Using Turnitin as a tool for attribution in cases of contract cheating

Thomas Lancaster
Birmingham City University
Millennium Point
Birmingham B4 7XG
thomas.lancaster@bcu.ac.uk
http://thomaslancaster.co.uk

Robert Clarke
Birmingham City University
Millennium Point
Birmingham B4 7XG
robert.clarke@bcu.ac.uk
http://uk.linkedin.com/pub/robert-clarke/16/12a/967

Abstract
Although methods for automatically detecting contract cheating, that is finding students who have outsourced the production of their assessed work to a third party, have been proposed, successful implementations of these detection methods have not yet been reported. This paper instead reports on an investigation to make use of a database of known work for this purpose. The work is accessed through the non-originality engine Turnitin, against which attempts at contract cheating found on agency websites are matched.

369 assignment specifications found on online agency contract cheating sites, such as Freelancer.com, were collected between January and November 2013. These were all assignment specifications for which attempts to attribute these with any level of certainty to an academic institution had proved impossible to a contract cheating detective. The assignment specifications all represented cases that looked likely to belong a UK educational institution. The assignment specifications were run through the Turnitin database in use within the UK and the results analysed as part of a process attempting to notify tutors that one of their students may be attempting to cheat.

The initial indications were that the use of Turnitin was of value to the contract cheating detection process, with 105 out of 369 (28.5%) initially identified. 2 out of 369 (0.5%) were subsequently found by a tutor at the institution concerned as a result of being in the database. However, several challenges were identified that will require the Science, Technology, Engineering and Mathematics (STEM) communities to come together and work to improve the use of Turnitin within the contract cheating detection process. This paper explores the results of this study and the wider issues surrounding the use of Turnitin for the detection of contract cheating.

Keywords
contract cheating, attribution, academic integrity, Turnitin, plagiarism
1. Introduction

The body of literature on contract cheating, the type of academic misconduct where students outsource the production of coursework which they then submit to be assessed, has shown that this form of cheating is both prevalent and difficult to detect (Clarke and Lancaster, 2006; Lancaster and Clarke, 2007a; Lancaster and Culwin, 2007). An investigation into contract cheating on an MBA Marketing module in the UK provided an estimate from students that 40% of them had committed contract cheating (Wellman and Fallon, 2012).

Whilst plagiarism from external sources can usually be detected by non-originality engines such as Turnitin (2013), contract cheating cannot be easily detected in this way. This is because students are submitting work that is original, but which they have not themselves written.

Much of the quantitative research into contract cheating has investigated the use of agency websites, of which the current market-leader seems to be Freelancer (2013). These are sites where students can place their requests to have assessed work completed for them amongst legitimate requests by businesses. The students can take advantage of a competitive marketplace of high and low bidders, all of whom are rated for the quality of work they provide. Whilst programming work from these sites has not always been found to be of high quality, this unoriginal work has been shown to not raise alarm with markers (Jenkins and Helmore, 2006). The financial commission generated for agency sites and other contract cheating sites has been found to be substantial (Clarke and Lancaster, 2013; Matthews, 2013).

Well-defined methods, both software and human-led, exist to prevent and identify student plagiarism, but not contract cheating (Culwin and Lancaster, 2001a; Culwin and Lancaster, 2001b). Methods of setting assessment to make contract cheating more difficult, such as using viva voce examinations and practical tests, have been identified (Lancaster and Clarke, 2007b; O’Malley and Roberts, 2011). Current detection of contract cheating relies heavily on the human role of the detective. This is someone who monitors known agency sites, and tries to extract information from postings and match this through search engines to identify academic institutions and cheating students (Lancaster and Clarke, 2012).

The attributability of assignments to institutions and individual students has been shown to be a problem, with only 23.7% of postings from one agency site able to be attributed (Lancaster and Clarke, 2012). Some assignments may be too generic. In other cases, it may be possible to identify which academic institution an assignment was from, but finding out which student submitted the request from a large cohort has been shown to be impossible.

Partially automated methods of detecting contract cheating, which work by automatically capturing information about student requests and the context in which they were made, have been proposed (Clarke and Lancaster, 2007; Lancaster, 2013; Koumantaris, 2011). However, for successful implementation, these techniques would depend on the availability of massive databases of known and identifiable assignment specifications. Such databases are not yet available.

This paper reports on a study of an alternative approach for attributing cases of contract cheating to academic institutions using Turnitin. This allowed a detective to alert those institutions that monitoring of their student cohort was necessary. The success of this new approach to attribution is evaluated. Methods to increase the attributability of assignment specifications found online, deploying Turnitin, are proposed.
2. The Use Of Turnitin In The Contract Cheating Detective Process

Traditionally, the process used by a detective to find contract cheating is ad-hoc in nature. Due to the voluntary nature of the role, which fits around other responsibilities, not every suspect posting on an agency website can be investigated. Where a posting does appear to be an assignment specification and is investigated, successful attribution depends on the quality of search engine results, the availability of assignment details on the visible web, and often forensic analysis of trace evidence within the files uploaded by a student.

This investigation looks at the use of Turnitin with 369 items that were not originally attributed by a detective after the normal process of searching online and using trace evidence proved inconclusive. The data relates to contract cheating requests submitted during the period from January to November 2013. The assignment specifications processed through Turnitin included requests for original work. The items also included requests for extensions and rewrites of other work. This was either polish for the student to gain a higher mark, or to have the raw similarity percentage in copy and pasted work reduced.

The 369 items analysed through Turnitin were all selected because they showed associations with the UK. The association might be UK references such as UK locations, or UK currency. The association might also have been due to the use of UK English, or because the requester or the person who bid successfully to complete the work stated their location as being in the UK. The restriction to using likely UK sources was because the detective had access only the UK version of the Turnitin database. There are several versions of Turnitin used around the world with difference databases, presumably due to legal issues of data protection.

The decision to use Turnitin in this manner related to previous observations that sometimes students included a copy of their assignment specification along with the answers. Alternatively, other trace evidence may have ended up in the private Turnitin database that would otherwise be restricted to a university intranet or otherwise not available on the indexed web. Previous recommendations by Clarke and Lancaster (2007) had also asked that assignment specifications be directly uploaded to Turnitin, so as to provide sources for detectives to directly match contract cheating requests against, but this does not yet seem to have taken place with any degree of regularity.

3. Results Of The Initial Turnitin Investigation

The results of the initial investigation of the use of Turnitin to detect contract cheating that would otherwise not otherwise be attributable show some promise.

Out of the 369 items submitted, 105 of these (28.5%) had the source institution identified to the satisfaction of the detective. Sometimes this was the result of several very small matches to the same academic institution, which boosted the confidence of the detective that they had obtained a successful attribution. In other cases, this combined with suspicions that already existed. For instance, the student requesting the work may have provided a location near an academic institution identified through Turnitin.

When Turnitin identifies a match, it will not always allow direct access to the source document for a comparison by a detective. The usual rule of Turnitin use is that information collected from the web is immediately accessible. Where a document has been uploaded from another institution, the institution name will be shown, but the work submitted by that institution will not be directly accessible. An automated email request to the tutor responsible for that work has to be made, requesting the release of that work.
50 Turnitin requests for access to work were successfully returned. This provided many direct contacts for tutors to notify about contract cheating, and in some cases the detective was then also able to notify the source institution of other postings made by the same student alias. Some of these requests were for small percentage matches, for instance when a tutor setting assignments includes common text within these, which then became part of a student submission.

The low response rate to Turnitin requests needs to be investigated. This may be because of institutional decisions not to release the work of their students, the need to release this information may not be understood by the staff receiving the requests, or it may be that the tutors in charge of the assignment no longer held the post they then did, and the e-mail requests therefore vanished into a *black hole*. Some of the Turnitin requests made were for matches to work uploaded more than three years ago.

Figure 1 summarises the success of the contract cheating attribution process using Turnitin.

![Figure 1](image_url)

**Figure 1 – Categories Of Matches On Turnitin For Contract Cheating Cases**

13 of the 369 items were requests to have work rewritten. In 5 of these 13 cases, the source institution was positively identified.

12 out of the 369 items were judged to likely be Edexcel sub-degree courses, offered by many colleges. This meant that it would be very difficult to identify the originating institution unless other evidence was available.

In two cases, Turnitin provided matches to the web-based learning objectives of two universities in Australia. These learning objectives were not visible to Google since the information on the university websites had since been replaced, but Turnitin still had the previous items stored in its database.

Two other cases were not originally identified, despite being uploaded to Turnitin. However, the detective subsequently received Turnitin requests from other UK universities, since this matched work that they themselves uploaded at a later date. These are shown in Figure 1 as inbound requests. In one of those two cases, the member of staff has stated that they are now uploading their own assignment specifications as well as the student solutions. It is likely that, were both assignment specifications and the assignment details found on agency sites uploaded as a matter of source, the long-term attribution rates would be much
higher than they are at present.

4. Recommendations For The Use Of Turnitin

This initial study of the use of Turnitin has shown it to be an effective way of attributing cases of contract cheating that would normally be intractable. As well as identifying likely institutions in many cases, when Turnitin requests are returned, this also provides the direct contact email for a member of staff. This person likely set the assignment. If not, they are likely in the same department as the person who did the set the assignment. Such contact details are particularly helpful when many institutions restrict external access to email addresses.

Problems identified when using Turnitin included the situation where a group of staff teach on a module, and a paper view request is sent to multiple people. This is particularly the case for a cohort of final year project or dissertation supervisors, for example. There is also an issue when no response to a request is received. In such a case, an alternative contact mechanism needs to be established and publicised by Turnitin.

The geographical restrictions set for Turnitin use mean that many opportunities to match assignment specifications on agency sites to Turnitin content are missed. Since Freelancer, and other agency sites operates sans frontière, and many countries now offer tuition in English, alternative solutions are needed. This could be a single global database, or a way to ensure that contract cheating attempts are uploaded to all global versions of Turnitin.

Staff uploading their assignment specifications to Turnitin should now be considered necessary as a routine part of the assignment setting quality assurance processes. It would be helpful if both academics and external examiners insist upon this. This also provides a level of protection of the intellectual property rights of authors beyond that which would normally be in place if their assignment specifications were available to direct web searches.

There are still some academic institutions which do not require that all student work is uploaded to Turnitin. Even when work is uploaded, there are also institutions lacking a robust quality process to ensure that all such work is manually checked for issues of academic integrity. Contract cheating matches can be small percentages, and it is not sufficient for tutors to simply look at the high percentage matches that are emphasised by the colour-coding system used on Turnitin. The continued education of staff is necessary to show them the indicators to look out which may indicate a match between student work and the assignment specifications captured from agency websites.

Likewise, robust anti-contract cheating software is needed to assist detectives. This should ensure that the content of contract cheating requests is automatically archived locally and uploaded to all versions of Turnitin. It may well be that the Turnitin team are best placed to directly grab this content to include in their global databases, and to subsequently provide the opportunity for a match when tutors and their students later upload work to be checked for undue similarity. There may also be a mechanism through which Turnitin can automatically highlight matches to known contract cheating agency sites.

The success of these anti-contract cheating processes is contingent on the sector continuing to work to raise awareness of the issues associated with contract cheating. This initial study of the use of Turnitin by contract cheating detectives has shown that successful detection is no longer impossible, even without designated software. The Science, Technology, Engineering and Mathematics (STEM) community needs to come together to ensure the continued value of the academic qualifications that they are awarding.
References


Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

© 2014 The Higher Education Academy