data from three participants was removed from the analysis, as the lengths of time for which they were logged into the system were substantially longer than the other participants, leading to a high likelihood that they were doing other things whilst still logged into the system.

Correlational analysis of this data revealed three statistically significant relationships. First, there is a positive correlation ( $r = +0.41$ ) between marks for essay 1 and number of drafts. This could mean that submitting more drafts leads to higher marks, or simply that brighter students submit more drafts. Second, there is another positive correlation ( $r = +0.65$ ) between the number of visits and the number of drafts. Finally, there is a positive correlation ( $r = +0.60$ ) between the mean time and the total time. The latter two are perhaps not surprising: in that people who visited the system on more occasions submitted more drafts, and those who were logged onto the system for a longer overall time also spent longer logged onto the system per visit.

Importantly, there was no significant difference found between the mean mark for essay 1 and the mean mark for essay 2. For the 13 students who submitted both TMAs, their mean marks were 70.9 for essay 1 and 73.3 for essay 2. This suggests that participants performed consistently on both essays.

### ***Comments from interview with a student about use of OpenEssayist***

We were able to conduct one interview with a student who had used OpenEssayist on this course. The student had used the system for both module assignments, and had used it for a further two since then. His first expectation was that it would give him an initial kind of feedback as a mark, before submission, so it was not initially what he had expected. In light of this he felt it perhaps was not as helpful to him as he hoped on his first essay. He continued to explore it however, and realised the value of a few of the features, and now finds it very helpful in clearing up structure, use of key



phrases and sentences in particular. He felt it has altered how he goes about essay writing, in terms of structuring and dividing the essay among the different sections as well as dividing his time across the various elements of the essay. He says his marks have improved too.

He also thought the system was potentially more useful to students who were not so confident at essay writing. He was disappointed however that once he finishes H817 he will not have access to OpenEssayist anymore, and feels it should be made more widely available:

Obviously I am using this, I've got access to this approach, simply because I am an Open University student doing a particular module. Knowing what I know about it now in some ways I think it's a pity because I think it's a more generally applicable tool... So I would just say I think it might be more widely available. Because after this module I won't have access to it, whereas actually I might have gone on and used it again in further assignments and courses.

It is also promising to report that one of the tutors aligned with this course commented that the cohort of students participating in this study had performed better than expected in their assignments. Again we can only speculate that OpenEssayist may have had a supportive role in this.

## **Discussion**

As part of the SAFeSEA project, we created a technical system to offer automated feedback on draft academic essays, called OpenEssayist. The system is based on research findings on feedback features that have been found to support students' understanding, and implementation of feedback suggestions (Nelson & Schunn, 2009). It was considered particularly important that students be offered feedback at the draft stage of writing, when they still had the opportunity to act upon suggestions made.

There are numerous potential benefits of using automated systems that can offer feedback, for instance in reducing the demands on tutor time, and in providing a support system that is available when students need it. Much research has been carried out on systems providing automated summative feedback or just assessment, but not formative feedback that allows for revision before submission. This is where OpenEssayist is unique in its ability to support students in writing better essays first time around, rather than receiving feedback when it is potentially too late for them to do anything.

From the data collected on students' use of OpenEssayist during their course assignments, there was wide variety in how frequently and for how long students used the system, and also which features they accessed and returned to. It is interesting to note that some students accessed OpenEssayist after the period in which it was used for their course, and even submitted drafts after this point. We can only guess that they were perhaps using it for other assignments. A comment made in interview suggests that this may well have been the case. The majority of students accessed at least seven system features, and those who accessed more features also tended to return to more features. Features concerning key words – highlighted on the essay text, extracted, and shown as a word cloud or dispersion – seemed to be particularly popular, followed by highlighting key sentences and extracting these as a summary of the essay.

From interview data, a student indicated that he felt OpenEssayist would be more useful to those less familiar with academic writing, but when he invested more time to explore the features and trial them with his own work he could see more potential benefits and uses for himself.

Correlation analysis of students' essay marks and system use revealed some interesting findings. A significant positive correlation was found between students' marks for essay 1 and the number of drafts they submitted. Number of drafts submitted was also positively correlated with number of visits, and the mean time spent in the system per visit was positively correlated with the total time spent logged onto OpenEssayist. Whilst we cannot establish causal relationships with correlation analysis, the first of these findings could be particularly important, in that perhaps those students who submitted more drafts gained higher marks, or that those students who tend to get higher marks also engaged more with the process of submitting drafts. This is an interesting issue for further research, and a thread we will pursue in our ongoing work.

### **Conclusions and implications**

OpenEssayist is unique in being an automated feedback system that has been developed to offer feedback on students' draft essays, rather than assessment on their finished work. This is therefore a system that offers opportunities for students to engage with and reflect on their work, and to improve their work through understanding of the requirements of academic essay writing. In trialling use of the system in a genuine Open University course, we found that students made use of it to varying degrees, which is perhaps likely with any study resource. Those who took the time to explore system affordances and what they could be used for however tended to report more positively on its perceived value. From our analysis we were also able to conclude that a significant positive correlation exists in this sample of students between marks on essay 1 and the number of drafts submitted. We could speculate as to what this may mean for this set of students, or more widely, but it seems clear that use of a system such as OpenEssayist has many potential advantages to students and tutors, which will benefit from further research and exploration.

Moving forward, as OpenEssayist is designed to offer feedback to students during the drafting process, this has considerable implications for supporting students to improve their work, and also supporting students to *believe* that they can improve their academic work. This is no small feat for learners who may often feel isolated and stretched trying to squeeze study around other commitments and demands on their time.

### **References**

Alden Rivers, B., Whitelock, D., Richardson, J.T.E., Field, D. & Pulman, S. (2014). Functional, frustrating and full of potential: learners' experiences of a prototype for automated essay feedback. Proceedings of *Computer Assisted Assessment international conference: Research into e-assessment*, Zeist, The Netherlands, June 30<sup>th</sup>-July 1st (pp. 40-52).

Bean, M. <http://www.open.ac.uk/about/main/mission>. Accessed 5th June 2014

Black, P. & Wiliam, D. (1998) Assessment and classroom learning, *Assessment in Education*, 5(1), pp. 7-74.

Butler, D. L. & Winne, P. H. (1995) Feedback and self-regulated learning: a theoretical synthesis. *Review of Educational Research*, 65(3), 245-281.

Chickering, A. W. and Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education, Racine, WI, *Wingspread Journal*.

Ferris, D.R. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly*, 31(2), 315–339.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254-284.

Nelson, M.M. & Schunn, C.D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science*, 37(4), 375-401.

Paulovich, F.V., Toledo, F.M.B., Telles, G.P., Minghim, R. & Nonato, L.G. (2012). Semantic wordification of document collections. *Computer Graphics Forum*, 31(3), 1145-1153.

Simpson, O. (2003). *Student Retention in Online, Open and Distance Learning*. London; Sterling, VA: Kogan Page.

Taras, M. (2003). To feedback or not to feedback in student self-assessment. *Assessment & Evaluation in Higher Education*, 28(5), 549-565.

Whitelock, D., Field, D., Pulman, S., Richardson, J.T.E. & Van Labeke, N. (2013). OpenEssayist: An automated feedback system that supports university students as they write summative essays. *The 1st International Conference on Open Learning: Role, Challenges and Aspirations*, The Arab Open University, Kuwait, 25-27 November 2013.