

1-1-2014

Detecting the work of essay mills and file swapping sites: some clues they leave behind

Ann M. Rogerson
University of Wollongong, annr@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/gsbpapers>



Part of the [Business Commons](#)

Recommended Citation

Rogerson, Ann M.: Detecting the work of essay mills and file swapping sites: some clues they leave behind 2014, 1-9.
<https://ro.uow.edu.au/gsbpapers/434>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Detecting the work of essay mills and file swapping sites: some clues they leave behind

Abstract

Essay mills are commercial enterprises delivering assessable work on a fee-for-service basis. File-swapping sites encourage students to upload graded work and institutional materials to exchange for work or solutions submitted by others. The number of both types of sites continues to grow, indicating the issue is unlikely to disappear. Plagiarism preventative measures such as promoting academic integrity practices, and including recent real-world events in assessment design do not provide a disincentive to stop students purchasing or repurposing assignments due to the seemingly low risk of detection. This paper summarises the results of a content analysis study of detailed interview notes compiled while investigating students with irregularities in their assignment submissions. A pattern of clues were identified within the irregularities. Clues indicating essay mill purchases included the misrepresentation of bibliographic material where information was omitted or changed, inappropriate references out of the subject area, the style and types of references used, in addition to generalised discussion that did not answer a specific question. File swapping site usage was identified through the inclusion of inappropriate and irrelevant material, bibliographic 'mashups' where academic journal and book information was blended into a single reference, and cross matching identified by Turnitin where students relied on the same file swapping source. Once the clues were identified patterns of use in other assignment work was easier to detect. Part of the key to detection is awareness, and knowing the patterns and signs evident in non-original work can assist in identifying other submissions warranting closer inspection.

Keywords

leave, behind, essay, mills, file, swapping, work, sites, detecting, clues, they

Disciplines

Business

Publication Details

Rogerson, A. (2014). Detecting the work of essay mills and file swapping sites: some clues they leave behind. 6th International Integrity & Plagiarism Conference: Conference Proceedings (pp. 1-9). Newcastle-on-Tyne: Plagiarismadvice.org.

Detecting the work of essay mills and file swapping sites: some clues they leave behind

Ann Rogerson

University of Wollongong, Australia

annr@uow.edu.au

Keywords: plagiarism, detection, Turnitin, essay mills

Essay mills are commercial enterprises delivering assessable work on a fee-for-service basis. File-swapping sites encourage students to upload graded work and institutional materials to exchange for work or solutions submitted by others. The number of both types of sites continues to grow, indicating the issue is unlikely to disappear. Plagiarism preventative measures such as promoting academic integrity practices, and including recent real-world events in assessment design do not provide a disincentive to stop students purchasing or repurposing assignments due to the seemingly low risk of detection.

This paper summarises the results of a content analysis study of detailed interview notes compiled while investigating students with irregularities in their assignment submissions. A pattern of clues were identified within the irregularities. Clues indicating essay mill purchases included the misrepresentation of bibliographic material where information was omitted or changed, inappropriate references out of the subject area, the style and types of references used, in addition to generalised discussion that did not answer a specific question. File swapping site usage was identified through the inclusion of inappropriate and irrelevant material, bibliographic 'mashups' where academic journal and book information was blended into a single reference, and cross matching identified by Turnitin where students relied on the same file swapping source.

Once the clues were identified patterns of use in other assignment work was easier to detect. Part of the key to detection is awareness, and knowing the patterns and signs evident in non-original work can assist in identifying other submissions warranting closer inspection.

Introduction

Despite the best efforts of institutions and academics to educate students in academic skills including the appropriate acknowledgement of sources, students are still identified as having submitted plagiarised materials for assessment. Park (2003:479) outlines that there are "multiple and contingent motives for plagiarism by students" including unfamiliarity with academic acknowledgement practices, time pressures and student attitudes. One of the other motives cited by Park (2003:479) is the opportunity afforded by ease and accessibility of information on the Internet. Internet based resources have been cut and pasted into assignments for some time, but now commercially run enterprises promote services to write papers for students on a fee-for-service basis, which Page (2004) has labelled cyber-pseudepigraphy. File swapping sites based on peer to peer protocols popular in (illegally) sharing entertainment content (Cenite, Wang, Peiwen, & Chan, 2009) have evolved to 'share' educational content. The convenience of access and the availability of purchasable and shared information via Internet based sites create issues in the promotion of academic integrity in educational institutions (Sutherland-Smith, 2008). The need to appropriately acknowledge sources of information and authorship may not be as easily understood or accepted by a generation of students raised in a period where sharing, building and swapping information through websites, portals, blogs and file sharing domains have become the new norm.

Essay mills or paper mills, where students pay a fee to have assessment material prepared on their behalf are growing at an alarming rate (Bartlett, 2009). Sites are easily found through Internet searches besides being actively promoted by students to other students. These sites facilitate

contract cheating (Lancaster & Clarke, 2007) where students pay a contractor to write assessable work for them. No longer a hidden industry, essay mills are going as far as promoting their services by targeting students attached to institutional Facebook sites at the start of academic sessions. Sites are not confined to particular countries or continents, with some essay mills operating on a global scale offering multiple language options, telephone and chat support, enticing students by offering to ease their study burden and help them achieve better grades. This can be attractive to students who wish to avoid the process of writing and meet assessment due dates (Evans, 2006). Some essay mills go so far as to indicate to students that it is acceptable to purchase essays, just as you would hire a tradesperson to complete work on your behalf. Worse still are the sites promoted as essay mills but are actually scams that do not provide essays at all, taking money in advance without providing an assignment to the purchaser and/or provide the same essay to multiple students.

File swapping or trading sites like www.thinkswap.com.au, and <http://wenku.baidu.com> encourage students to upload graded work, test questions and answers, copyrighted and even confidential materials to exchange for submissions uploaded by others. Usage of these types of sites is often masked as uploaded material is held in file formats not currently examined by Turnitin © such as .gif or .jpg. File-swapping sites take the notion of information sharing into new territory, particularly from an academic integrity perspective. Creative-commons licences allow sharing with author acknowledgement, whereas file-swapping sites seek to share information without acknowledgement for personal gain in the form of passing assessment tasks. Some file swapping sites operate by a credit or barter system (a student uploads material and is then entitled to download material) while others operate with a membership fee. Students can download parts of assignments and repurpose them, altering words and phrases to avoid writing part or all of a paper. These websites exist in many countries, touting the student testimonials about the latest test and assignment papers uploaded. While a student may own the copyright in their own assignment, they do not own the copyright in materials prepared by the institution, and unfortunately as the home page of websites such as www.thinkswap.com.au demonstrate university owned materials not only appear but are also promoted.

The number of sites promoting fee-for-service assignment writing and file swapping continues to grow, indicating that the issue is not going to disappear. Plagiarism preventative measures including the promotion of academic integrity practices (Bretag, 2013; McCarthy & Rogerson, 2009), and using recent real-world events as a basis for written assessment design only go so far. These approaches do not provide a disincentive to stop the practice of purchasing or repurposing assignments particularly where the motivation of the student is tied to gaining the degree without little concern for the associated knowledge. A lack of institutional support can also prove a disincentive for academics to pursue cases due to appeals by international students being upheld (Bretag, 2007). Experience, plus an exposure to a wide range of material assists in the identification of material (Bretag & Mahmud, 2009). However, there are some clues that can assist in identifying work where students should be interviewed about the source or questionable authorship of their assessable submissions.

The problem

Suspicious arose about the originality of some assessment submissions over 2 classes (total 102 students) forming part of post-graduate degree programs for international students. While grading the first assignments a large number of anomalies were detected in the standards of referencing and English used, which resulted in a higher number of fail results compared to other instances of the subject over the previous five years. Short discussions (3-5 minutes) were conducted with students where inconsistencies were highlighted, and learning support referrals initiated. During two of these interviews students admitted to not writing their own assignments. Both were identified through a comparison of the submissions from their previous attempt at the subject and the apparent vast improvement in English language expression in the new submission. The students were duly

reported, investigated and penalised under University of Wollongong academic integrity and plagiarism policy processes <http://www.uow.edu.au/about/policy/UOW058648.html>.

After the poorer than expected results in the first assessment task, a large amount of lecture and tutorial time was devoted to preparing students for the second assignment. This included outlining how to address errors highlighted in the first assignments, particularly in the identification, use and recording of quality references and sources. Students were also offered additional support in consultation times, and via email, although few availed themselves of the opportunities. Despite the increased intervention proposals, a review of the Turnitin originality reports indicated that a large range of discrepancies still existed in the second assessment task submissions. On closer examination the inconsistencies could be categorised into distinct groups. Some were the result of varying standards of English language grammar and phrasing indicating a vast improvement between the first and second assignments despite the apparent lack of use of learning support or consultations, yet were so generic the specific assignment question had not been addressed. Matches between students were also found on Turnitin originality reports, in addition to large matches to Internet based materials. Others were the result of reference details that were in conflict with the information discussed in the assignment, or clearly incorrect due to the information presented at the end of the assignment.

The range of irregularities was identified early on in the grading process, and to confirm the observations the assignments were double marked by the lecturer to re-examine the papers given the emerging patterns. Due to the volume and complexity of issues, an interview template was prepared to note concerns relating to each assignment such as the “telling cues as obvious shifts in diction and surprising levels of sophistication” (Abasi, Akbari, & Graves, 2006:104). Other issues documented were the misuse or misrepresentation of references and failure to use specific examples to support arguments as required in the assignment question. During the second marking the patterns were confirmed, and further examples identified. As a result, and in accordance with university policy when irregularities are identified within an assignment submission, students had to be interviewed to discuss any concerns. The student interviews were to determine if there was a reasonable explanation for variances, or, if further action was required under the University of Wollongong academic integrity and plagiarism policy processes. The grading notes made on the interview forms allowed for specific follow up with each student, and an opportunity to record their responses for consideration and review.

Only assignments that were considered to be actually written by the student were handed back, and all other students advised that due to concerns an individual meeting with the lecturer was required before their assignment would be returned. This resulted in total of N=70 x15 minute interviews conducted with students where the anomalies were highlighted and a discussion ensued. Students were asked about any learning development activities they had undertaken since the first assignment, their process of writing and in some cases asked to demonstrate how they searched for references. In addition, some background questions about pathways to study were asked. Basic questions about the theory related to the assignment were asked to assess whether key concepts had been understood, an approach to investigate cases where there is the suspicion of plagiarism particularly amongst international students (Sowden, 2005). Further notes were made during the student discussions, so that determinations about grades or consequences could be made after the interviews. These interviews generated a rich body of data that was held securely until the students involved completed their course of study to support further enquiries resulted from the higher than normal failure rate.

Methodology

The opportunity to examine the interview data through content analysis was considered important to analyse and report the patterns identified to add to the body of knowledge about academic integrity issues. Ethics approval was sought and granted for a retrospective examination of the data once it

was de-identified. First, the notes related to each assignment interview were matched to a downloaded copy (word.docx file) of the relevant assignment captured in Turnitin. After matching the interview notes to the assignment, each set of student interview papers and associated assignment were allocated a case number as a reference point for analysis and discussion. The case number was then applied to the related set of student data in a spreadsheet (grades and assignment feedback). After the case numbers were applied any student identifiers (such as name/student number/class) were removed by erasing the student information on the paper copies, and deleting any student identifiers in the spreadsheet and word.docx file downloads, and document properties. Case by case the interview data was then transposed into the spreadsheet, with comments noted under the pattern categories identified so key data elements could be analysed.

Findings

The summary of issues relating to the second assignment submissions is presented in Table 1.

Table 1 Summary Data of Assignment Issues

Number of Students	High similarity in Turnitin	Differences in English expression	Did not answer the question	Reference 'mashups'	References misrepresented in list	References inappropriate or irrelevant
N=70	8	70	48	3	16	21

The key issue that flagged assignments for closer review was the differences in English expression. The next biggest issue was the fact that the report submitted did not address the assignment question (n=48). The next sections discuss the interview data in relation to the categories of issues.

Issues identified in Turnitin

Eight (8) cases had high similarity percentages in Turnitin. The majority of these cases related to the cutting and pasting of information off the Internet, resulting in discussions about appropriate academic integrity practices. One case matched to a student submission for the first assignment, and the student thought that papers were only checked against the current assessment. Two of the eight cases had a 16% match to each other in the introduction section [cases 1 and 37]. The students were interviewed separately to determine who copied from who, however the answer was different to what was expected. The first student admitted to taking the information off a Chinese file swapping site. When the second student was questioned about the cross match, the student was confused why it would match to another student. The source of the match was eventually suggested and the student admitted to using the same file swapping site, however both students refused to disclose which site they used, and appeared to have genuinely worked independently.

Papers reflecting high quality English expression

The majority of papers in the study had either a high level of English expression (good use of grammar, spelling and phrasing) or variations in language between good and poor English expression. When students were interviewed about their process of writing, a great difference was identified between the use and understanding of spoken English, and written English. There were some students who did not appear to understand English language very well, despite requiring an IELTS academic level of 6.5 to undertake the degree. One student [case 17] arranged for a friend to accompany her as I was advised "she did not speak English well". The friend ended up translating my questions into Mandarin Chinese, and translating the responses back into English. Despite this, the student insisted that they had written the assignment themselves in perfect English.

Assignments with writing variations between good and poor English comprised areas matched off Internet resources, or appear to have been sourced from textbooks currently not examined by Turnitin. Clues in this area relate to older dated citations (pre 1990 and as far back as 1871), and

mixes in presentations of intext citations, for example where a correct intext citation is presented, yet the reference did not appear in the reference list. A few of these could also have been sourced from file-swapping sites, although the evidence here is inconclusive.

Interviews revealed a wide range of English language capabilities. Some students were quite articulate, but in one example [case 22] the student had completed an Australian undergraduate degree in pure mathematics but had no experience in writing essays. This student was interested in learning what to do and this interview turned into individual tutoring sessions to demonstrate how to improve their future assignments. There were other interviews of a similar nature which accounted for one in every five interviews.

Reviewing the assignments where the student appeared to have a lower command of English language, and the entire assignment demonstrated a high level of English expression, the following pattern was identified. Firstly, the papers were very generic, and did not answer the assignment question set in a business context (n=48). Assessment tasks were designed to minimise the potential for plagiarism by referring to recent events, and seeking a comparative analysis between countries. Some related theory was described at a high level, but the required supporting examples were not provided, and in some cases the countries in question were not even mentioned. Secondly, the papers showed a Turnitin similarity score between 0% and 3% overall similarity, despite the fact the bibliography was included in the submission. Other students in the class who had used quality references including the main theorists papers had overall similarity matches ranging between 18% and 26%, which reduced to an overall score between 3% and 5% (individual matches 1% or less) once the bibliographic filter was switched on. The lower originality scores, coupled with a consistent high level of English expression seemed to indicate that the essays had been either purchased, or written by a third party. Students were then asked to demonstrate how they searched for their references. This area of the interview shed some light on the actions of the essay mills and ghost writers.

Misrepresentations in referencing

The reference lists in the n=48 generic responses appeared to be legitimate, and both the reference list and intext citations were consistent in format and presentation in Harvard format. The question then arose why the overall matches were so low. Misrepresentations in referencing comprised a few forms. Some could be the result of poor scholarship, but the interviews also identified that some misrepresentations are deliberate attempts to take short cuts, in order to appear to meet grading criteria, while limiting similarity matches in Turnitin. Where referencing issues were identified, students were asked to locate some of the references they had used, and when they could not locate them, or they did not appear as listed, students were forthcoming about the actions they had taken. When questioned about why the information discussed did not match the reference source, comments such as "I just put one in" [Case 44], "I search and put in first reference and not check it" [sic Case 49] portraying the issue as poor scholarship. This may or may not be the case, or could be the 'story' that students have used in the past to cover their actions.

Other students found the reference, however there were author names missing, and finer details such as issue numbers, volume numbers, and page numbers were different to the original papers as reflected in the journal databases (see Figures 1 and 2).

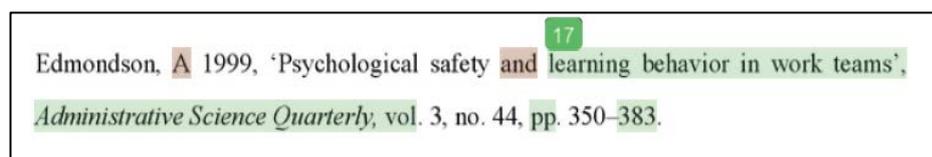


Figure 1
Misrepresented
reference in student
submission
(Turnitin highlights)



Figure 2 Actual reference details (volume has been switched to issue number and an incorrect volume number shown) <http://www.jstor.org/stable/2666999>.

The students seemed genuinely shocked, but had no explanation for the differences. It seems that some ghost writers, or mills may alter aspects of legitimate references, so that the

reference appears correct (see Figure 1), but in actual fact is misrepresented (see Figure 2). These misrepresentations are not always obvious in Turnitin, and cannot be found by visual examination. People reading papers do not have the time to individually check references to detect this, and for this

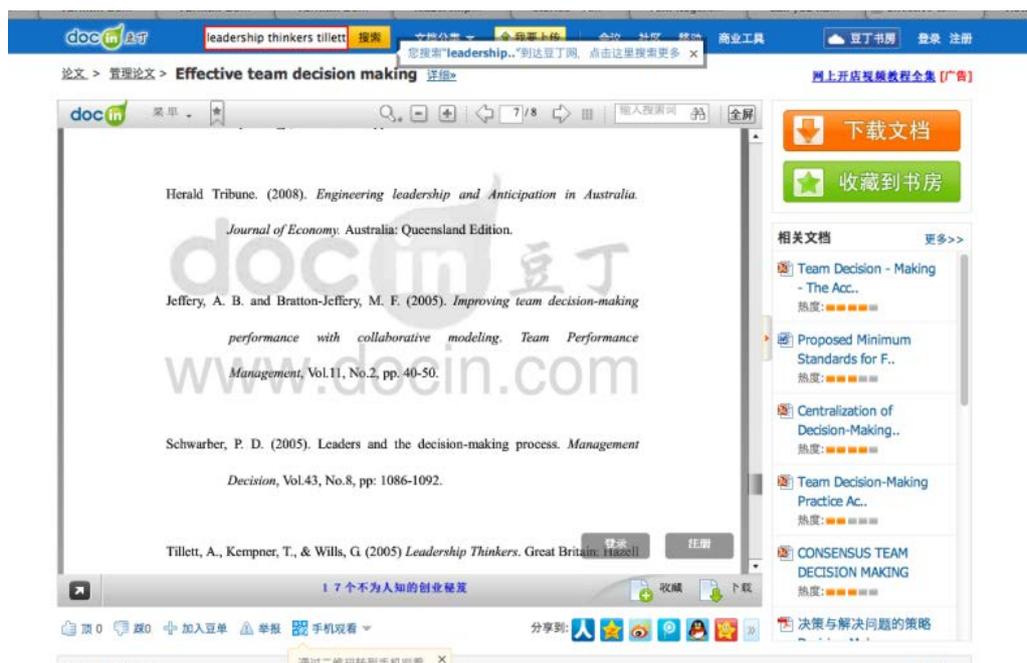
Schwarber, P.D., 2005, Leaders and the decision-making process. *Management Decision*, 43(8), pp. 1086-1092.

Sukthakar, G. & Sycara, K., 2010, Analyzing Team Decision-making in Tactical Scenarios. *The computer Journal*, 53(5), pp. 503-512.

Tillett, A. & Kempner, T. & Wills, G., 2005, *Leadership Thinkers*. Great Britain: Hazell Watson and Viney.

Tribune, H., 2008, Engineering leadership and Anticipation in Australia. *Journal of Economy. Australia: Queensland Edition*.

Figure 4 File-swapping website match



The student was asked to find the reference during the interview, and tried searching through the University of Wollongong library databases, and became very frustrated when she could not locate the reference. When the student was shown the link to the file-swapping site, the argument changed. They stated that there was a Mr Tribune and it was a real reference. The student did not want to believe the lecturer that the Herald Tribune was a newspaper. There were two other students who had references that appeared with a similar style of 'bibliographic mashup' [cases 13, and 62], however searches via Google, Firefox and Internet Explorer could not detect the source.

Throughout the course of the interviews with students a pattern of clues emerged which facilitated easier detection of material that had not been authored by the student, had been written by someone else, or repurposed by re-engineering prior student submissions. Any issues that required further action under the UOW Academic Integrity and Plagiarism policy, took place under due process with appropriate penalties applied. Others where the interviews could not confirm or deny the use of unauthorised materials were marked at face value – subsequently many of these students failed the subject due to their performance in the final exam demonstrating that the students had little understanding of the content despite 11 weeks of lectures and tutorials.

Discussion

Griffiths and Brophy (2005) noted that most students use Google as the search engine of choice and that their use of academic resources was low. Students traded the quality of results based on how much time it took to find them. While their research was conducted in 2005, the discussions with the majority of students involved in the study supported this view. Due to the advertising of essay mill type sites, and the high positioning on search terms students are more likely to come across an essay mill when searching for 'assignment information' if using an engine like Google. Peer information is likely to support the use or access of a file-swapping site, although the quality of the results is likely to be lower. This view is supported into influences on academic integrity whose regression analysis demonstrated that peer behaviour has the greatest impact on influencing students (McCabe, Butterfield, & Trevino, 2006). It became clearly apparent during the discussions that the students were totally focussed on achieving an overall similarity score of zero in Turnitin. Despite preparatory sessions on the use of Turnitin and how originality reports worked, interviews revealed still believed that a zero similarity score would ensure their assignment would not be subject to detailed examination.

There was a large investment of time in the double marking, preliminary investigations and interviews, however, once issues were identified due process had to be followed. The time and effort required to pursue cases and lack of institutional support discourages many academics from following up on cases where discrepancies are evident (Hughes & McCabe, 2006). It should be noted that these classes have been identified as extreme cases relating to a particular cohort that overall had higher than normal failure rates. The prime underlying reason for this situation is considered to be the poor level of English language capabilities evident in this particular group. Subsequent classes demonstrated that these particular classes had a lower level of English language capabilities when compared to other cohorts, which could have been a prime influencing factor on the volume of cases with questionable authorship. However the opportunity afforded by the investigations and interviews have led to a greater awareness of the existence of clues and easier identification of material not authored directly by the student.

Conclusion

A combination of certain aspects of similarity reports, and an awareness of the patterns indicating certain discrepancies can assist in the detection of essay mill usage or file-swapping sites. Turnitin matches between students in a class, or previous classes may not be the result of person to person sharing of materials, instead it can indicate the use of the same file swapping site. Bibliographic mashups may evade Turnitin, but it is falsification and should be discussed with the student. Zero

matches in Turnitin may be of greater concern than previously thought. Poor levels of English skills places students at risk (Park, 2003). Firstly the students do not understand requirements or do not have a level of English where they are capable of completing the required assessment tasks, therefore in order to pass and comply with family obligations, they seek a way of surviving, with little regard of the rules or ethics relevant to purchasing or copying (Costigan, Hua, & Su, 2010). Secondly, when English levels are low students purchasing essays or repurposing materials from file-swapping sites are not in a position to adequately assess the appropriateness, correctness or validity of materials purchased or used. Promoting the idea to students that bibliographic matches to quality academic papers means a higher similarity score (which can be checked through the filters) and is actually better than a zero may be a positive step. Following the process means that students are fully aware of the consequences, and instead of peer endorsement of file swapping and essay purchasing, peer confirmation that issues will be identified and pursued and become a deterrent in themselves. While students may see this as an effective way to achieve a passing grade for a written assessment task, it defeats the true purpose of studying, which is at post graduate level to demonstrate the acquisition and application of knowledge.

References:

- Abasi, A. R., Akbari, N., & Graves, B. (2006). Discourse appropriation, construction of identities, and the complex issue of plagiarism: ESL students writing in graduate school. *Journal of Second Language Writing, 15*(2), 102-117. doi: 10.1016/j.jslw.2006.05.001
- Bartlett, T. (2009). Cheating Goes Global as Essay Mills Multiply. *The Chronicle of Higher Education, 55*(28), A.1.
- Bretag, T. (2007). The Emperor's New Clothes: Yes, There Is a Link Between English Language Competence and Academic Standards. *People and Place, 15*(1), 13-21.
- Bretag, T. (2013). Challenges in addressing plagiarism in education. *PLoS medicine, 10*(12), e1001574. doi: 10.1371/journal.pmed.1001574
- Bretag, T., & Mahmud, S. (2009). A model for determining student plagiarism: Electronic detection and academic judgement: University of Wollongong Research Online
- Cenite, M., Wang, M. W., Peiwen, C., & Chan, G. S. (2009). More Than Just Free Content: Motivations of Peer-to-Peer File Sharers. *Journal of Communication Inquiry, 33*(3), 206-221. doi: 10.1177/0196859909333697
- Costigan, C., Hua, J., & Su, T. (2010). Living up to Expectations: The Strengths and Challenges Experienced by Chinese Canadian Students. *Canadian Journal of School Psychology, 25*(3), 223-245. doi: 10.1177/0829573510368941
- 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*(1), 87.
- Griffiths, J. R., & Brophy, P. (2005). Student Searching Behavior and the Web: Use of Academic Resources and Google. *Library Trends, 53*(4), 539-554.
- Hughes, J., & McCabe, D. (2006). Understanding Academic Misconduct. *The Canadian Journal of Higher Education, 36*(1), 49.
- Lancaster, T., & Clarke, R. (2007). *Assessing Contract Cheating through Auction Sites—A Computing Perspective*. Paper presented at the Proceedings of 8th Annual Conference for Information and Computer Sciences.
- McCabe, D., Butterfield, K., & Trevino, L. (2006). Academic Dishonesty in Graduate Business Programs: Prevalence, Causes, and Proposed Action. *Academy of Management Learning & Education, 5*(3), 294.
- McCarthy, G., & Rogerson, A. M. (2009). Links are not enough: using originality reports to improve academic standards, compliance and learning outcomes among postgraduate students. *International Journal Educational Integrity, 5*(2), 47-57.
- Page, J. S. (2004). Cyber-pseudepigraphy: A New Challenge for Higher Education Policy and Management. *Journal of Higher Education Policy and Management, 26*(3), 429-433. doi: 10.1080/1360080042000290267
- Park, C. (2003). In other (people's) words: Plagiarism by university students--Literature and lessons. *Assessment and Evaluation in Higher Education, 28*(5), 471-488. doi: 10.1080/02602930301677

Detecting the work of essay mills and file swapping sites: some clues they leave behind

- Sowden, C. (2005). Plagiarism and the culture of multilingual students in higher education abroad. *ELT Journal*, 59(3), 226-233. doi: 10.1093/elt/cci042
- Sutherland-Smith, W. (2008). *Plagiarism, the Internet, and student learning; improving academic integrity* (Vol. 23). Portland: Book News, Inc.