

Detecting plagiarism and taking appropriate action

Written by Jude Carroll, March 2009

These are my personal views. They do not represent those of any institution or to represent advice which has any legal standing. If you quote this, these reservations should be made clear.

Detection alone will not sort out the problem of plagiarism and could threaten the experience of both teachers and students. Detection needs to be part of an integrated approach, combining detection efforts with better course design and induction of students, with effective teaching as to how to avoid plagiarism or collusion and with instruction on how to uphold academic values and conventions. However, a holistic approach does rely on detection as one vital aspect of the overall package – ignore it and the rest is less effective. If you don't show that you are taking steps to identify instances where students have not followed the rules, then they are likely to invest less effort themselves in doing so.

This paper deals with detection in that spirit.

What are teachers looking for?

Teachers are not looking for plagiarism. Instead, a teacher should be looking for instances where the student has submitted work that is not 'his or her own work'. This could be instances where work is copied, where the student draws on others' work without acknowledgement, where the student did not work sufficiently independently, and/or where the student's submission does not allow the teacher to make a judgment as to whether or not the student has met the learning outcomes.

The teacher is also looking for deception. Has the student deliberately tried to fool the teacher or to make the teacher think that the work is that of the student who submitted it? Has the student taken steps to create a false impression in the teacher as to whose work is being assessed?

The teacher who starts to wonder, 'Who did this work?' or 'Is this the student's own work?' or 'Where did these ideas and these words come from?' should then be identifying things which help him or her decide one way or the other, 'Is this plagiarism?'. The statement, '*This is plagiarism*', can only be made by a person and can only be a judgment made after considering a range of facts, findings and events. This paper describes some of them.

Detecting plagiarism before the student's work is graded

It is possible to screen students' work prior to marking it. You can do this by using a range of electronic tools and via other, non-electronic methods.

Electronic tools require electronic submission. Electronic submission is usually fairly straightforward via the University's Virtual Learning platform or through a central computing facility. This process also creates a record of the time and date when the work has been submitted and means that it is available for other types of inspection alongside text-matching. University systems are also designed to protect against viruses etc and ensure much more efficient processing than when student hardware (discs, CDs etc) are used.

Screening using commercial software looks for similarity between students' work and that in the databases and within the 'reach' of that particular software.

Databases usually include:

- offerings on the current and archived Web;
- previously submitted student work;
- and a collection of protected and paper-based texts to which the tool has gained access.

Database coverage for particular tools varies widely and will therefore yield different results depending on, for example:

- the type of student submission,
- the time when the search is attempted,
- the ability of the commercial provider to negotiate access to private or copyright sources,
- how deeply embedded in a particular website the original source was located,
- whether or not the search robot has been excluded from a site

and so on. The corpus of work held in different text-matching databases will also be more or less relevant to different disciplines. In general, arts and humanities sources are more available than science and technology – but this is only a generalization. When deciding to purchase this software, the reach and coverage of any one tool plus a range of questions should be asked in order to compare different tools. Where others have attempted to answer the question, 'Which tool is best?' in the past, the results are often very different depending on the criteria used. In my view, no text-matching tool is rated as 'very good' when compared to others but several are rated as 'good' by researchers and by users and a few are certainly more widely used than others.

Most text-matching tools use a similar process:

- **The text is digitalized** then 'chopped' into segments. Providers typically will not say how long these digital segments might be.
- **The tool searches** the various databases and the Web, looking for matching text for a number of these digital segments.

- **Most software adjusts the search when a match or matches are found to ‘look’ more deeply.** Most also have ways to disregard trivial matches.
- **A report is generated.** Reports differ but most will state the percentage of the student’s text which matches text in its database; most state which sources provided the match. Some rank the sources according to their percentage of matching text in the student’s work. Many show how the student has altered the original text as long as the alterations fall within the boundaries of the digitalized ‘chunk’.

Is text-matching useful?

Advocates of text-matching software say the tools are useful, authoritative and fast. [Note: all reports will need to be reviewed to check that the matches are relevant so even though checking a report is usually a quick process, the time requirement can accumulate in a large class.] As more and more students and institutions submit their work, then the databases can provide a national/international corpus of material. The tool also copes with the difficulties of having many markers in a large class who, because they see only some scripts, could not identify instances where students have copied from each other. A further benefit is that some teachers use submission for formative feedback – that is, as part of their teaching strategy for improving students’ academic writing. Those in favour of using text-matching software claim it is more objective than relying on changes in the student’s writing style or language capability in order to identify copying. If only change of language is used to spot copying, then the worry is that some groups of students are likely to be disproportionately represented amongst those identified for potential disciplinary action. International students writing in English are often mentioned in this context. Finally, advocates of text-matching software use claim a deterrent effect since students know their work will be checked.¹

There are also numerous detractors for using text-matching software. Reasons include the limits of their ‘reach’ which means that any tool is unlikely to find all or even a significant number of instances of copying in any one student’s text. According to some studies, even the top-rated tools only find at most, 50% of the instances of copying in a student’s paper. Tools cannot cope with many kinds of copying – for example, it disregards diagrams, images, text which is held behind ‘firewalls’ or on protected databases. Though the number of journals and textbooks which do allow their text to be used is growing, most databases hold a small or even tiny proportion of the sources which students might use. Tools cannot identify copy that has been translated, cannot identify examples where students have commissioned others to write the work, and many have a language-specific bias. Despite this significant limitation in capability, teachers can become less vigilant themselves and therefore miss a significant number of instances of copying or failing to attribute work.

¹ Studies to support this claim vary. Some show this is the case, others show the opposite. The claim remains unproven.

Further difficulties with text-matching include the time it takes to load and process reports and, since most students are not plagiarizing, the pointless use of this time. One of the biggest dangers, many say, is that teachers have too high expectations for 'plagiarism detection' since such tools can only match text. Managers, too, hold unrealistic expectations about how such tools can detect and deal with plagiarism cases. Any decision and judgment as to whether or not the match is significant remains with the teacher.

Looking for possible cases without using electronic tools

One marker of a small number of papers can usually remember and link similar ones. Also, students who have copied each other's work often hand their papers in together so maintaining the order of submission may help. You may also wish to compare exam marks with those gained via coursework and scrutinize scripts where the discrepancy is very large.

Detection of plagiarism when marking work

This is rarely a clear-cut decision that 'Yes, this is plagiarism'. Instead, the decision develops from collecting together several signs or signals that could mean 'this is not the student's own work'. This section describes several ways in which these signs and signals might be collected.

1. **Checking for Web matches.** If you come across text and think, 'Where did this come from?' then you might search the Web for similar text. The easiest way is via Google using the Advanced Search function. First, choose a phrase of up to 10 words (4 or 5 well-chosen ones is often better) which strikes you as unlikely to have been written by the student. On the Google homepage, click on the small hyperlink to the right of the search box that opens the Advanced Search facility where you put in exactly the phrase you have identified. If the phrase comes from the Web, the resulting list of url's usually has the likely source amongst the first few items. To find the phrase on the site, use the 'cache' button under the Google listing and the phrase will appear in (highlighted) yellow so you need not scroll through pages of text.

A Google Advanced Search is not language dependent and probably produces the most relevant sites of all the meta-search engines. Of course, it cannot search the parts of the Web that are protected or out of the public domain; it cannot search for text-based copying; and it only finds exact matches (compared to text-matching which can cope with some variation within the electronic 'chunk') . However, many teachers say that often, the student could not reach such sites either and support the utility of this approach.

2. **Checking through commercial software.** The previous section described how to use text-matching software to screen a whole class. Here, one student's paper that has triggered questions about its authorship could be submitted to see if matches are found. This requires electronic submission or it would require you to ask a student who has submitted a hard copy to supply an electronic version. Asking a student to do so within a reasonably short time frame might also generate additional information on authenticity since a student who submitted a word-processed document but could not provide an electronic version might be asked why this was the case.
3. **Checking for 'hidden' characteristics.** If the student has submitted electronically and the document is in Word, you can investigate the authorship of a document via the 'Properties' function under the 'File' button. This will tell you the name of the person using the programme, dates when the document was created and revised, and (under 'statistics'), the total edit time for the piece of work and the number of revisions. If these functions are without entries, students could be asked for drafts – but only if they were asked to keep them in the Course Handbook.

You might also inspect the formatting to see if a passage that seems to be downloaded remains in HTML or whether a range of formats are used, often a sign of 'patch writing'.

4. Noting the visible signs within the student's document

The most common way that teachers detect plagiarism is by noting characteristics of the student's work which are cause for concern.

These include:

- Change of language or of discourse style
- Change in level or complexity
- Obvious signs of copying such as urls left in the text, different fonts, formatting changes or grey-outs where hyperlinks are left in
- Bibliographic clues - mixed referencing systems, dated references; in-text citations not listed in the references; hard-to-access documents claimed
- Individual words, formulae, statistical manipulation, diagrams or images which would not be expected in a student paper; perfect punctuation
- Variation in pronoun use ('I' to 'we' to 'I')
- Out-of-character level of work for this student
- A feeling that "this reads as strangely familiar"

Any one of these characteristics, on its own, might have a perfectly reasonable explanation. If you find two or three of these

characteristics, it becomes less likely that the work is 'the student's original work' and more and more likely that levels of unacceptable copying are involved. **IMPORTANT:** *all of these are acceptable in a student's work if they are also cited and referenced using the normal conventions for acknowledging others' work. 'Acceptable', however, does not mean, they become worthy of a passing grade.*

Signals of plagiarism which are not in the work itself

A teacher who marks an assignment and gives it a grade is signaling that the student has met the learning outcomes and therefore deserves to have academic credit for this piece of work. A teacher who is not confident that the work is that of the submitting student will need to check further before the teacher can award a grade and before the teacher can award academic credit. Therefore, checking 'Who did this work?' is not a precursor to taking action under the disciplinary rules; it is a necessary requirement for awarding a grade.

Actions to check the authenticity and authorship of a piece of work could include:

- **Asking the student to come in for a discussion.** The discussion could cover the content of the work or the way in which the student went about creating it. Remember that these questions need to be open, neutral, and respectful. You are seeking information to be able to grade the paper - not assembling a case for discipline.
- **Asking the student to produce evidence of the process used to make the work.** This could mean asking to see drafts, copies of significant research articles referred to in the text, or notes. These can only be requested if the Course Handbook or PM included the requirement that they be kept for possible inspection.

Authentication can also be sought through asking others who might have been involved in the work about the claims in the student's submission. For example, a placement supervisor could verify that the incidents described by the student had occurred. Again, this kind of action would need to be specified as a possibility in the Course Handbook or PM.

Authentication that the work is indeed that of the student is especially necessary in some cases. Examples include where students execute work independently, perhaps over a long period of time, and where additional help is easily available such as when students are creating computer programmes. Authentication is necessary where students have avoided or not used monitoring opportunities such as supervision meetings or tutorial discussions then they hand in a fully-finished piece of work. It is needed where work should have been created in a public forum such as a workshop but where the student instead, created the work

independently and beyond the observation of the marker. In all these cases, a teacher might wonder 'Who did this work?' and must then take steps to find out, before awarding a grade.

Suspicion of deception

Cases must be reported for disciplinary action when the teacher suspects there is deception. **IMPORTANT:** the teacher is not reporting plagiarism, the teacher is reporting deception and the way in which that deception was carried out was by submitting someone else's work , i. e. by plagiarism.

It is difficult to create a list of signs and signals of student deception. The most helpful way this can be done is through teachers discussing and collecting examples in their own departments, usually over time. Typical examples might be:

- Where students took steps to cover up or hide their unacceptable behaviour. For example, if a student shows he or she understands referencing conventions by using them correctly in most of the text then omits all citation and attribution signals for a subsection which contains copied text, this probably shows deliberate deception.
- Where students alter text so that it conveys a meaning different from that it held in the original version and where the alterations are designed to make the student appear more skillful, recent, authoritative or credible. For example, a student may use a document from 1992 and cite it as being from 2007.
- Where work includes lies or fabricated materials.
- Where no reasonable person would consider that the action would be acceptable. So, for example, a student who has only recently enrolled in a Swedish Masters programme can probably support a claim for misunderstanding of the Swedish requirement to 'do your own work' if he or she creates a text by using 8 or 10 sources and patching them together into a piece of work where the student has written linking text. The same student, submitting a text which is 100% copied from a cousin who followed the course the previous year, could almost certainly not support the claim that the submitting student thought of the work as 'my own'.

Enough to take action; enough to report

All examples of plagiarism must be managed rather than overlooked. Those where the student is still learning the rules and the necessary skills to meet the requirement for independent work can be managed by the teacher – this is a learning issue. Those who break the rules but where there is no suspicion of deception must be managed within the department, either by awarding a lower grade or by choosing from agreed consequences. Consequences could include asking the student to resubmit the work following corrections, to resubmit another piece of work or to award a ‘fail’ grade.

It is assumed that most cases or indeed, all but a very few cases will fall into the previous two categories.

Those submissions where there is reason to suspect deception must be reported. All Universities will have stated procedures for reporting cases of deception and these must be followed exactly. The grounds for reporting will also be stated in any reporting document and will include the sorts of evidence gathered through actions listed in this document. It is for the Disciplinary Board to decide whether that is sufficient to support a case for imposing a disciplinary penalty.