Supporting student writing with an intelligent tutoring system for assignment checking

Peter Vitartas
La Trobe University
Australia

Sarah Midford
La Trobe University
Australia

Anastasia Kanjere
La Trobe University
Australia

In this paper we present the results of a prototype system designed as a draft assignment checker that students can use prior to the submission of their assignments. The tool was designed to provide descriptive timely feedback to students on their digitally submitted text-based assignments. This process allows students to submit draft versions of their assignments, obtain feedback and improve them before they make a final submission for marking. Students are able to access the results and descriptive feedback generated for the assignments they have uploaded and the software allows customisation of the evaluation measures based on the type of assignment and expectations of the academic staff. Findings from a survey of student feedback on the system are presented. Overall students found the system useful, but the tool needed to be incorporated into the assignment preparation process more closely to be effective.

Keywords: Data analytics, Assignment feedback, Evaluation

Introduction

The incorporation of technology into education has quickly become a serious consideration for educators and educational designers. Learning Management Systems, CD ROMS and online learning software have been leaders in the move from chalk and whiteboards to more serious considerations of education data mining and machine learning that enlist the power of computing for the purpose of teaching and learning design. The ability to analyse student behaviour in terms of LMS page views, log-ons, keystrokes and automated processes that deliver adaptive learning are just part of the current arsenal at the disposal of modern educators. Indeed, learning analytics is now considered a “thing” with the proceedings of the annual Learning Analytics and Knowledge conference (LAK) now ranking in the top six publications in the 2019 Google Scholar metrics for Education Technology (SOLAR, 2019).

An early development in the Education Technology area has been the emergence of Automated Writing Evaluation (AWE) systems and more recently the opportunity it has provided as a forerunner to intelligent tutor systems (Vitartas et. al., 2016). Most of the existing AWE systems have been designed for schools or as tools to assist in graduate entry testing rather than for use in higher education. In this paper we present the results of a prototype system designed as a draft assignment checker that students can use prior to the submission of their assignments in the higher education context. We start by providing a brief introduction to AWE’s then provide details of the prototype trialed in this research before presenting preliminary results and feedback from students who have utilized the system.

Automated Writing Evaluation Systems

AWE is also known by the acronyms AES (Automated Essay Scoring), AEG (Automated Essay Grading) and AEE (Automated Essay Evaluation) (Hockly, 2018). Commentators with a less optimistic orientation towards the use of technology for this purpose are more likely to use the term ‘machine scoring,’ as, for example, in (Herrington & Moran, 2012). A history of AWE generally begins with the work of English teacher turned researcher Ellis Page, and his Project Essay Grade (PEG) beginning in the 1960s. In a 1966 article in The Phi Delta Kappan, Page insisted that “we will soon be grading essays by computer, and this development will have an astonishing impact on the educational world” (Page, 1966, p. 238). Ellis had an optimistic vision for writing feedback being provided to students much more extensively and in a timely manner than could be achieved by English teachers in school or college.

Page’s vision is now much closer to reality in various automated formative tools, offering immediate descriptive feedback on writing (e.g. WriteLab, Turnitin’s Revision Assistant, Pearson’s WriteToLearn™, ETS’ Criterion®, and Vantage Learning’s MyAccess!). In addition to these proprietary products, a number of freely available services from the academic domain are also available to examine text and extract phrases including Coh-Metrix, WordNet, TerMine, MALLET Stanford Core NLP and Natural Language Toolkit. In 1999, E-rater, a tool developed by the Educational Testing Service, was used in the General Management Admissions Test, making it the first AWE to be used in a high-stakes assessment situation (Zhang in O'Leary, Scully, Karakolidis, & Pitsia, 2018, p. 162)

The field of AWE has made significant advancements and has attracted an impressive body of both scholarly and commercial interest. In 2003, researchers Mark Shermis and Jill Burstein edited a collection of work on the subject called Automated essay scoring: A cross-disciplinary perspective (Mark D. Shermis & Burstein, 2003). In their updated version, published in 2013, the title shifted to Automated Essay Evaluation: Current Applications and Directions. As Carl Whithaus points out in the foreword, “The shift indicates that feedback, interaction, and an altogether wider range of possibilities for software is being envisioned in 2012 than was seen in 2003” (in M. D. Shermis & Burstein, 2013, p. viii).

Newer possibilities for the application of AWE software include an increased focus on more complex forms of feedback and ‘feed-forward’ which support the learning process. Systems like WriteLab and Turnitin’s Revision Assistant have focused on the iterative nature of writing and on providing formative feedback, rather than grades, in order to encourage students to revise and rewrite their work. The most recent scholarship is all focused on using this technology as a learning tool which provides feedback and encourages revision (Ajetunmobi & Daramola, 2017; Allen, Likens, & McNamara, 2018; Bektik, 2017; Knight, Buckingham Shum, Ryan, Sándor, & Wang, 2018; Nathawitharana et al., 2017; Roscoe, Wilson, Johnson, & Mayra, 2017; Shibani, Knight, Buckingham Shum, & Ryan, 2017; Vitartas et al., 2016). Automated feedback can be embedded into discipline-related skills in order to become a valuable teaching and learning tool to develop writing (Shibani et al., 2017). Significant research interest has also been devoted to the use of AWE in teaching and finessing the acquisition of second languages, particular in relation to English as a Foreign Language (Bai & Hu, 2017; Huang & Renandya, 2018; Ranalli, 2018; Ranalli, Link, & Chukharev-Hudilainen, 2017).

Methods

The Next Generation Rubrics (NGR) project (Vitartas et al. 2016) was established at an Australian university initially as a proof of concept but then further developed into a working prototype. The tool was designed to provide descriptive timely feedback to students on their digitally submitted text-based assignments. Drawing on the concept of a marking rubric, or a “a scoring guide used to evaluate the quality of students’ constructed responses” (Popham, 1997, p.1), a set of evaluative criterion and guidance on expectations for the criterion were incorporated into the tool.

The main functionality of the software is to assess digitally submitted assignments and providing descriptive feedback to students. This process allows students to submit draft versions of their assignments, obtain feedback and improve them before they make a final submission for marking. Students are able to make multiple submissions to the assignment checking tool prior to submission. Upon submission of an assignment to the system, the software conducts content analysis, evaluates the assignment based on an evaluation criterion developed in association with the academic, and generates feedback via graphically based dashboard that highlights areas that could require further improvements. Students are able to access the results and descriptive feedback generated for the assignments they have uploaded, and the software allows customisation of the evaluation measures based on the type of assignment and expectations of the academic staff.

Study Design

The tool was made available to students enrolled in two subjects. The first was an introductory first year subject in the Bachelor of Arts (BA) that included 383 enrolments. As part of the assessment tasks, students were required to submit a 1500-word critical assignment in essay format. The second subject was a first-year subject from the Bachelor of Business with 271 enrolments. It required students to submit a 1500-word report on a macroenvironmental analysis of a manufacturing industry.

Information on the tool, instructions for its use and links to an external site hosting the software were posted on the subject’s learning management system (Moodle) assignment page. The introduction of the software was supported by internal emails to students and a briefing with staff undertaking the tutorials for the subjects. As the tool was relatively new and only in the prototype stage of development the use of the tool was made optional and there were no incentives or requirements for students to use the system. This may have limited the uptake of the tool, but it also provided insight into the interest and support among students for such a system. Ethics approval was obtained from the La Trobe University Human Research Ethics Committee.

Participants and Setting

A total of 35 students, 19 from the Business School and 16 from the School of Humanities and Social Sciences (HUSS) used the tool for one of their assignments during the semester. This represented a relatively low take-up as the subjects were relatively large. For example, the response rates were 7% and 4% for the Business and HUSS students respectively. However, it should be noted that there was no compulsion or promotion of the Tool's availability to students other than having an information link on the LMS site and so it was only those students who were self-motivated that engaged with the tool. In addition, the HUSS students were spread across regional campuses which may also have accounted for a lower take-up rate.

Students were able to submit their draft as many times as they liked. This allowed them to make adjustments to their assignment, then to recheck it. In this way a type of learning takes place by having students made aware of potential errors in their work and checking if the changes they make address the issues. This type of immediate feedback also has the advantage of learning in context as they have the feedback immediately and can see improvements based on their actions.

The total number of submissions by the HUSS students was 41, an average of 2.7 submissions for each student. However, several students only submitted once while others took advantage of the tool for multiple submissions. A similar approach was found with the Business Students, where there were 37 submissions, an average of 2.0 per student. The majority of students submitted only once, however a small number of students submitted more than five times. In the case of HUSS one student submitted six times while a Business Student submitted ten times.

Data Collection

At the end of the semester a survey was sent to the 35 students who had used the tool seeking feedback on their experience. Twelve students responded – a response rate of 34%, four from each subject and four who sought to remain anonymous. The following section reports on the findings to an online survey administered through qualtrics. An inducement to go in the draw for one of five $30 shopping vouchers was used as an incentive for the sample to respond. The survey consisted of 11 multi-part fixed and open-ended questions and took an average of 11 minutes to complete. The questions sought to identify the type of computer systems students used when accessing the tool, their experience with four features of the tool, the type of feedback they used, ratings for the feedback elements and opinions about the tool's usefulness and whether they would recommend the tool.
The majority of students (58%) used PC's to access the system while two (17%) used Mac's and three (25%) used a mobile device (phone or tablet). There appeared to be no differences in responses or difficulties encountered based on the system they used to access the feedback tool.

Results

Respondents were asked 'How easy or difficult did you find…' four aspects of using the system. These included their initial login to the tool, locating where to submit their assignment, uploading their assignment, and interpreting the feedback. Responses were recorded on a five point 'Extremely easy' through to 'Difficult' scale. The responses are reported in Table 1. The majority of students found the tool easy to use on all four aspects. However the majority of students found interpreting the NGR feedback only moderately easy rather than extremely easy. This may have been because some of the statistics were new to them and they had to read the detail on the feedback report to interpret the information. Only one student found the aspects of the initial login and interpreting the results as difficult. It is believed that some students encountered problems with the system because they were using the system off campus or had not logged on through the University system.

Table 1: Ease of use for steps in system use

How easy or difficult did you find:	Extremely Easy	Moderately Easy	Neither Easy nor Difficult	Slightly Difficult	Difficult
...your initial login to the Assignment Checker	41.7	41.7	8.3	8.3	0.0
...locating where to submit your assignment in the tool	50.0	33.3	16.7	0.0	0.0
...uploading your assignment(s)	66.7	25.0	8.3	0.0	0.0
...interpreting the feedback	41.7	50.0	0.0	8.3	0.0

Students were asked "What aspects of the feedback did you find most helpful" for three aspects of the tool. These included statistics on the assignment, the gauge showing target performance and the description of rubric criterion. Responses were reported on a five-point scale of 'Very helpful' through to 'Very unhelpful'. See Table 3 for the results. The majority of respondents indicated all aspects were either very helpful or somewhat helpful. One student indicated statistics on the assignment and the gauge showing target performance very unhelpful while another considered the gauge and description of the rubric criterion neither helpful nor unhelpful. It would appear the description of the rubric criterion was indicated as the most helpful aspect of the feedback.

Table 2: Most helpful features of the tool

What aspects of the feedback did you find most helpful?	Very helpful	Somewhat helpful	Neither helpful nor unhelpful	Somewhat unhelpful	Very unhelpful
Statistics on the assignment	33.3	58.3	0.0	0.0	8.3
Gauge showing target performance	33.3	50.0	0.0	8.3	8.3
Description of rubric criterion	58.3	25.0	8.3	8.3	0.0

Table 3 presents the results for student's ratings of ten statistics provided to students as part of the feedback. Responses were recorded on a five-point scale of 'Extremely Useful' through to 'Not useful'. Students found the word count, grammar error count and count of formatted in-text references to be the most useful of the statistics provided by the tool. The readability and spelling error count were also rated highly. The least useful measures were the discipline coverage and critical thinking term coverage. It is believed there may have been some misunderstanding among students of these tools based on qualitative comments and feedback on the tool.

Table 3: Usefullness of writing statistics

How useful did you find each of the following measures:	Mean	Extremely useful	Moderately useful	Neither useful nor not useful	Slightly not useful	Not useful	N/A
Word count	1.91	54.5	18.2	18.2	9.1	0.0	0.0
Spelling error count	2.00	63.6	9.1	9.1	9.1	0.0	9.1
Grammar error count	1.91	63.6	18.2	0.0	9.1	0.0	9.1
Referencing - count of formatted intext references	1.91	36.4	54.5	0.0	0.0	9.1	0.0
Referencing - count of number of references	2.36	27.3	54.5	0.0	0.0	9.1	9.1
Critical thinking coverage	2.45	27.3	27.3	27.3	9.1	9.1	0.0
Readability score	2.00	45.5	18.2	27.3	9.1	0.0	0.0
Discipline content coverage	2.18	45.5	18.2	18.2	9.1	9.1	0.0

Qualitative responses

Students were also asked if there was a measure they would have found useful that wasn't included in the tool. One student indicated that it would be helpful to have some tips to help improve the work and an interpretation of the gauges. Another thought they would like to have the tool look more like the Turnitin tool which highlights areas where there is similarity detected.

When asked whether they thought their reworked assignment was improved after using the assignment checker the majority said yes. For example, "Yes it helped me see where I was going well and where the mistakes were" (HUSS Student). Another replied "Yes, helped stay on point, made me re-think what I needed to change to meet the criteria" (Business Student). While a HUSS student noted "It made me realise I hadn't used enough critical thinking and topic terms, so I kept improving until I got the desired outcome on the gauge."

Students were also asked two summary questions about the tools usefulness and whether they would recommend the tool to others. For both questions the majority of students indicated they found it either extremely useful or very useful (64%) and most would recommend the tool to others – the net promoter score was 45.5 indicating a positive score.

Discussion and Conclusions

For the majority of the small sample of students who completed this study it was found the assignment checker was either very useful or extremely useful and they would recommend it to their colleagues. This is despite some students indicating that they did not find the tool useful. Of particular interest in this study was the small number of students who took the opportunity to use the tool. While the tool was available well before the assignment due date, it would appear that many students did not complete their assignments in time to use the tool, or could only use it once before the due date. The benefits from using assignment feedback tools such as the assignment checker will only occur if the tools are incorporated into the planning and development phases for the assignment preparation. This would require students understanding the benefits of using and editing their work prior to submission dates.

References

Ajetunmobi, S. A., & Daramola, O. (2017). Ontology-based information extraction for subject-focused automatic essay evaluation. Paper presented at the 2017 International Conference on Computing Networking and Informatics (ICCNI) 29-31 Oct.https://doi.org/10.1109/ICCNI.2017.8123781

Bektik, D. (2017). Learning Analytics for Academic Writing through Automatic Identification of Meta-discourse. The Open University

Herrington, A., & Moran, C. (2012). Writing to a machine is not writing at all. In N. Elliot & L. Perelman (Eds.), Writing assessment in the 21st century: Essays in honor of Edward M. White (pp. 219-232). New York: Hampton Press.

Nathawitharana, N., Huang, Q., Ong, K.-L., Vitartas, P., Jayaratne, M., Alahakoon, D., . . . Ahmed, T. (2017). Towards Next Generation Rubrics: An Automated Assignment Feedback System. Australasian Journal of Information Systems, 21. https://doi.org/10.3127/ajis.v21i0.1553

Page, E. B. (1966). "The Imminence Of... Grading Essays by Computer," The Phi Delta Kappan (47:5), pp. 238- 243.

Shermis, M. D., & Burstein, J. C. (Eds.). (2003). Automated essay scoring: A cross-disciplinary perspective. Hillsdale, N.J.: Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410606860

Vitartas, P., Ahmed, T., Alahakoon, D., Midford, S., Nathawitharana, N., Ong, K. L., and Sullivan-Mort, G. (2016). "Using Learning Analytics to Guide Learning: An Analysis of Marketing Assignments," Proceedings of the Australian and New Zealand Marketing Academy Conference (ANZMAC), Christchurch, pp. 623-629.

Other references available from the first author on request.

Please cite as: Vitartas, P., Midford, S. & Kanjere, A. (2019). Supporting student writing with an intelligent tutoring system for assignment checking. In Y. W. Chew, K. M. Chan, and A. Alphonso (Eds.), *Personalised Learning. Diverse Goals. One Heart. ASCILITE 2019 Singapore* (pp. 588-592).