Evaluating an electronic plagiarism detection service
Evans, Robert

Postprint / Postprint
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
www.peerproject.eu

Empfohlene Zitierung / Suggested Citation:

Nutzungsbedingungen:

Terms of use:
This document is made available under the "PEER Licence Agreement". For more Information regarding the PEER-project see: http://www.peerproject.eu. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public. By using this particular document, you accept the above-stated conditions of use.

Diese Version ist zitierbar unter / This version is citable under: https://nbn-resolving.org/urn:nbn:de:0168-ssoar-231150
Evaluating an electronic plagiarism detection service

The importance of trust and the difficulty of proving students don’t cheat

ROBERT EVANS  Cardiff School of Social Sciences, Wales

ABSTRACT  Plagiarism by students is seen as an increasing problem. The fear is that students will use the internet to obtain analysis, interpretation or even complete assignments and then submit these as their own work. Electronic plagiarism detection services may help to prevent such unfair practice but, in doing so, they create a new problem: certifying the absence of plagiarism. This article reports the results of an evaluation of one such service within an interdisciplinary school of social sciences. The article describes how the system works and the experiences of staff and students in using the service, together with an evaluation of the data generated. The key findings are that the service did identify examples of poor scholarship and unfair practice that had been missed under the usual marking system but that rigorously checking every script for plagiarism was impractical. Trust and student honesty thus remain central to a successful academic system.

KEYWORDS: cut and paste culture, electronic plagiarism detection service, plagiarism, trust

Introduction

The threat posed to the integrity of academic awards by internet essay banks and other web resources is seen as an increasingly serious one for higher education. The ease with which text, numbers and computer codes can be moved between students and institutions has the potential to undermine traditional forms of learning and assessment. Evidence from students and staff suggests that these dangers are not merely hypothetical, with stories about the rising incidence of plagiarism occurring in both the professional and popular media (e.g. Adenekan, 2003; Das, 2003; Diekhoff et al., 1996;
Gibelman et al., 1999). The emergent scenario is of a ‘new plagiarism’ in which learning suffers as students ‘wield an Electronic Shovel . . . to find and save huge chunks of information with little reading, effort or originality’ (McKenzie, 1998).

Tackling this problem is a serious challenge. Teaching, assessment and marking practices are all implicated in preventing plagiarism but credible attempts to detect it and, if necessary, impose sanctions are also important (Braumoeller and Gaines, 2001; Woessner, 2003). In this context, the internet may not be all bad. Just as it provides new opportunities for cheating so it also provides new opportunities for detecting plagiarism. Indeed, the irony of the on-line source is that if students can find it, there is a fair chance that those assessing their work can as well (Auer and Krupar, 2001; Smith, 2003). Electronic plagiarism detection systems have the potential to extend and automate this work but their use is largely unreported in the academic literature. This article begins to fill this gap by reporting the experiences of the Cardiff School of Social Sciences in implementing an electronic plagiarism detection service as part of a wider strategy to prevent plagiarism and improve standards of scholarship amongst its students.

**Literature review**

McLafferty and Foust neatly capture the importance of plagiarism as a breach of trust when they say that ‘Words are academics’ currency and bond’ (2004: 186). If we cannot trust that academics’ words are really their own, then traditions of teaching, peer review and authorship are undermined in the same way that economic activity is undermined if we are unable to trust the purchasing power given to bits of paper, metal and plastic (cf. Giddens, 1990: 26). In trying to address plagiarism by students, research has mainly focussed on three main areas: estimating its prevalence; identifying its causes; and developing strategies to prevent and/or detect it. This last element is the one that is typically assumed to be most troubled by the availability of internet sources, though the potential of the same technology to detect plagiarism may mean that the situation is more finely balanced than the hyperbolic rhetoric of the ‘cut and paste’ culture suggests.

In estimating the prevalence of plagiarism, it is important to recognize that the term itself conceals a variety of misdemeanours, ranging from poor referencing and paraphrasing, to deliberate attempts to cheat by copying whole paragraphs or even essays. In addition, the source of the plagiarized material can vary: it may be the student’s own work, where assignments are ‘recycled’ for different modules; other students, through deliberate or accidental collusion; as well as the more widely discussed academic or
commercial sources (Park, 2003: 475–6). In distinguishing between different types of plagiarism, some authors (e.g. Braumoeller and Gaines, 2001: 837) thus distinguish between ‘casual’ and ‘blatant’ plagiarism, with the former denoting poor scholarship and the latter a more deliberate attempt to cheat. Taking a fairly broad definition of plagiarism, research generally suggests that a large proportion of students are willing to admit to either having engaged in some form of plagiarism or being able to imagine circumstances in which they would. For example, Underwood and Szabo (2003) report that about 1 in 2 students said they would engage in plagiarism to avoid failing a module, while a more recent article by the same authors suggests a wider range of reasons might legitimate plagiarism (Szabo and Underwood, 2004). Other studies give different estimates, but the general picture is of a significant, though perhaps not yet rampant, problem. For example, Franklyn-Stokes and Newstead (1995) put the proportion of students who engage in plagiarism at between one-third and one-half, while a number of other studies put the proportion of students who are prepared to admit having engaged in some form of plagiarism or cheating (broadly defined) at over 50 per cent. Examples of these studies, which are all cited in Park (2003), include Haines et al. (1986), Hollinger and Lanza-Kaduce (1996), Brown (1995), and Stern and Havlicek (1986). Finally, it should be noted that although some studies appear to replicate the finding of very high rates of plagiarism others report lower rates. Thus, for example, Jensen et al. (2002) report that up to 90 per cent of US students have engaged in some form of malpractice or cheating while Scanlan and Neumann (2002, cited in Kellogg, 2002) report lower rates of around 25 to 30 per cent.

There appears, therefore, to be little doubt amongst the research community that plagiarism, at least in its minor forms, is fairly widespread. But what is the reason for this? Although ready access to pre-prepared information is clearly a factor in making plagiarism easier and arguably more prevalent (Baty, 2000) it would be wrong to think that it is the only one. Research on the causes of plagiarism paints a much more complex picture. A simple typology, based on Park (2003), includes psychological determinants such as hidden memory (e.g. Macrae et al., 1999) and personality (e.g. Raffetto, 1985; Buckley et al., 1998); demographic factors such as age (e.g. Haines et al., 1986; Straw, 2002), gender (Calabrese and Cochran, 1990), country (Diekhoff et al., 1999; Lupton et al., 2000), ethnicity and culture (Burnett, 2002; Deckert, 1993), and student lifestyles (Straw, 2002); as well as factors related to the course of study (Meade, 1992), the style of teaching (Burnett, 2002; Gerdeman, 2000), lack of training in how to reference (Roig, 2001), the perceived chance of being caught and the consequences of this (Braumoeller and Gaines, 2001; Davis and Ludvigson,
and the changing nature of higher education from elite to mass participation (Ashworth and Bannister, 1997).

It is also worth noting that the expectations of academics, the so-called ‘hidden curriculum’ (Snyder, 1971), will also influence the ways students approach their assessments. Whilst some studies suggest that student use of information and communications technologies (ICT) is tacitly discouraged by assessment strategies (Marriott et al., 1999; Selwyn, 2002; Selwyn et al., 1999), the rhetoric surrounding internet plagiarism suggests an alternative scenario in which the use of at least one form of ICT is rampant (e.g. Laird, 2001). In this context, the moral nature of assessment as an activity through which academic standards are reproduced (Knight, 1995) becomes central. The signals that academics send to students about the relevance of the courses they are taking (Gerdeman, 2000), the care with which they will be marked (Burnett, 2002) and what counts as acceptable use of the internet and other electronic resources (Boud, 1995; Macfarlane, 1998) are all important influences on students. In other words, although students are rightly encouraged to take advantage of the internet as a source of information, academics must ensure, through their own actions, that students realize ‘learning requires more than high-speed connections and a good search engine’ (Carnie 2001, quoted in Park, 2003: 481).

Despite this complexity, however, it would be wrong to say that academics are helpless. There is, in fact, a great deal that we can do, including maintaining good relations with students, designing assessments so that they are difficult to plagiarize in the first place and providing clear and consistent advice to students (Harris, 2001; Lathrop and Foss, 2000; Pearson, 1999). Nevertheless, just as it would be wrong to despair, so too it would be wrong to be complacent. Carefully worded policy statements (Brown and Howell, 2001), consistent training (Auer and Krupar, 2001; Wilhoit, 1994) and developing credible means of detecting plagiarism (Maramark and Maline, 1993) are also necessary. Intriguingly, it now seems that the very same internet that threatened to be the scourge of education could now be its saviour. Where it might once have taken many hours to document and prove suspected plagiarism search engines and dedicated electronic plagiarism detection services can do this work automatically and in minutes. Detecting electronic plagiarism now appears to be almost as easy as committing it.

The use of such systems is still largely unreported in the literature, however (though Braumoeller and Gaines (2001) provide an exception). Whilst using such a system is not a panacea, it may provide a useful resource for an important part of the academic job – the grading of coursework and the awarding of degrees to students. In using these services,
however, markers are introducing a new element into the assessment process that has the potential to change the relationship between staff and students from one of trust to one of distrust and surveillance. The relative balance between these different elements is crucial to the successful use of plagiarism detection services and the evaluation described in this article thus considers student and staff experiences of the system as well as the usefulness of the data produced.

**Evaluating the plagiarism detection system**

The plagiarism detection service used in this evaluation was developed by iParadigms and made available via the UK Joint Information Services Committee (JISC). In order to use the system, coursework must be prepared electronically and submitted via a web interface. Each assignment is then checked against a series of databases and an ‘originality report’ produced for the marker(s). The originality reports indicate the proportion of the text in the assessment that has been found to match text available elsewhere as well as highlighting this ‘matching text’ and indicating where the ‘original’ text is to be found. It is important to note that the originality reports do not identify plagiarism per se. All they do is identify text that matches other text. Markers must decide for themselves whether the text represents plagiarism, a properly referenced quotation or something in between.

The JISC electronic plagiarism detection service was piloted on year two undergraduate students and the evaluation was run in parallel with the normal paper-based assessment process. Students were thus required to submit paper copies of their assessed coursework as normal and then asked to submit the same work electronically. Although not ideal in terms of response rate and self-selection bias this approach did ensure that students were protected from any disadvantage if the system did not work as expected or hoped. The first part of the trial involved training the students to use the system. The training was organized in groups of 20–25, with students receiving a brief introduction to the service and a more detailed guide to enrolling on the system, signing up for classes and submitting coursework. During this training session students were also given a copy of the JISC Student User Guide and a feedback questionnaire to record their first impressions of the system and their level of IT use. When the students returned to university after the Christmas vacation approximately 85 per cent of those who had registered before Christmas continued to participate in the trial, submitting a total of 513 pieces of coursework.
Results

Student co-operation was generally good. Of a possible 240 students, 170 turned up during the last teaching week of the semester and registered on the system. As shown in Table 1, the feedback questionnaires revealed that students’ assessments were largely positive, with the median response for most items being ‘agree’. The only exceptions were the statement that ‘plagiarism was a problem’ and that the student would ‘prefer to submit coursework electronically’. In both these cases, the median response was the more neutral ‘neither agree nor disagree’.

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Neither agree nor disagree</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>User guide was clear</td>
<td>46</td>
<td>101</td>
<td>14</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Enrolling on the JISC system was easy</td>
<td>64</td>
<td>81</td>
<td>14</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Enrolling on modules was easy</td>
<td>68</td>
<td>80</td>
<td>10</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Submitting coursework was easy</td>
<td>32</td>
<td>80</td>
<td>27</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Plagiarism by students is a problem</td>
<td>13</td>
<td>45</td>
<td>85</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Perceptions of plagiarism are a problem</td>
<td>10</td>
<td>74</td>
<td>64</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>I know how to avoid plagiarism</td>
<td>25</td>
<td>109</td>
<td>19</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>I know how to cite sources correctly</td>
<td>24</td>
<td>118</td>
<td>12</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>I would prefer to submit c/w electronically</td>
<td>18</td>
<td>36</td>
<td>58</td>
<td>39</td>
<td>12</td>
</tr>
</tbody>
</table>

These two items are particularly significant, not least because they fit uneasily with the literature cited earlier. For example, although students tended to agree that perceptions of plagiarism were a problem they did not share this view themselves. This may be because they genuinely believed, correctly as it turned out, that there was not a significant amount of plagiarism going on and/or that they viewed the plagiarism they did know about as minor and therefore ‘unproblematic’. In any case, they, again
correctly, rated their own ability to avoid plagiarism quite highly, something that is also somewhat at odds with the literature, which typically suggest that students are uncertain about how to reference work. These feelings and beliefs no doubt feed into the responses to the final question, which asked whether or not they would prefer to submit their coursework electronically. Although about one-third agreed they would prefer to submit coursework in this way, and thus change from the current practice, another third disagreed so that, depending how or where you count from, the survey provided equal evidence of support for the new system and for maintaining the status quo.

Given that the students generally found the system easy to use, and had the basic IT skills and access needed to use it, explanations for this opposition are to be found in students' perceptions of the system and, in particular, its effect on the way they experience assessments. At an individual level, submitting coursework electronically requires trusting computers, and the absent others who program and maintain them, to do a task that is currently done physically and visibly by students and staff. Some evidence of the role this face-to-face interaction, and the trust it engenders, plays in the handing in of assignments was given by one of the students, who wrote on the back of the feedback form that submitting coursework electronically was:

Not a nice experience if computers scare you!! Wouldn’t feel I could rely on it. Not same feeling of relief as physically handing it in.

No doubt other students are less concerned with the embodied performance of handing in coursework but this is not the only way in which submitting coursework via an electronic plagiarism detection service raises issues of trust and distrust. Changing the way coursework is submitted changes the way students interact with the School as the new technology reconfigures student identities and their position with respect to the university (cf. Bijker et al., 1987; Woolgar, 1991). Making the detection of plagiarism routine also makes distrust the norm and, although many students seemed happy to accept the increased convenience of submitting coursework electronically, it would not be surprising if a minority resented the implication that they needed to have their honesty demonstrated.

Submitting coursework is only part of the problem. A more significant part of the evaluation related to the way the plagiarism detection service changes marking practices and, in particular, the extra work that is created by the originality reports, which now have to be examined as well as the essays. At its most basic, the originality report is an electronic version of the student’s essay with the words that match text that can be found elsewhere highlighted in different colours. The total number of matching words is used to generate a score through which the originality reports are
themselves colour-coded into bands that represent different degrees of potential plagiarism. Blue thus represents assignments in which five words or fewer match the data base, whereas red denotes an originality report in which over 75 per cent of the assignment text matches text found elsewhere. These ranges, together with the number of scripts that were detected in each category for our sample, are illustrated in Table 2:

<table>
<thead>
<tr>
<th>Colour</th>
<th>Blue</th>
<th>Green</th>
<th>Yellow</th>
<th>Orange</th>
<th>Red</th>
</tr>
</thead>
<tbody>
<tr>
<td>% matching text</td>
<td>&lt;5 words</td>
<td>0–24%</td>
<td>25–49%</td>
<td>50–74%</td>
<td>75–100%</td>
</tr>
<tr>
<td>No. of scripts detected</td>
<td>31</td>
<td>471</td>
<td>7</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

In addition to the overall proportion of matching text, the originality report also provides information about where the matching text is to be found. Perhaps unsurprisingly, given that participation was voluntary, we found that very few of the essays submitted (11 out of 513 or 2 per cent) contained more than 25 per cent of matching text, supporting our students’ belief that plagiarism was not a problem. In fact, closer inspection showed that it was even less of a problem than the table suggests, as all three ‘red’ essays were attributable to students mistakenly submitting the same essay twice. Of the others in this more serious category, that is, the orange and yellow rated essays, examining the originality reports revealed a variety of practices ranging from plagiarism to poor scholarship and referencing and including one instance of submitting very similar essays for two separate modules. Significantly all of these essays would have justified some disciplinary or remedial action had they been detected as part of the normal assessment cycle.

This outcome clearly indicates the success of the system in identifying plagiarism that had been missed by markers and provides a positive argument for using the system. Nevertheless, there is also a downside. Because the system cannot tell plagiarized text from properly attributed quotations it is possible that the simple quantitative score will not provide an accurate guide to the amount of plagiarized text, particularly in cases where the plagiarism is minor. This is because a well-written essay, which starts with a restatement of the question, contains some properly referenced quotations, and ends with a comprehensive bibliography will also contain text flagged by the system as potential plagiarism. As a result, the system has the potential to undermine itself by generating so much data that, if every originality report is checked, marking will become an even more time-consuming and cumbersome process than it already is.
In our experience, however, this outcome is unlikely. As shown in the table, the green category includes over 90 per cent of all the scripts submitted and to check each report would require a significant amount of work. To see if this would actually be necessary, we picked a sample of approximately 50 essays coded as green and examined what text was being matched and whether or not it constituted malpractice or plagiarism of the kind that was evident in the other essays. In general, we found that essays coded as green contained relatively low proportions of matching text and were clustered around the lower end of the band (average = 8%). What is more, within these essays, the vast majority of matching text came from things like the essay title, references in the bibliography and attributed quotations. In another module, where the assignment consisted of a report on practical work, commonly matched text included things like experimental instructions, protocols and the data. In other words, in most cases, there was very little cause for concern about the work presented and little that would count as even ‘casual’ plagiarism. Of course, there were some exceptions to this pattern. Some essays coded as green were towards the upper end of the band and displayed weakness that one would want to address. These included poor paraphrasing and referencing of sources and, in one case, a close reliance on another student’s interpretation of statistics, though whether this was uncertainty about the material or a deliberate attempt to cheat is unclear.

The result was thus somewhat paradoxical. Although we found that the JISC system clearly does have a useful role to play in identifying plagiarism, its wholesale use is problematic. Whereas much of the literature addresses the problem of identifying cheats, our biggest problem was proving their absence. The small number of cases coded as yellow or worse meant that it would be relatively simple for markers to check all originality reports where a match of 25 per cent or greater was found. Instead, the problem arises in the case of those students – the vast majority in our sample – who did not attempt to cheat. Checking all reports coded as green does not seem a viable strategy so that demanding proof of honesty – trying to prove the negative of ‘no plagiarism’ – has the potential to overwhelm the marking process. In other words, some way of reinserting ‘trust’ back into the marking process is needed as, to put it bluntly, the cost of the false negatives that are allowed to slip through by not checking for minor plagiarism is not enough to justify the effort it would take to prevent them.

Discussion
This article has examined how the electronic plagiarism detection service offered in the UK by JISC worked in the context of an interdisciplinary
school of social sciences. We found that student views on the system were evenly distributed between support, opposition and something in-between. Perhaps surprisingly, given the voluntary nature of the trial, which offered nothing except the chance of being caught cheating, we achieved a response rate of about 85 per cent from those who registered for modules, with this figure itself representing about two-thirds of the possible students. Based on these experiences we found that the vast majority of those who did take part were quite capable of using the system properly and that many of the problems experienced could have been avoided if better instructions had been provided earlier. Against this positive experience, it was also noted that issues of trust and distrust – both in computer systems and students – were also implicated in their responses and some further research to address these issues would be useful.

From a marking perspective, the service appeared to perform effectively with regard to major or blatant plagiarism. Originality reports were generated quickly and appeared to show that the vast majority of students did not engage in activities that could be classed as plagiarism. In addition, we were also able to identify eight pieces of work where standards of scholarship were unacceptably low and where remedial or disciplinary action would have been appropriate. This is important as acting on these cases, which were missed in the normal marking process, would show the honest majority that those students who do cheat get caught and thus also have a deterrent effect in future years. For minor plagiarism, however, the findings are more ambivalent. Although, in principle, it would be possible to check every originality report, and identify even the most casual or minor plagiarism, this is not practical. Instead, there has to be a cut-off point below which originality reports are themselves simply taken ‘on trust’ and are not checked further. As a result, and despite the appearance of total surveillance created by the rhetoric of text matching, the practical implementation of the system on any substantial scale requires a degree of trust and tolerance within which minor or casual plagiarism may go un-sanctioned, if not undetected.

In summary, therefore, the technology of electronic plagiarism detection software does not undermine trust in the way that it might appear. Although the routine checking of all assignments appears to signal an institutional distrust, this strategy is unsustainable on a large scale. Checking every report for originality has the same effect as checking every coin or cheque for authenticity – the system simply grinds to a halt. Instead, only major cases can be fully investigated, so that, if the system is to work, and if academics are to have time to teach, then the integrity of students remains essential.
References


**Biographical note**

ROBERT EVANS is a Senior Lecturer in Sociology in the Cardiff School of Social Sciences. His research interests focus on Science and Technology Studies (STS) and, in particular, around the relationships between expert knowledge and public participation.

Address: Glamorgan Building, King Edward VII Avenue, Cardiff CF10 3WT, UK. [email: EvansRJ1@Cardiff.ac.uk]