

April 2005

**An Active Introduction to Academic Misconduct &
the Measured Demographics of Misconduct**

Fintan Culwin

Faculty of Business, Computing and Information Management
London South Bank University
Borough Road
London SE1 0AA

Email: fintan@lsbu.ac.uk
Telephone: +44 20 7815 7434

Fintan Culwin is a Professor of Software Engineering Education at London South Bank University, which he joined fifteen years ago after a career as a secondary school computing teacher. He has authored six programming text books, holds his PhD in software engineering education and has current interests in forensic document analysis, academic misconduct, automated assessment and learning objects for initial software development education.

Abstract

An activity designed to actively introduce first year students to issues related to academic misconduct is described. The activity involves computing students writing a short essay on a topic related to the history of computing. The essays are subsequently automatically checked for non-originality and the outcomes made available to the students. The results of approximately 350 students from two different sessions are analysed for any demographic influences. The only significant findings were that non-originality was a predictor of non-completion of first year studies and of a lower percentage outcome.

Introduction

The best practice guidance for managing academic misconduct as summarised by Carroll (Carroll & Appleton 2001, Carroll 2002) includes the recommendation that all students are explicitly educated at the outset of their higher education about what constitutes misconduct, why it is unacceptable and the institutional processes for managing it. Much of the general advice on how to accomplish this (Culwin & Lancaster 2001, Carroll 2002, Harris 2004) seems to confuse the issues of: educating students per se, detecting non-original content in student work, institutional management of plagiarism and academic misconduct, and designing assessment processes that are more resistant to misconduct.

The section of Carroll's good practice guide titled 'Active learning methods to teach students' (Carroll 2001 Ch 4.3) is only two paragraphs in length and offers little in the way of detailed suggestion, but recommends unspecified active learning techniques as most effective. The general advice given elsewhere regarding educating students seems largely to consist of informing students as they are inducted into higher education as to the issues involved and assuming that thereafter this knowledge will be understood and available. The availability of the information is guaranteed by the institution hosting it somewhere in its web presence, and advertising this location to the student body.

However the processes of acculturation into higher education need to embed values related to the ownership of intellectual property in order to foster attitudes that will respect the importance of correct attribution of non-original material. It seems unlikely that passive instruction and abstract discussion will embed such values into a 'napster generation' (Barbrook, 2002) who do not accept the legitimacy of intellectual property rights relating to popular music. This attitude can be expected to generalise to textual, and other forms of media, that are Web hosted. An attitude that is given some academic credibility by some postmodernist theorists (Ross 2004) who deny the concept of authorship and hence plagiarism. The resulting behaviour in students is evidenced by the extent of non-original content in submitted work (Dick et. al 2003) despite the now ubiquitous required student signature upon submission confirming that the material is original or correctly attributed.

More active approaches to educating students involve explicitly illustrating the acceptable and unacceptable ways in which non-original material can be included in student writing (Culwin 2005). Providing case studies of plagiarism in practice, quizzes and interactive on-line learning objects relating to plagiarism (Szondi 2004, Boling & Theodore Undated). Having the students obtain an essay from an essay bank and submitting it to a non-originality detection service (Weller 2003). However there seem to be very few such examples in the general literature.

Accordingly the first part of this paper describes an activity in which first year computing students at London South Bank University (LSBU)

participated during the 2002/3 and 2003/4 academic sessions. The activity was intended to provide a positive learning experience, allowing those who engaged in misconduct to be identified for educative rather than punitive attention. A by-product of these activities was direct measurement of the degree of non-originality in a corpus of student work. This data was subsequently demographically analysed in an attempt to obtain empirical evidence to support various conclusions from existing studies based upon less direct evidence.

The UK study that initiated the current interest in plagiarism and academic misconduct was conducted in 1995 by Franklyn-Stokes and Newstead (Franklyn-Stokes & Newstead 1995) and, significantly, was a questionnaire survey. As such it did not attempt to measure behaviour or the attributes of an artefact of behaviour; but to record self-reported behaviour. It can be argued that this would result in the true incidence being under-reported as individuals are to some extent practising self deception when they engage in misconduct.

Subsequently in order to preserve this deceit they will deny, even anonymously, that they have done so. Conversely it can be argued that males are more subject to anti social behaviour and bravado and so might over report the level of their misconduct.

Since that time the questionnaire study has remained the prevalent technique for assessing the level of misconduct in student cohorts (Marsden et al 2005). Some have asked students to report the level of behaviour in their cohort rather than their own (Sheard & Martin 2003) in the hope that this would yield more accurate data. Other studies

asked students to respond to hypothetical scenarios (Underwood 2003). While some asked tutors to report the amount of misconduct (Culwin et al 2002) although tutor's perceptions have been shown to be an under estimate (Franklyn-Stokes & Newstead 1995).

The reported rates of misconduct vary widely between studies (Stubblings & Brine 2003, Dick et al 2003, Hart 2004) which adds confusion to attempts to state if the amount of misconduct is increasing or decreasing. This confusion is caused by differences in the phrasing of the questions, differences in the implicit institutional or explicit questionnaire definitions of the terms used, differences in the way measurements were made (Likert, nominal, interval) and differences in the gross demographics of the cohorts studied (discipline area, gender balances, age profiles, level of study). Hence although it is often stated that the amount of misconduct is increasing there is very little, if any, wide ranging and longitudinal evidence that this is the case. What undoubtedly is the case is that the level of academic and stakeholder interest has increased dramatically over the last decade.

The studies have generally reported that the level of reported misconduct varies with gender, with males reporting higher levels than females; and age, with younger students reporting more misconduct than older students (Whitley et al 1999, Franklyn-Stokes. and Newstead 1995, Hart 2004, Marsden et al 2005, Simon et al. 2004). Other demographic factors are less clear-cut but, there is support for the suggestion that misconduct is associated with lower levels of ability and with students who have an instrumental attitude towards higher education (Hart 2004, McCabe & Trevino 1997).

The alternative to asking respondents to comment or report upon the level of misconduct is to make a direct measurement of non-originality within a corpus of collected student work. This method of measurement can be capricious in at least two ways. Firstly the reported level of non-originality will be lower than the true level. A study by Satterwhite & Gerein (Satterwhite & Gerein 2001) took samples from 146 different publicly available sources and submitted them to a number of different detection services and search engines. The various services returned between 20% and 58% hits, although the extent of detection that was classified as a hit is not clearly stated. Secondly the tools report only the extent of detected non-originality, this will include legitimately cited material as well as illicitly reused material. However given these caveats the levels of non-originality reported from such studies can be regarded as a low water mark for the true extent of non-originality in a corpus.

Other methodological problems relating to using direct measurement in this way include the stability of the tools used and the meaning of the values returned. Many of the tools are not stable and their performance is continually being upgraded in the light of experience. This together with the considerations that the tools have not been available for a significant amount of time and that some that once existed no longer appear to be available, makes longitudinal studies of cohorts or cohorts passing through an institution problematic.

In addition the meaning of the metric returned by the tool is not always obvious or comparable, or in some cases even explained. It might be assumed that a non-originality metric of 100% between two documents would indicate that they were word for word identical. However some tools, including the one described below, are fuzzy and a reported metric of 100% does not mean that the documents are necessarily totally identical. In a similar way a reported metric of 0% does not mean that the documents had no words or phrases in common. Accordingly direct comparisons between studies that use different tools, or even the same tool on different occasions, may not be possible.

Despite these problems there are a small number of studies that attempt to report upon direct measurement of behaviour as expressed in the proportion of detected non-originality in a submission. Many of these studies report upon computer program source code plagiarism as these tools have been available for a longer period of time and their operation is simpler than those that are able to measure free text non-originality (Chen et al 2002, Byrne et al 2004, Vamplew & Dermoudy 2005,). A smaller number of studies have investigated free text student submissions (Knight et al 2004, Johnson et al 2004, Weinstein & Dobkin 2002) and a related study by Clough (Clough 2001) investigated non-originality in a corpus of press articles.

With the exception of the Weinstein study the papers are largely more concerned with describing the operation, tuning and deployment of the tools and only incidentally report upon the extent of the non-originality detected and none relate it to any demographic factors. The Weinstein

study is notable as it manually filtered out false hits by visual inspection but again attempted no demographic analysis of the data obtained. It did however present objective evidence that explicitly warning students that their work would be subject to non-originality analysis caused the level of non-attributed non-original content to fall.

The Weinstein study reports the underlying Internet plagiarism rate at 17% and claims that this is comparable with other, uncited, studies. It divides these into 7% for 'large scale plagiarism' and 10% for 'plagiarism small in magnitude'. However, the percentage non-originality value, as provided by TurnItIn (TurnItIn 2005), upon which this distinction is made is not stated. The Knight study is largely concerned with optimising automated Google searches (Google 2005) and reported 4 possible hits from 480 submissions. The Johnson study used CopyCatch (CopyCatch 2005) and reported 4 plagiarised papers from a corpus of 590 submissions.

It is against this background that the results obtained from this study of direct measurement were related to demographic considerations in an attempt to triangulate the general conclusions of the questionnaire studies.

The Activity

The 2003/4 activity was designed following the experience in the 2002/3 session. This is a description of the latter activity with significant changes from the previous activity commented upon. The activity formed part of a level one professional skills unit taken by all first year

computing and IT students and which is perceived by the staff as an extended induction to higher education. A part of the unit involved a lecture on the history of computing, which was followed up with a homework task to write a related essay. A list of about 200 topics taken from the history of computing had been prepared, allowing each student to have an individualised task; which is also in line with good practice guidelines. The instructions for the task defined the outputs that the student had to produce which were: a title, some keywords, a 1000 word essay and two URLs where further information could be obtained. The instructions advised the students to use World Wide Web search engines to locate sources and included the explicit advice:

Although you are being encouraged to use the Web to search for information you ***MUST*** consult several sources of information and the essay you submit ***MUST*** be entirely in your own words drawing only facts and ideas from your sources.

The students were allowed two weeks to complete the activity and were advised that they could contact the activities' co-ordinator if they could not locate any suitable material. A small number of students made use of this support, most of whom needed assistance with constructing effective search terms but some of whom had been given a topic that was too obscure and needed to have a different topic allocated.

The students were required to submit their essays and other outputs using a Web form. Their essays were then made visible on the University Intranet, indexed in various ways and students were

encouraged to read the essays that other members of their tutor group had produced.

Once the deadline for submission had passed the essays were analysed for non-originality using a suite of programs based upon OrCheck (Lancaster & Culwin 2004) technology. Essentially this technology allows automated Google searches to be made, downloads the documents indicated by the search engine and analyses the degree of similarity between each downloaded document and the document under investigation. The results of the analysis include a visualisation that shows which parts of each located document have been shown to be similar to parts of the document under investigation. Copies of the target and located documents with the similarities highlighted in different colours are also produced.

The results of the analysis were communicated to each student by means of a pro-forma report which included the address of a hidden URL, where the student's submission and the most similar document were shown side by side with the essentially identical segments highlighted in red. An example of the web page produced for each student is illustrated in Figure 1.

These results were made available to the students at the first tutorial following a lecture on academic misconduct. That lecture illustrated the process of academic misconduct by the lecturer pretending that they had been required to write a 1000 essay on a historical figure called "Tony Blair". The first essay was prepared by downloading the biography of the prime minister from the official Number 10 web site, measuring about 1000 words in Microsoft Word and inserting the name

of the lecturer as the author. This essay was then analysed live by the OrCheck tool with the source readily identified. A number of other pre-prepared essays were then produced exhibiting different attempts at disguise with each being submitted live to the tool with the sources again being readily identified and the visualisation showing exactly how each source had been used to produce the finished essay.

At this point the students were told that their essays had been processed by the tool and that their tutors would be giving them the results of the analysis and discussing the implications with them at the next tutorial. It was also readily admitted that the tool was not foolproof and it was quite possible that the headline reported percentage non-originality value was lower, or even much lower, than the true degree of non-originality. The lecture then continued by explaining what other activities constituted academic misconduct, how and why these activities damaged both the individual and, via the damage to the institution, damaged everyone. The lecture concluded by outlining the processes that would be followed if academic misconduct was suspected by a tutor.

Figure 1 On-line non-originality report

The screenshot shows a web browser window titled "PSK History Report". The main content area displays the following information:

- Gross non-Originality 49.2%**
- The first URL you supplied contributed 43.7% of your article**
- The second URL you supplied contributed 24.3% of your article**
- 8.4% of <http://ftp.arl.mil/~mike/comphist/61ordnance/chap4.html> was used as 43.7% of this submission.

Below this, a note states: "This is an orcheck image of your file, the colours indicate non-originality." This is followed by a small image showing a document with colored bars indicating non-original content.

The bottom section of the report lists several ballistic problems that ORDVAC is designed to solve:

Here are a few things that ORDVAC was designed to do e.g. solve the following problems:-

1. Exterior ballistics problems like high altitudes, solar and lunar path, and guidance control data for Ordnance weapons, including guided missiles.
2. Interior ballistic problems, propellant and launcher behaviour, e.g. properties of rocket propellants, computation of detonation waves for reflected shock waves, vibration of gun barrels and the flow of fluids in porous media.
3. Terminal ballistic problems, including nuclear, fragmentation effects in such areas as explosion kinetics, shaped charge behaviour, and heat transfer.
4. Ballistic measurement problems such as photogram metric, ionosphere, and damping of satellite spin calculations, reduction of satellite Doppler tracking data.

On the right side of the window, there is a list of specific ballistic problems:

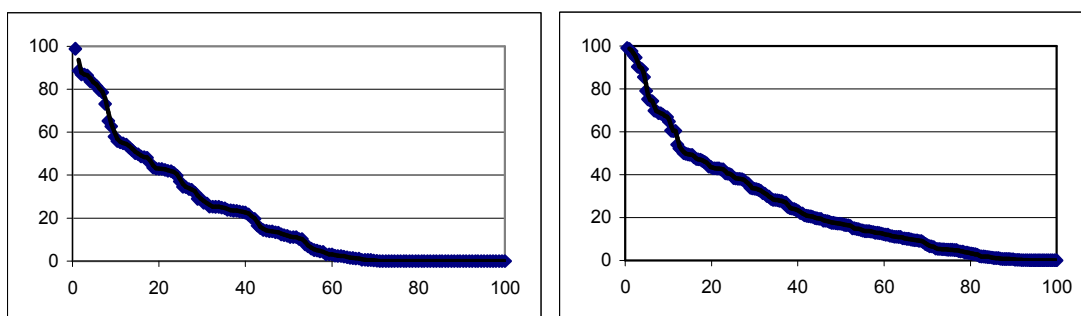
1. Exterior ballistics problems such as high altitudes, solar and lunar trajectories, computation for the preparation of firing tables and guidance control data for Ordnance weapons, including free flight and guided missiles.
2. Interior ballistic problems, including projectile, propellant and launcher behavior, e.g., physical characteristics of solid propellants, equilibrium composition and thermodynamic properties of rocket propellants, computation of detonation waves for reflected shock waves, vibration of gun barrels and the flow of fluids in porous media.
3. Terminal ballistic problems, including nuclear, fragmentation and penetration effects in such areas as explosion kinetics, shaped charge behavior, ignition, and heat transfer.

There was no formal evaluation of the activity as such by the students. However most of the tutors commented upon how amused and engaged the students were, even by those who had unseemly amounts of non-originality detected. A follow-up activity was planned which involved a structured interview with a sample of the students, with the sample biased towards those who had shown large amounts of non-originality. However due to a combination of staff illness and administrative confusion this activity did not take place.

The gross measurements

Figure 2 shows the cumulative non-originality graphs of the amount of detected non-originality for the 145 students who participated in the 2002/3 activity and the 207 who participated in 2003/4.

Figure 2 Cumulative non-originality 2002/3 left and 2003/4 right



On each graph the x axis is the percentage of students who had at least the degree of non-originality shown on the y axis. For example in 2002/3 20% of students had approximately 40% or more non-originality in their essays. A small adjustment was made in the process of

collecting and analysing for 2003/4 which had the effect of making the process a little more sensitive at the lower end of the scale. Allowing for this the graphs can be considered very similar. The mean in 2002/3 was 21.2% with a large standard deviation of 25.6%; in 2003/4 the mean was 24.7% and coincidentally the standard deviation was also 24.7%. A t test confirmed that there was no significant difference between the data from the two years.

When percentage non-originality values are stated in studies it is not always clear how the value was obtained. The meaning of 100% non-originality seems intuitively clear, all of the document can be shown to be identical with parts of one or more other documents. The meaning of 0% is less clear, naively it might be taken to indicate that there is no similarity between the document under investigation and any other documents considered in the analysis. However as all of the documents are written in the same natural language there must be some overlap in common vocabulary, even before any technical vocabulary issues are considered. Values intermediate between 0% and 100% are correspondingly even less clear.

The values reported here are computed by first removing the 200 most common words in everyday English as reported by the British National Corpus (BNC 2005) from all documents. The first word in the resulting target document is then considered and the longest sequential matching sequence of words starting with that word in any other document is located. If that word does not exist in any other document it scores 0 and the next word is considered. If the longest matching string is two words, each word scores 0.25; three words score 0.5; four

words 0.75 and five words or more 1.0. After scoring the sequence the first word following it is considered. This process continues until all the words in the target document have been scored.

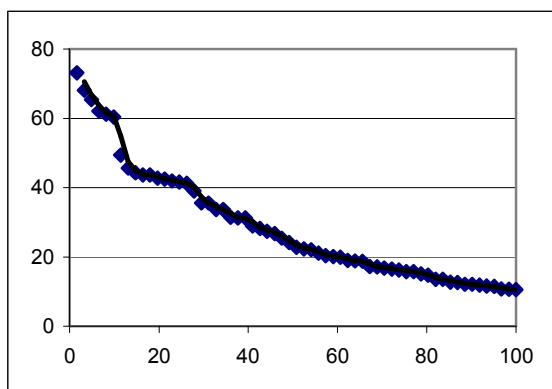
Accordingly each word in the document can score between 0.0 and 1.0 with the gross reported percentage being the average score of all non-common words in the document. The scoring system will not be unduly influenced by a common technical vocabulary as terms or short phrases, score lower than longer phrases or sections of matching text.

As the 200 most common words are removed before scoring it is possible that a score of 100% could be obtained between the document under investigation and a source or sources, despite them not being word for word identical. This fuzzy matching of documents is capable of seeing through some attempted disguise of the source material by the plagiarist.

What was somewhat surprising was the number of students who submitted a URL which subsequently proved to be the most significant match. A total of 136 of the 207 students were shown to have more than 10% non originality, a fairly arbitrary cut off value to exclude the possibility of a common technical vocabulary and other non-significant matching phrases. Of these 136, 61 that is approximately half, submitted a URL that proved to be the greatest contributor of non-originality. The average non-originality for these contributions was

28.8% with a standard deviation of 16.4, the minimum was 11% and the maximum 73%. The cumulative frequency graph for this group is shown in Figure 3.

Figure 3 Cumulative non-originality attributable to a supplied URL



The graph indicates that about 40% of this group, about 54 students, supplied a URL which contributed 25% or more of their essay; with about 10% of the group, about 14 students, supplying a URL which contributed 60% or more. The implications of this are either that the students did not think that anyone would read their essays in conjunction with the link supplied and note the similarity or that, despite the clear injunction, they did not believe that this behaviour was unacceptable.

The Demographic Analysis

Accepting the limitations of the measurement process the gross degree of non-originality as observed can be used to investigate demographic and other divisions within the sample. As the terms were randomly

distributed through the sample the efficacy of the non-originality measurement can be assumed to be non-capricious with respect demographic and other variables. Consequently any observed differences between sub-groups within the sample can be safely attributed to membership of that group, rather than being an artefact of the measuring process.

Gender

Table 1 shows the overall mean level of detected non-originality for males and females. The left hand table sows the data from the 2002/3 session and the right hand table the data from the 2003/4 session. The first row of the table shows the number of males and females, the second row shows the level of non-originality for the entire sample and the third row the level of non-originality excluding those who had less than 10% non-original material detected.

Table 1. Analysis by gender 2002/3 left and 2003/4 right

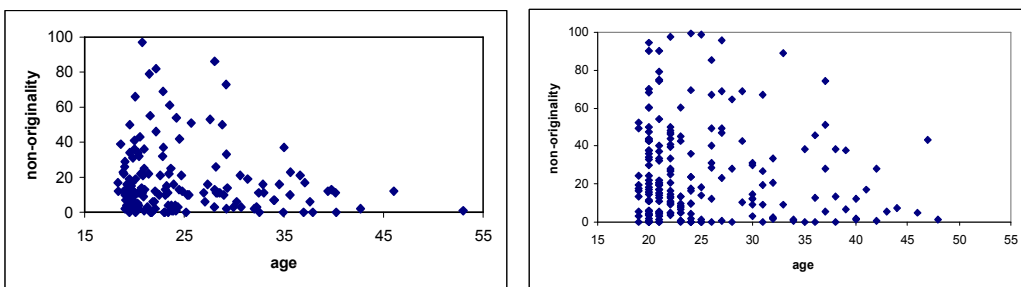
	female	male		female	male
n	36	119		67	138
all	15.1	17.6		22.3	26.2
10 plus	17.1	20.5		35.3	36.7

Standard socio-biological theory states that as females are more risk averse and general educational studies suggests that females are more committed and organized. Both of these factors might suggest that male non-originality would be higher than female non-originality, and this is indicated by the data. However the difference was not shown to be significant.

Age

Figure 4 shows the relationship between age and non-originality. There is no clear pattern in the graphs and the correlation coefficients indicated no systematic relationship.

Figure 4 Scattergraph of age against non-originality 2002/3 left and 2003/4 right



When the sample was divided into those younger than 22 and those older than 22, a t test showed no significant differences. The value 22 was chosen as is about the minimum age for individuals to graduate in UK universities. As with gender it might be hypothesised that older students would be less likely to engage in misconduct and although this is indicated by the raw data it cannot be concluded. The data is presented in Table 2.

Table 2. Analysis by age 2002/3 left and 2003/4 right

	<22	>=22		<22	>=22
n	83	73		101	103
all	18.9	15.4		27.0	22.8
10 plus	28.1	23.4		34.9	38.8

Outcome

Outcome can be measured in two ways. The students submitted the essay in November with the end of the academic year the following June. Some students withdrew before completing the year and so would have an incomplete set of results. The remaining students' outcome could be measured by taking the overall percentage score as recorded on their first year transcript.

Table 3. Analysis by completion 2002/3 left and 2003/4 right

	complete	withdrawn	complete	withdrawn
n	129	27	182	25
all	16.3	21.9	23.8	31.3
10 plus	24.3	34.2	34.8	45.0

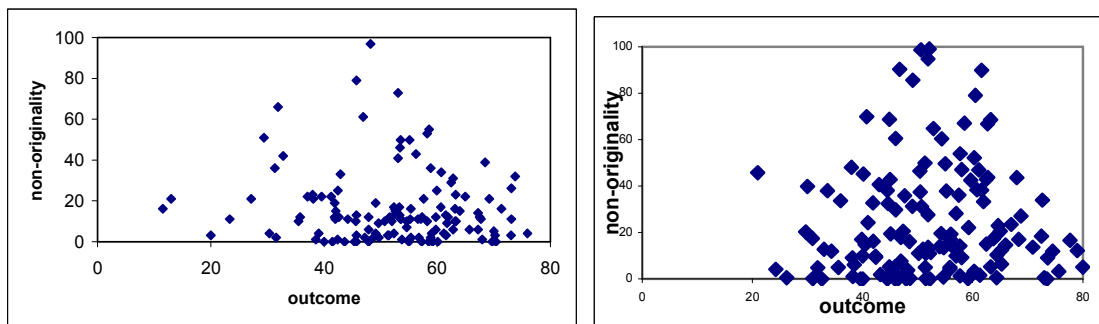
Table 3 shows the average non-originality scores of the minority of students who did not complete their first year of study and of those who did. A t test on the entire cohorts showed the difference to be significant 0.1 confidence level for the whole groups in both years (2002/3 $p=0.083$, 2003/4 $p=0.078$). When the analysis was restricted to those with more than 10% non-originality the differences for both years were significant at the 0.05 confidence level (2002/3 $p=0.033$, 2003/4 $p=0.047$).

The data suggests that the amount of non-originality is indicative of not completing the year. That is those first year students who submitted higher levels of non-original material in November were significantly less likely to present a full set of marks to the examination board the following July. This result does not seem surprising as students who have to resort to cheating so early in their higher education academic

progression would seem to be ill prepared for the demands of studying at this level. It might also be that they are less committed to studying and so are more likely to withdraw.

Figure 5 presents scattergraphs of the students who completed the year against non-originality. There is no clear pattern in the graphs and the correlation coefficient indicated no systematic relationship.

Figure 5. Scattergraph of outcome against non-originality



When the sample was divided into those whose outcome was less than 40% and those above 40%. The 2002/3 averages were 20.4% and 15.5% respectively, a difference which was not significant. The 2003/4 averages were 25.9 and 15.8, a difference which was shown by a t-test to be significant at the 0.05 level ($p=0.042$). The value of 40% was chosen as this is regarded as the pass/fail boundary at LSBU.

Following on from the non-completion argument it would appear that there is some evidence that weaker students are more prone to engage in misconduct.

Course

The students in the sample came from one of two programmes: Computing, which includes both computing and internet computing and Business Information Technology (BIT), which includes both BIT and e-commerce. The computing programme is regarded as the more technical and students entering it tend to have a higher level of qualification. Table 4 shows the non-originality related to course, the number of students in this analysis is smaller than that in previous tables as the course data for a number of students was clearly erroneous. A t test on the 2002/3 data showed significance at the 0.05 level, both for the entire sample and for those with more than 10% non-originality ($t=1.74$, $t\text{ crit}=1.66$, $p=0.042$ & $t=1.89$, $t\text{ crit}=1.66$, $p=0.031$). The data from the 2003/4 cohort was non-significant.

Table 4. Analysis by course 2002/3 left and 2003/4 right

	Computing	BIT		Computing	BIT
n	64	84		84	115
all	14.0	19.5		25.9	22.9
ten plus	21.9	29.4		33.9	37.6

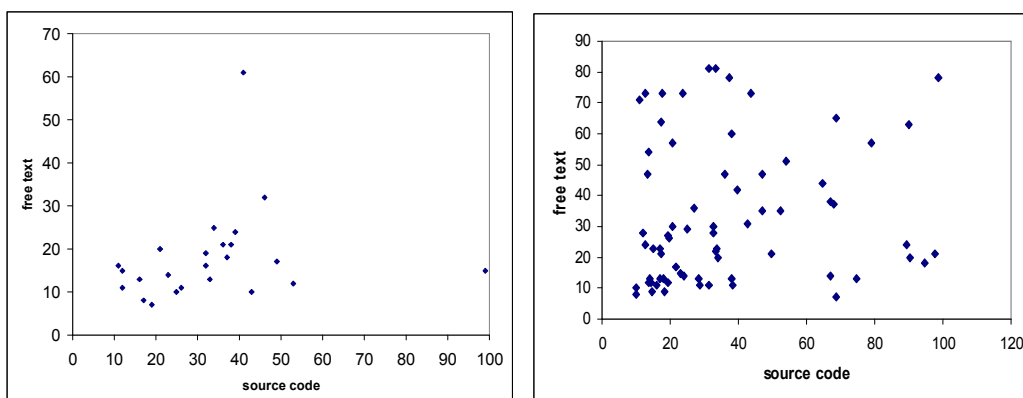
A tentative conclusion from the 2002/3 data might be that there is some evidence that less technical courses are associated with higher

levels of misconduct. This conclusion would be in keeping with the amount of plagiarism noted on final year projects that are organised and administered identically across both programmes. However it was not possible to triangulate this conclusion from the 2003/4 data.

Triangulation with MOSS

In the 2002/3 session the computing students were required to complete a Java programming coursework; in 2003/4 all students were required to do so. The program listings were collected in both years and submitted to the Measure Of Software Similarity (MOSS 2005) service, which measures the extent of intra-corporal similarity in the source code listings. This provides an independent measure of non-original material and so can be triangulated with the data from the essay activity. The scattergraphs of the 24 students from the 2002/3 cohort and the 69 students from the 2003/4 cohort, whose free text non-originality was greater than 10% and for whom MOSS measures were available are given in Figure 5.

Figure 5. Scattergraph of source code against free text non-originality, 2002/3 left and 2003/4 right



Although the correlation from the 2002/3 data was significant at the 0.01 level, the correlation from the 2003/4 data was non-significant. The implication from the significant correlation would have been that presenting non-original work appears to be consistent between subjects and modalities, however this was not confirmed from the larger 2003/4 sample.

Conclusions and further work

This paper gives an example of how first year undergraduate students can be actively introduced to the issues of academic misconduct. Although no formal evaluation of the activity was conducted the reports from the tutors and from the students themselves would seem to indicate that this activity was more favourably received than a more passive exposition which was presented in previous years.

The study also shows that it is possible to systematically measure the amount of non-originality in student submissions and from that investigate the demographics of academic misconduct. The relatively short length of the student submissions in this study and the somewhat artificial nature of the task may have affected the outcome. It is also possible that an institutional memory is in the process of being developed whereby first year students are pre-warned that the history essays will be subject to non-originality investigation. If this is the case then the amount of non-originality contained within the submissions would decline; however this is exactly what is wanted from a programme to educate students about academic misconduct.

Several factors which have been suggested as possibly significant, were indicated but were not shown to be significant. The only conclusion that is as yet unambiguously concluded is that students who subsequently did not complete their first year were more likely to present non-original material. This is partly coupled with the conclusion that weaker students are significantly more likely to engage in cheating. If these findings can be shown to be replicable then it would indicate that activities such as this can be used to identify students who might benefit from additional support. It would be worthwhile for the study to be repeated in subject areas other than computing and in institutions other than LSBU. Not only would this further triangulate the data, if common measurement techniques were applied, would allow the demographic influences of broad subject area and type of institution to be investigated.

References

Barbrook, R. (2002), *The Napsterization of Everything*, *Science as Culture*, 11(2) 277-285.

Boling Elizabeth & Frick Theodore (undated), How to Recognize Plagiarism, Indiana University, School of Education, available at <http://www.w.indiana.edu/~istd/practice.html> (accessed 1 April 2005).

BNC (2005), *British National Corpus*, available at www.natcorp.ox.ac.uk (accessed 1 April 2005).

Byrne Siobhán, Charles Daly, Jane Horgan & John Waldron (2004), A technique for measuring plagiarism in automated & distance learning environments, Proceedings of the winter international symposium on Information and communication technologies, Mexico 2004

Carroll Jude & Appleton Jon (2001), *Plagiarism A Good Practice Guide*, Report commissioned by the Joint Information Systems Committee (JISC) available at http://online.northumbria.ac.uk/faculties/art/information_studies/lmri/Jiscpas/docs/brookes/brookes.pdf (accessed 1 April 2005).

Carroll Jude (2002), *A Handbook for Deterring Plagiarism in Higher Education*, Oxford Brookes University, ISBN 1873576560

Chen Xin, Li Ming, Mckinnon Brian & Seke Amit (2002), A Theory of Uncheatable Program Plagiarism Detection and Its Practical Implementation, available at www.cs.ucsb.edu/~mli/sid.ps, (accessed 1 April 2005).

Clough Paul (2001), *Measuring text reuse in a journalistic domain*, Proceedings of the 4th Annual CLUK Colloquium, Sheffield, UK, 53-63

CopyCatch (2005), <http://www.copycatchgold.com>, (accessed 1 April 2005).

Culwin, Fintan & Lancaster Thomas (2001), Plagiarism, prevention, deterrence and detection, report commissioned by the Institute for Learning & Teaching (ILT) available at http://www.heacademy.ac.uk/embedded_object.asp?id=21635 (accessed 1 April 2005).

Culwin, F., MacLeod, A. & Lancaster, T. (2002), *Source Code Plagiarism in UK HE Computing Schools: Issues, Attitudes and Tools*, South Bank University Technical Report SBU-CISM-01-02.

Culwin Fintan (2005), PlagPoster available at <http://cise.lsbu.ac.uk/plagposter> (accessed 1 April 2005).

Dick Martin, Sheard Judy, Bareiss Cathy, Carter Janet, Joyce Donald, Harding Trevor, Laxer Gary (2003), *Addressing student cheating: definitions and solutions*, ACM SIGCSE Bulletin 35(2) 172-184

Franklyn-Stokes, A. and Newstead, S. (1995), *Undergraduate cheating: who does what and why?*, Studies in Higher Education, 20(2), 159-172.

Goolge (2005), <http://www.google.com/apis/> (accessed 1 April 2005).

Harris Robert (2004), *Anti-Plagiarism Strategies for Research Papers*, available at <http://www.virtualsalt.com/antiplag.htm> (accessed 1 April 2005).

Hart Mike (2004), Plagiarism and Poor Academic Practice – A Threat to the Extension of e-Learning in Higher Education?, Electronic Journal

of E-Learning (EJEL) 2 (2), available at <http://www.ejel.org/volume-2/vol2-issue1/issue1-art25.htm> 2004 (accessed 1 April 2005).

Johnson D, Patton R, Bimber B, Almeroth K & Michaels G (2004), *Technology and Plagiarism in the University: Brief Report of a Trial in Detecting Cheating*, Association for the Advancement of Computing in Education (AACE) Journal, 12(3) 281-299.

Knight Alan, Almeroth Kevin & Bimber Bruce (2004), *An Automated System for Plagiarism Detection Using the Internet*, World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED MEDIA), Switzerland 2004.

Lancaster Thomas & Culwin Fintan, *Using freely available tools to produce a partially automated plagiarism detection process*, Proc ASCILITE 2004 pp520-529, available at <http://www.ascilite.org.au/conferences/perth04/procs/pdf/lancaster.pdf> (accessed 1 April 2005).

McCabe Donald L. & Trevino Linda Klebe (1997), *Individual and Contextual Influences on Academic Dishonesty: A Multicampus Investigation*, Research in Higher Education, 38(3) 379 - 396

Marsden Helen, Carroll Marie & Neill James T, *Who cheats at university? A self-report study of dishonest academic behaviours in a sample of Australian university students*, Australian Journal of Psychology, 57(1) 1-10

MOSS (2005), *Measure Of Software Similarity*, available at <http://www.cs.berkeley.edu/~aiken/moss.html> (accessed 1 April 2005).

Ross, George MacDonald (2004), *Plagiarism really is a crime: a counterblast against anarchists and postmodernists*, Proc. Plagiarism:

Prevention, Practice & Policy conference, Northumbria University Press, 69-70

Satterwhite, Robin & Marla Gerein. (2001) *Downloading detectives: searching for on-line plagiarism*, available at <http://www.coloradocollege.edu/library/course/webplag2.html> (accessed 1 April 2005).

Sheard Judy & Dick Martin, *Influences on cheating practice of graduate students in IT courses: what are the factors?*, ACM SIGCSE Bulletin, 35(3) 45-49

Simon Christopher A., Carr Jim R., Mccullough Sesi M., Morgan Sally Oleson J., Ted & Ressel Maggie (2004), *Gender, student perceptions, institutional commitments and academic dishonesty: who reports in academic dishonesty cases?*, Assessment & Evaluation in Higher Education 29(1) 75-90

Stubbings Ruth & Brine Alan (2003), *Plagiarism and the role of the library*, Update (magazine of the Chartered Institute of Library and Information Professionals) Issue December 2003 available at <http://www.cilip.org.uk/publications/updatemagazine/archive/archive2003/december/update0312c.htm> (accessed 1 April 2005).

Szondi Juliet & Martindale Chris (2004), *PLATO: PLAGiarism Teaching Online* available at <http://www.derby.ac.uk/focus-on/archive/internal.asp> (accessed 1 April 2005).

Turnitin, <http://www.turnitin.com> (accessed 1 April 2005).

Underwood Jean D. M. (2003), *Student attitudes towards socially acceptable and unacceptable group working practices*, British Journal of Psychology 94(3), 319-317

Vamplew, P.W. & Dermoudy, J.(2005), *An Anti-Plagiarism Editor for Software Development Courses*, Proceedings of the Australasian Computing Education Conference, Newcastle, Australia

Weinstein J.W. & Dobkin CE, *Plagiarism in US Higher Education: Estimating Internet Plagiarism Rates and Testing a Means of Deterrence*, available at <http://webdisk.berkeley.edu/~Weinstein/Weinstein-JobMarketPaper.pdf> (accessed 1 April 2005).

Weller Martin J (2003), Suggested activities for students, Open University, available at http://online.northumbria.ac.uk/faculties/art/information_studies/lmri/Jiscpas/site/adv_academicprac.asp (accessed 1 April 2005).

Whitley Bernard E., Bichlmeier-Nelson Amanda and Jones Curtis J. (1999) Gender Differences in Cheating Attitudes and Classroom Cheating Behavior: A Meta-Analysis, *Sex Roles* 41(9) 657 - 680