A model for determining student plagiarism: Electronic detection and academic judgement

Tracey Bretag  
University of South Australia, tracey.bretag@unisa.edu.au

Saadia Mahmud  
University of South Australia, saadia.mahmud@unisa.edu.au

Follow this and additional works at: https://ro.uow.edu.au/jutlp

Recommended Citation
Available at:https://ro.uow.edu.au/jutlp/vol6/iss1/6

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
A model for determining student plagiarism: Electronic detection and academic judgement

Abstract
This paper provides insights based on the authors’ own practice as university instructors, researchers and arbitrators of student plagiarism. Recognising the difficulty in defining plagiarism while still acknowledging the practical importance of doing so, the authors find the common element between the various types of plagiarism to be the lack of appropriate attribution to the original source. The use of electronic text-matching software to detect different types of plagiarism is explored, and a model presented for identifying potential plagiarism in students’ work. The authors conclude that despite its shortcomings, electronic detection in combination with manual analysis, nuanced academic judgement and clear processes provide the means to determine if plagiarism has occurred.

Cover Page Footnote
Saadia Mahmud, previously known as Saadia Carapiet.
A model for determining student plagiarism: 
Electronic detection and academic judgement.

Tracey Bretag and Saadia Mahmud

School of Management, University of South Australia
tracey.bretag@unisa.edu.au; saadia.mahmud@unisa.edu.au

Abstract
This paper provides insights based on the authors’ own practice as university instructors, researchers and arbitrators of student plagiarism. Recognising the difficulty in defining plagiarism while still acknowledging the practical importance of doing so, the authors find the common element between the various types of plagiarism to be the lack of appropriate attribution to the original source. The use of electronic text-matching software to detect different types of plagiarism is explored, and a model presented for identifying potential plagiarism in students’ work. The authors conclude that despite its shortcomings, electronic detection in combination with manual analysis, nuanced academic judgement and clear processes provide the means to determine if plagiarism has occurred.

\^Saadia Mahmud, previously known as Saadia Carapiet.
Introduction

The public perception that plagiarism is on the rise has resulted in a proliferation of discussion and research on student plagiarism (Angélil-Carter, 2000; Bretag, 2005, 2007; Carroll, 2003a; Devlin, 2003; McCabe, Butterfield & Trevino, 2006; Zobel & Hamilton, 2002). Internationally recognised author and practitioner, Jude Carroll, has publicly stated that plagiarism exists in all educational institutions, and those that do not admit to the problem are not being honest (Carroll, 2003b; see also Piety, 2002). What seems to be emerging from commentary and research is the understanding that student plagiarism is here to stay.

As Carroll has noted, “definitions matter and agreeing on a good one is harder than you think” (2003a, p. 12). However, despite the difficulties of defining plagiarism, there is a significant body of literature providing detailed advice on how to respond to student plagiarism, in terms of preventative, punitive and educative measures. Researchers have provided practical recommendations at the institutional level, and strategies at the individual educator level to deal with plagiarism, and many of these suggestions overlap. One of the strategies to detect and deter student plagiarism has been the development of electronic software tools. When these tools first became available many educators believed that they would provide an easily implemented solution, which would cut down on the hours of tedious manual detection. However, within a short period of time, it became clear that electronic detection software “is not a magic bullet” (Carroll, 2003a) and that it is just one tool among many to be used within an educative framework. Using our own research and practice as a basis, we argue that despite its shortcomings, electronic detection in combination with manual analysis, nuanced academic judgement and clear processes provide the means to determine if plagiarism has occurred. This paper extends this understanding and provides a simple model to assist in the identification of student plagiarism.

In the initial sections of this paper, the authors review a wide variety of definitions and types of ‘plagiarisms’ identified in the literature. While acknowledging the complex nature of plagiarism, we conclude that the common element in the various ‘plagiarisms’ is the lack of appropriate attribution to the original source. The paper then explores the various responses to student plagiarism recommended in the literature, the use of electronic tools to detect plagiarism, and the way that such tools can be used within an educative framework to enhance student learning.

Defining plagiarism

We agree with Carroll’s (2003a) observation that agreeing to a definition of plagiarism is difficult, but nonetheless necessary if students are to avoid breaching academic integrity. One of the impediments to providing a conclusive definition of plagiarism is the complexity of factors influencing its occurrence, which then creates an understandable reluctance on the part of educators and researchers to provide a simple, once-and-for-all definition. Bretag (2005) refers to plagiarism as an issue potentially relating to various factors, including linguistic competence, academic literacy, culture, racism, academic integrity, media scandal and institutional governance. Further complicating the issue is the fact that plagiarism is often conflated with cheating and academic misconduct generally.

Derived from the Latin word meaning ‘kidnap’ or ‘plunder’, the concept of plagiarism has criminal connotations (Angélil-Carter, 2000, p. 16). For this reason, Angélil-Carter suggests that the whole concept of plagiarism needs to undergo “substantial transformation” (2000, p. 17), as there are “varying levels of plagiarism, and varying reasons for plagiarism, most of which are not ‘cheating’ or intentionally fraudulent” (p. 116). According to James, McInnes and Devlin (2002), “plagiarism varies in both intent and extent, ranging from deliberate fraud, to negligent or accidental failure to acknowledge sources of paraphrased material and misunderstandings about the conventions of authorship” (p. 5). Harris (2001) maintains that it is difficult to provide a single definition of plagiarism...
and provides seven sample definitions taking into account variations in the “the use of data, the permissibility of collaboration, requirements for citation, and even what constitutes common knowledge” (Harris, 2004).

The Council of Writing Program Administrators, based in the United States, refers to plagiarism as “a multifaceted and ethically complex problem”. However, as an aid to educators, administrators and students, they provide a pragmatic definition as follows: “…plagiarism occurs when a writer deliberately uses someone else’s language, ideas, or other original (not common-knowledge) material without acknowledging its source” (Defining and avoiding plagiarism: The WPA Statement on Best Practices, 2008). Most university websites, while also acknowledging the complexity of the issue, attempt to provide a practical definition of plagiarism along similar lines (for example, the University of South Australia (Plagiarism – teaching strategies, 2008), Oxford Brookes University (Plagiarism, 2008b), Clemson University (Plagiarism, 2008a)).

Types of plagiarism

Given the lack of consensus regarding a single definition of plagiarism, it is not surprising that a wide variety of (often overlapping) plagiarism types have been identified. Table 1 provides an overview of some of the types of plagiarism mentioned by authors working in the broad field of academic literacy and integrity, with the common element identified being the lack of appropriate attribution to the original source.

Table 1: Types of plagiarism

<table>
<thead>
<tr>
<th>Author</th>
<th>Type of plagiarism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Martin (1994)</td>
<td>Word-for-word&lt;br&gt;Paraphrasing&lt;br&gt;Secondary sources&lt;br&gt;Form of a source (structure of an argument)&lt;br&gt;Ideas&lt;br&gt;Authorship</td>
</tr>
<tr>
<td>Howard (1995)</td>
<td>Cheating (borrowing, purchasing or otherwise obtaining another’s work)&lt;br&gt;Non-attribution of sources&lt;br&gt;‘Patch-writing’</td>
</tr>
<tr>
<td>Klausman (1999)</td>
<td>Direct&lt;br&gt;Paraphrasing&lt;br&gt;‘Patchwork’</td>
</tr>
<tr>
<td>Evans (2000)</td>
<td>Quotation&lt;br&gt;Paraphrasing&lt;br&gt;Auto-plagiarism (failure to cite oneself)&lt;br&gt;Self plagiarism (submitting the same document several times)&lt;br&gt;Cryptomnesia (where hidden memory plays a key role in lack of citation).</td>
</tr>
<tr>
<td>Harris (2001)</td>
<td>Downloading free papers from the Internet&lt;br&gt;Buying a paper from a commercial paper mill&lt;br&gt;Copying an article from the Internet or online database&lt;br&gt;Translating foreign article into English/another language&lt;br&gt;Copying a paper from another student&lt;br&gt;Cutting and pasting from several sources&lt;br&gt;Quoting less than all the words copied&lt;br&gt;Changing some words but copying whole phrases&lt;br&gt;Paraphrasing without attribution&lt;br&gt;Summarising without attribution&lt;br&gt;Faking citations</td>
</tr>
</tbody>
</table>
Despite the wide variety in the types of plagiarism identified in Table 1, it is important not to equate plagiarism with all forms of cheating and academic misconduct; nor is all plagiarism necessarily word-for-word copying. What unites the different types of plagiarism is a lack of appropriate attribution to the original source.

**Responses to plagiarism**

Extensive and explicit information relating to student plagiarism and academic misconduct in assessment has a prominent place on university websites; see for example, the University of Adelaide (Academic integrity policy principles, 2008), Murdoch University (Dishonesty in assessment, 2008) and the University of Western Australia (Academic dishonesty, 2009). Rather than intimating that plagiarism can be prevented, Carroll (2003a) maintains that institutions need to commit to the three Ds: deterring, detecting and dealing with it fairly. Carroll (2003a, p. 19) provides a useful framework for determining penalties for plagiarism, with four criteria, given in descending order of priority: extent of the plagiarism, the student's year level, the student’s knowledge of the institution’s academic conventions and regulations, and the rules of the specific discipline.

The Centre for the Study of Higher Education (James, McInnes & Devlin, 2002, p. 39) presents three aspects of plagiarism that first need to be considered by academics and administrators pursuing potential academic misconduct. The first is the student’s “intent to cheat”, with “deliberately presenting the work of others as one’s own” placed at the extreme, punishable end of a continuum. The second aspect is “the extent of plagiarism” with “downloaded essay handed in as own paraphrasing” again representing the extreme end of the continuum. The third consideration is the “possible responses to plagiarism” that involve the first two aspects, and take either the form of educative or punitive strategies (James, McInnes & Devlin, 2002, p. 39). Devlin (2002) further provides a set of 36 strategies to minimise plagiarism on The Centre for the Study of Higher Education (CSHE) website, which includes advice regarding assessment development, timing and feedback, academic skills education, student honesty, academic vigilance, student collaboration, detection of plagiarism and penalties.
Using electronic text-matching software to detect plagiarism

Manual detection of plagiarism can be difficult and time consuming and many commentators maintain that the lengthy process of detecting plagiarism is one of the reasons some academics are reluctant to pursue potential cases (Carroll, 2003a; Devlin, 2003; Duggan, 2003; McCabe, 2005; Zobel & Hamilton, 2002). Martin (1994) cites Bjaaland and Lederman (1973), who recommend that teachers/markers read essays four times as part of the process of detecting plagiarism. At a more practical level, Harris (2004) offers teachers a number of strategies for detection of plagiarism in research papers such as mixed citation styles, lack of references or quotations, unusual formatting, being off topic, signs of datedness, anachronisms, anomalies of diction and style and clear indicators of plagiarism which he calls “blunders of the clueless … since it includes obvious indicators of copying”. Importantly, these strategies for detecting plagiarism are merely the starting point and require further investigation to find the original source.

A current review of electronic plagiarism detection (Scaife, 2007) compared eleven products on the market and classified these into four categories depending on the area of plagiarism addressed. The four categories were:

1. Only perform checks against Internet based material, typically using one of the major Internet search engines
2. Check for similarity within a batch of documents
3. Educational suite of products offering complete student submission/plagiarism detection solutions

The Scaife report found that almost half the products (five of the eleven) checked text against Internet based material only, while only two of the products (Turnitin and Urkund) offered an educational suite of products that checked against a range of sources, including the Internet, previously submitted papers, and journals. According to the Scaife report, Turnitin was the top scoring product with 10 million users in over 80 countries. Second was Urkund, established in autumn 2000, and used by several hundred schools and departments in Europe.

The existing electronic plagiarism detection services focus on ‘text-matching’ of the paper under review with other material found on the Internet, previous papers submitted and journals. A limitation of the services is that a vast amount of material which could be plagiarised is paper-based, such as published books, paper journals and conference papers. The exclusion of non-electronic material clearly limits the ability of the software to comprehensively detect plagiarism. Of the wide variety of plagiarisms identified in Table 1, the text-matching facility in electronic plagiarism detection software is only suited to detect ‘word-for-word’ or ‘direct’ plagiarism and then only in electronic form. The more subtle forms of plagiarism, plus all types of plagiarism from paper-based sources, are not able to be detected at present.

Educators and researchers working in the field of academic integrity agree that electronic detection is not the solution to eliminating plagiarism. As early as 2001, Carroll and Appleton argued that “electronic detection can only be an adjunct to the normal exercise of academic judgement” (2001, p. 25). Purdy (2005) insists that “As with any technology, plagiarism detection technology requires human application and interpretation” (p. 286). Barrett and Malcolm (2006, p. 41) concur that the software indicates possible plagiarism rather than providing complete certainty.
Determining student plagiarism using electronic detection

In our own practice and research, we have found that Turnitin ‘Originality Reports’ provide an excellent starting point, but what ultimately leads to determinations of plagiarism is considerable manual analysis and subjective judgement. The overall similarity index provided by Turnitin (the cumulative percentage of all the different sources which have been matched to the text under investigation) can potentially be deceptive and a high percentage of text-match is not necessarily an indicator of any form of plagiarism. For example, in one case from our own research (Bretag & Carapiet, 2007), there was an overall similarity index of 49%, but this text match was comprised of 39 separate items, the first six of which accounted for a 16% text-match, and the rest of the 33 items were just 1% matches each. The highest match was only 4%. In this case, after manual analysis, it was determined that no plagiarism had occurred. In another example, the overall similarity index was 56%, but a manual check of the sources showed that the author had appropriately cited sources in every instance, and therefore no plagiarism was determined to have occurred.

In addition to deceptively high text-matches, ‘anomalies’ can occur; for example, when a student has previously submitted their own work to Turnitin to safeguard against inadvertent plagiarism. In this case, when the same assignment is input by the instructor to Turnitin, a text-match of 100% is immediately shown and must be checked and disregarded by the instructor. Careful analysis based on pre-determined criteria is also important. For example, the instructor needs to determine which elements of a paper can reasonably be replicated without a charge of plagiarism (e.g. cover sheet, anti-plagiarism statement, topic, reference list). As Bretag (2008a) has argued previously, assessing potential plagiarism also requires nuanced academic judgement based on a range of subjective criteria, including but limited to the context (both academic and personal) within which the text has been written. See Figure 1 which provides a diagrammatic illustration of the manual process we use in conjunction with an electronic detection software program (in this case, Turnitin), to detect potential plagiarism in students’ submissions for assessment.
Figure 1: Determining student plagiarism using Turnitin
It is not only instructors and markers who need to understand the process of detecting plagiarism using electronic tools such as Turnitin. Importantly, the process described in Figure 1 needs to be shared and discussed with students prior to the submission of assignments, in the same way that marking criteria is explained and provided as a learning opportunity. As indicated by Figure 1, where the Overall Similarity Index is 50% or greater, we recommend to students that there is insufficient original material irrespective of references to sources, and that the paper needs to be redrafted so that large sections of cut and paste quotations are summarised or paraphrased, plus given appropriate citations. Presuming that the student has now submitted a paper with at least 50% of their own original work, the next stage of analysis uses criteria from our own research (Bretag & Carapiet, 2007), which applies the 10% copying guideline provided by the Copyright Act. We further advise students that quotations of up 10% are acceptable, as long as they are appropriately cited.

The plagiarism identification process detailed in Figure 1 cannot stand alone and needs to be integrated within an educative process (Carroll, 2003a; Devlin, 2003; Zobel & Hamilton, 2002). In addition to receiving clear, consistent and ongoing instruction relating to academic conventions, from note-taking to developing an academic argument using sources, students should be encouraged to use the capabilities of electronic detection software to improve their work before submitting it for evaluation.

In the course ‘Managing Communication in Business’ at the University of South Australia, coordinated by Bretag, students are required to use Turnitin as a means of checking their own drafts prior to submission. The following information is provided in the Course Information Booklet:

In this course, students are required to submit their assignment on two separate occasions to Turnitin prior to submitting the assignment for marking. The first submission will be a draft and the second submission will be your final revised assignment (the same one that you submit for marking and feedback). The purpose of submitting your work to Turnitin is to allow you to redraft your work, and prevent inadvertent plagiarism.

**Tips for using Turnitin**

Turnitin will create an ‘Originality Report’, which will show you an overall ‘Similarity Index’. This is the percentage of your essay that matches other electronic sources available on the Internet, databases and other students’ submissions. If the overall text-match is less than 10% from any one source, you do not need to revise the essay, but you should still check that each text-match is shown appropriately with “quotation marks” and with the full reference (both in-text and in the reference list).

If the overall text-match is greater than 10% from any one source, you will need to go through the essay carefully, and reduce the text-match by paraphrasing and summarising. Remember to still provide the in-text reference wherever you have used ideas from another source. All direct quotations must be shown clearly in “quotation marks” with the full reference.

The overall Similarity Index might be as high as 30-40%, but this is not necessarily a problem if each of the individual text-matches are less than 10%, and are appropriately referenced. The important thing is for you to use the Originality Report from your draft submission to carefully check that you have referenced all sources throughout your essay. If your essay has a very high overall Similarity Index (e.g. over 50%), this is probably an indication that you have not used your paraphrasing and summarising skills adequately. You will need to carefully recraft the essay so that you have not simply ‘cut and paste’ from sources without using your own words (Bretag, 2008a).
As the excerpt above indicates, electronic detection has the capacity to be part of the educative process, rather than a punitive response to plagiarism. Recent research has indicated that when electronic detection software is used in this way, students and instructors can work collaboratively towards improving academic literacies and learning outcomes. Keuskamp and Sliuzas (2007) maintain that text-matching software can provide educative opportunities, but suggest that students’ academic literacies may need further development to fully benefit from the quantitative reports generated, a position with which we concur. Despite some of her own reservations, McKeever (2006, p. 163) also concludes that electronic detection software can be a “beneficial educational tool”, enabling tutors and students to work collaboratively on the development of academic skills. Culwin (2006) demonstrated an imaginative use of detection software within one course as part of an overall plagiarism prevention strategy, and Morse (2006) has argued for: “...a proactive response to plagiarism, such as class discussions and assignments examining the many complexities and implications of plagiarism combined with students adopting Turnitin.com as a writing tool [which] allows instructors to engage students in responsible academic writing” (Morse in Donnelly, Ingalis, Morse, Castner & Stockdell-Giesler, 2006).

Students have also reported favourably on the use of Turnitin as an educative tool. Based on a survey of 152 students who had used Turnitin to draft assignments prior to final submission, Cheah and Bretag (2008b) found that 71.4% of students agreed that Turnitin helped them to identify sections of text that had been copied directly from sources; 74.2% stated that the program had helped them to identify areas that required editing to avoid plagiarism; and 70.5% believed that using Turnitin in this way had made them more aware of academic integrity. In the course in which the surveyed students were enrolled, one-on-one tutorial assistance was provided to every student to interpret the Originality Reports generated by Turnitin. Students were then encouraged to use the Reports to assist them in redrafting their papers prior to final submission. Where students were still concerned about the high text match indicated in the Overall Similarity Index, further consultation with their tutor was provided, without any suggestion that a high text-match equated to deliberate plagiarism or possible penalty.

Conclusion

In agreement with other theorists, we maintain that the very concept of ‘plagiarism’ is complex and therefore difficult to define. Having reviewed a wide variety of ‘plagiarisms’ identified in the literature, we conclude that the common element is the lack of appropriate attribution to the original source. An increased interest in plagiarism in recent years has led educators to find the most appropriate ways to deter, detect and deal with plagiarism (Carroll, 2003a). Electronic detection software potentially contributes to each of those three areas, although as the literature makes clear, it is not a ‘magic bullet’ (Carroll, 2003a), not least because it is only suited for identifying ‘word-for-word’ or ‘direct’ plagiarism and then only in electronic form. There are many other, more subtle forms of plagiarism which are not available for analysis by electronic detection software.

Using our own research and practice as a basis, we conclude that despite its shortcomings, electronic detection in combination with manual analysis, nuanced academic judgement and clear processes provide the means to determine if plagiarism has occurred. This paper has provided a simple model for identifying student plagiarism, and suggested that the most appropriate use of such a model is in collaboration with students, prior to the final submission of assignments. If used within an educative framework that encourages students to take responsibility for their own learning, while working hand in hand with instructors during the drafting stages, we believe that electronic detection software provides one means of ensuring academic integrity in students’ written assignments.
A model for determining student plagiarism: Electronic detection and academic judgement

Tracey Bretag and Saadia Mahmud

References


