"The score doesn't mean much": Students' and staff understandings of text matching software

Lee Adam, Carol Bond
Higher Education Development Centre
University of Otago

Ari Samaranayaka
Dunedin School of Medicine
University of Otago

We report on a 2012 study exploring students' and staff understandings of SafeAssign text matching software. A total of 326 students and 216 staff at a New Zealand university responded to a questionnaire asking about their knowledge of SafeAssign. Although SafeAssign does not identify plagiarism, 90% of students and 70% of staff thought that it did, indicating confusion about the difference between text matching and plagiarism. Moreover, both students and staff assumed that SafeAssign can do more than it does and responses indicated a lack of understanding of the sources with which SafeAssign is able to match text. Furthermore, many students were not aware that SafeAssign was being used in their papers, and some thought its function was simply an assignment submission platform. These results indicate that staff and students need more information about what SafeAssign does in order to make more effective use of the software.

Keywords: Plagiarism, SafeAssign, text matching software

Introduction

Many higher education institutions worldwide have turned to the use of text matching software (Badge, 2010) in response to concern over the increase in prevalence of student plagiarism (Atkinson & Yeoh, 2008; Park, 2003). The use of text matching software (TMS), designed for the specific purpose of providing a tool to assist in the detection of students' plagiarism, has increased over the last decade. There are a wide variety of tools available, including Turnitin, EVE2, and SafeAssign (SA). The use of TMS is based on users' assumptions about its function and usefulness. But how accurate are these assumptions? The current research exposes the assumptions of students and staff about the functions of SA, and shows that the software may be being (mis)used for functions beyond that which is intended (Bond, Adam, Samaranayaka, Finn, Senior, Cross & Deaker, 2013).

The emergence of TMS has occurred in parallel with an exponential increase in knowledge, increased access to information as a result of the Internet, plus the impact of increasing participation in higher education. One effect on academic staff is an increase in workloads, and particularly marking. Where once it was possible to recognise the writing styles of individual students and detect plagiarism manually, in many classes this is no longer the case (McKeever, 2006). The growth of the Internet has not only made it easier to plagiarise, but also easier to detect plagiarism using web-based plagiarism detection software (Evans, 2006; McKeever, 2006). However, the literature on such software tends to be limited to a small pool of empirically based research on effectiveness, institutionally based evaluation studies (Bond et al, 2013) students' and staff perceptions (e.g., Atkinson & Yeoh, 2008) and opinion pieces. Much of the literature focuses on the tool Turnitin® - a package that provides a similar service to SA, and is reportedly widely used (Badge, 2010).

Most TMS have at least three claimed or emphasised functions: text matching (sometimes called the 'detection' of plagiarism, or plagiarism prevention); the deterrence of plagiarism; and an educative tool useful in teaching academic skills such as paraphrasing and citation (Green et al, 2005). The text matching function of TMS requires the software to compare an electronic copy of an individual's work by searching for matches in an on-line environment. The software searches for "strings of text" (McKeever, 2006, p. 157) and produces an 'originality report' denoting the percentage of matched text, a hyper-link to the possible sources, and a percentage score allocated to the number of words that match another source. Thus, the name 'plagiarism detection software' is a misnomer as TMS does not identify plagiarism but only matches text submitted to it with text that it can access in its archives or on the Internet (Ledwith & Risquez, 2008). The deterrence function of TMS is reliant on the perception that implementing the software will reduce plagiarism as students' knowledge that the tool is being used acts as deterrence (Badge et al, 2007). The educative function of TMS is dependent on students' knowledge of plagiarism and what the TMS does, alongside using the TMS report to identify the problem, and the student being allowed the opportunity to resolve the problem and resubmit

(McKeever, 2006). However, research shows that few staff use TMS in an educative way (Löfström & Kupila, 2013).

Methodology

This mixed methods study explored students' and staff understandings of SA. The research was part of a larger study, requested by the University's Committee for the Advancement of Teaching and Learning (CALT), that also explored students' and staff understandings of plagiarism and their experiences with SA. The purpose of the research was to inform the management of plagiarism and SA the University. A sample of students and staff at the University were asked to respond to a questionnaire. The student sample was selected using one-stage cluster sampling. Students enrolled in seven papers using SA in 2012, representing a range of undergraduate paper levels from Health Sciences, Business, and Humanities, were surveyed. Although Science papers were not represented as SA is not commonly used in these papers, the sample included a first year service course that attracts students from a range of different programmes including sciences. Students were administered an in-class, paper-based questionnaire. Because the selected papers used SA, an assumption was made that all the students were familiar with SA. Students were assured that the survey was anonymous and they could choose whether to participate or not. Questionnaires were collected and data from the 326 respondents were manually entered into an Excel spreadsheet.

The staff sample comprised those who chose to respond to an open email invitation to complete an online questionnaire, making this a self-selected sample. Responses were received electronically and managed by a University staff member outside of the research team. Identifiers were removed and anonymised data from the 216 respondents were supplied to the research team electronically on an Excel spreadsheet.

Where possible similar questions were used for students and staff to allow a comparison of views. The wording of the two questionnaires was modified to suit the different experiences of students and staff. 'Likert like' scales that included a neutral position and a 'don't know' response were used. Open-ended questions were used to generate more qualitative data. General understandings of what SA did and how it worked were obtained by questions to all student and staff respondents whether they were users or not.

Analysis

The quantitative data generated by the two surveys were separately subjected to a simple descriptive analysis. Frequencies of Likert scale items were aggregated, for example, 'strongly agree' and 'agree' were treated as one. All analyses were completed using the 'svy' suite of commands in Stata12.1® software (2011). Descriptive statistics such as percentages of respondents with relevant characteristics were generated. Confidence intervals were calculated that accounted for the specificities of the corresponding sampling design such as 'sampling weights' and possible 'cluster effects' by students in the same paper. In relevant situations, percentages were compared for staff and students' samples (e.g., Austin & Hux, 2002). The qualitative data generated by the open-ended questions and requests for comment were analysed and loosely categorized according to topic using a simple text analysis (Bond et al, 2013).

Students' and staff understandings of SafeAssign

All the students in the student sample were enrolled in papers that used SA, however many students were unaware that SA was being used in their papers. Only 258 of the 326 student respondents (79%) identified that they had used SA, and some of these students indicated their belief that SA's main function was as a platform for assignment submission. Of the 216 staff who responded to the survey, only 97 (45%) used SA in the papers they taught, however all staff were asked about the functions of SA.

Students' and staff responses about the functions of SA are displayed in Figure 1 below. Most students and staff displayed awareness of the main functions of SA as text matching and production of a percentage score. Slightly more students (84%) than staff (79%) strongly agreed or agreed that SA identified words and phrases that matched words and phrases in published text. However, fewer staff (59%) than students (74%) were aware that SA provides a percentage score of matched text. In the staff responses, there was little difference between the responses of users and non-users for these two items, but all of the 'don't know' responses belonged to non-users. One user commented, "SA can also indicate very good referencing – quotations etcetera will 'match' so you have to look at the highlighted document – the 'score' doesn't mean much."

A very high proportion of students (90%) but fewer staff (70%) thought that SA identified plagiarism. Again

there was little difference between the responses of staff users and non-users, and all the 'don't know' responses belonged to non-users. One staff noted, "SA can detect other students' work but only if it has been previously submitted to SA. However it cannot always tell you the name of the original student." One student commented, "All I know is that it checks for plagiarism."

Students and staff were less aware of other aspects of SA, and their responses also differed from each other in varying degrees. About 43% of students and 32% of staff thought that SA was able to indicate the need for a reference, though many respondents disagreed with this assertion - approximately 16% of students and 31% of staff indicated that they did not know this was the case. Staff user 'don't know' responses accounted for 27%. Only 24% of students thought that SA provided information on parts of the assignment that could be changed, compared with 35% of staff. Slightly more staff (15%) than students (8%) thought that SA taught students about referencing, however 32% of staff indicated their belief that SA identified poor referencing. When asked if SA deterred students from plagiarising, 61% of staff either agreed or strongly agreed. One student responded, "I do not know anything about it, but would like to learn."
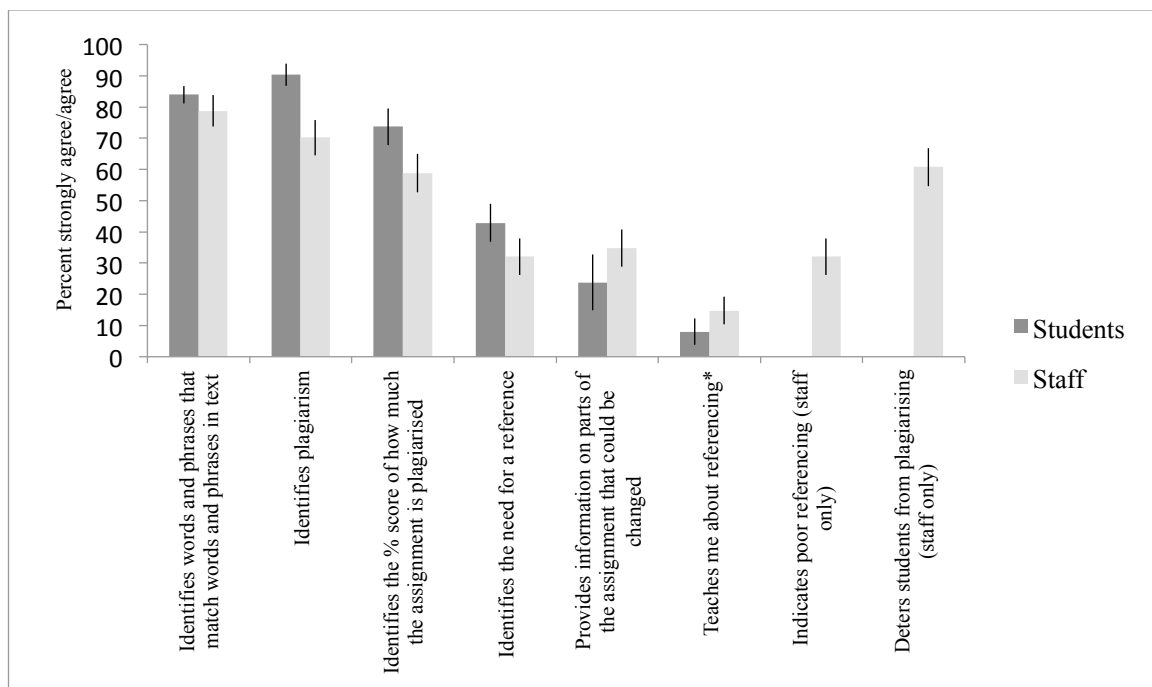

Figure 1: Students' and staff understandings of the functions of SafeAssign with 95%CI. *Staff were asked a slightly different version: Provides feedback about what the student needs to learn about referencing
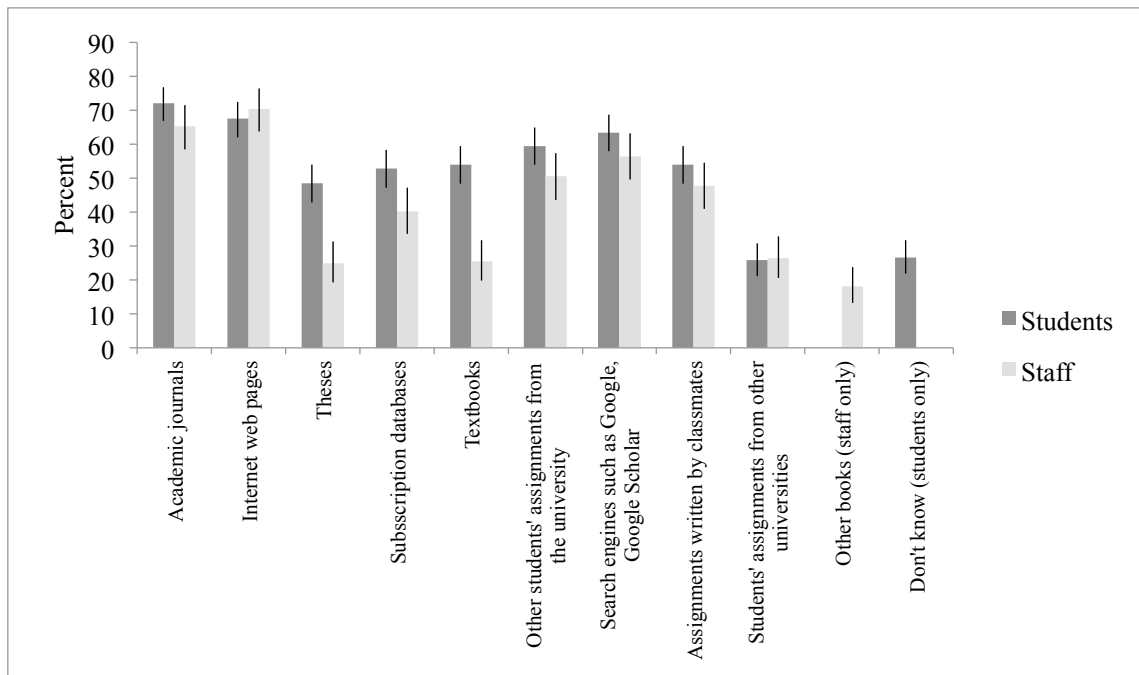
Figure 2: Students' and staff understandings of the sources available to SA to match text with 95% CI

All students and all staff were asked about the sources SA is able to access for text matching. Responses showed a lack of knowledge. The results are summarised in Figure 2 above. One of the main sources SA access are subscription data bases such as ProQuest, Academic One File or JStor, however only 53% of students and 40% of staff were aware of this. Although SA does not access theses or textbooks, approximately 50% of students and 25% of staff thought that it did. Furthermore, only 60% of students and 50% of staff were aware that SA checks against archives of previous students' assignments, and only 54% of students and 48% of staff identified that it also checks against assignments written by others in the same class. Both staff users and non-users indicated similar levels of agreement for these items.

Discussion

The results of this research show that students and staff believe that SA can do more than it does. Although SA does not identify plagiarism, 90% of the student respondents, and 70% of the staff respondents in this study thought that it did. These results indicate that both students and staff are confused about the difference between 'text matching' and plagiarism, and may be conflating them. This misconception may lead to SA being more likely to be perceived as a 'plagiarism detection' tool, thus ignoring the software's educative function. The lack of awareness of participants in this study of SA as a tool to identify areas of an assignment that could be re-worked or more effectively referenced provides further indication that the educative function of SA was not being fully utilised by students and staff. This finding is consistent with the literature that reports students in particular are unaware of the role TMS can play in academic writing (Löfström & Kupila, 2003). In order for TMS to function as an educative tool, staff must use the report when working with students to identify the problem and a possible resolution. In addition, students need to have the opportunity to resubmit the assignment to ensure the solution is effective (McKeever, 2006). In this way, staff can effectively work with students struggling with effective paraphrase practices or correct citation. However, this research suggests that staff and students focused on the percentage score produced by the software and may not understand that the SA report can be used in this way.

Despite overestimating the function of SA, neither students nor staff had an accurate understanding of the sources of text that SA can use for matching. In general, both students and staff underestimated the range of sources that SA uses for checking, and indicated their belief that it can check against sources that it cannot (for example textbooks).

One of the more concerning results of this research is that many students were unaware that SA was being used in their papers, and many students believed SA's main function was assignment submission. Although one of the reported features of TMS is its ease and convenience for the submission of work (Evans, 2006; Ledwith &

Risquez, 2008), submission of assignments is not a primary function of SA. Students in this study indicated that they should be made aware of the use of SA in their papers.

The results of this research indicate that both students and staff required more information on the functions and use of SA. In particular, they required more information on the educative function of SA as a means to reduce unintentional plagiarism. The results of the survey have been fed back to CALT, with a recommendation that the use of SA be continued, but that more resources are directed into education and support for student and staff using SA.

Acknowledgements

References

Atkinson, D., & Yeoh, S. (2008). Student and staff perceptions of the effectiveness of plagiarism detection software. *Australasian Journal of Educational Technology, 24*(2), 222-240.

Austin, P. & Hux, J. (2002). A brief note on overlapping confidence intervals. *Journal of Vascular Surgery,* 36, 1, 194-195.

Badge, J. (2010). *How effective are electronic plagiarism detection systems and does it matter how you use them? Reviewing the evidence.* Paper presented at the 4th International Plagiarism Conference "Towards an Authentic Future", Northumbria University. Retrieved from http://plagiarismadvice.com/documents/conference2010/papers/4IPC_0023_final.pdf.

Badge, J., Cann, A. J., & Scott, J. (2007). To cheat or not to cheat? A trial of the JISG plagiarism detection service with biological sciences students. *Assessment & Evaluation in Higher Education, 32*(4), 433-439.

Bond, C., Adam, L., Samaranayaka, A., Finn, K., Senior, A., Cross, D., & Deaker, L. (2013). Plagiarism and the use of SafeAssign at the University of Otago: A Report prepared for the Quality Advancement Office and the Committee for the Advancement of Learning and Teaching (CALT): University of Otago.

Evans, R. (2006). Evaluating an electronic plagiarism detection service: The importance of trust and the difficulty of proving students don't cheat. *Active Learning in Higher Education, 7*(87), 87-99.

Green, D., Lindemann, L., Marshall, K., & Wilkinson, G. (2005). Student perceptions of a trial of electronic text matching software: A preliminary investigation. *Journal of University Teaching and Learning Practice, 2*(3).

Ledwith, A., & Risquez, A. (2008). Using anti-plagiarism software to promote academic honesty in the context of peer reviewed assignments. *Studies in Higher Education, 33*(4), 371-384.

Löftström, E., & Kupila, P. (2013). The instructional challenges of student plagiarism. *Journal of Academic Ethics, 11*, 231-242.

McKeever, L. (2006). Online plagiarism detection services - saviour or scourge? *Assessment & Evaluation in Higher Education, 31*(2), 155-165.

Park, C. (2003). In other (people's) words: Plagiarism by university students - literature and lessons. *Assessment & Evaluation in Higher Education, 28*(5), 471-488.