Student Attitudes to Plagiarism and Collusion within Computer Science

Louise A. Dennis
School of Computer Science and Information Technology
University of Nottingham, Jubilee Campus, Wollaton Road, Nottingham NG8 1BB
ladcs.nott.ac.uk 0115 846 6568

December 7, 2004

This research was undertaken as part of the Postgraduate Certificate in Higher Education run by the University of Nottingham. Thanks go to Dr. Gordon Joyes for his encouragement and for work on a early abstract describing the results.

Abstract

There is a widespread perception among staff in Computer Science that plagiarism is a major problem particularly in the form of collusion in programming exercises. While departments often make use of electronic detection measures, the time consumed prosecuting plagiarism offences remains a problem. As a result departments continue to seek ways to reduce the amount of plagiarism and collusion that occurs.

This paper reports the findings of a questionnaire based study which attempted to assess the students' attitudes to the issues involved in the hope that such an understanding might result in practical measures for minimizing the problem.

The study revealed that while students did understand the definition of plagiarism in its most extreme cases they were often confused about less clear-cut situations. Changes in the previous experience of incoming students meeting modules originally designed on the assumption that students already had some programming background and were equipped for self-directed study would also appear to be a contributory factor in the extent of collusion in programming exercises.

Introduction

This paper reports an investigation into student attitudes to plagiarism and collusion in a Computer Science department in a pre-1992 university. The department has a large undergraduate body (approximately 300 students in each year of study) being taught in an environment which encourages computer use in all its forms. Teaching within the department is based around the lecture as the primary mode for delivering information (although there are some individual modules which take alternative approaches) backed up with widespread use of the web and the provision of supervised programming lab sessions and tutorials for specific modules. This means that while students have the opportunity to ask for individual assistance with specific problems they are primarily expected to work in a self-directed fashion within a large-group teaching environment.

The perception among staff is that plagiarism is a major problem and that it particularly manifests itself in the form of collusion in programming exercises. This is a more specialised example of the general problem of plagiarism that manifests itself across the HE sector. CS staff are particularly concerned about students either copying from each other or working together as a group to produce solutions. The issue of plagiarism in its more usual form tends to arise primarily in the form of project dissertations submitted in the final year where background material may be copied from external sources (specifically the world wide web).

The department makes extensive use of electronic plagiarism detection tools but would prefer to prevent the problem arising in the first place. While there is a large literature on devising assessments to minimise plagiarism (e.g. [1]) the recommendations are not always easy to implement in the context of the department: large class sizes make it difficult to individualise tasks set to students; in the case of programming, in particular, staff actually wish to assess

students' ability to produce a working computer program (rather than their ability to design an algorithm or understand some concept – although these are also learning outcomes in which staff are interested) . In fact many staff consider the ability to program to be a fundamental skill all graduates should possess so alternative forms of assessment and group working can not outweigh assessments designed to test a student's ability to produce a working program on their own. At present the department is investigating the use of programming examinations. Clearly these suffer from many of the drawbacks of more traditional exams and also reduce the number of non-examination based assessment methods used by the department.

In theory the situation should be much better in the case of project work. Each student has an individual task, performed in contact with a lecturer who is aware of the relevant background literature and so is likely to detect plagiarism from external sources. In practice a divergence in student interests (primarily a large subsection of the student community who, for in some cases sound professional reasons, wish to produce a functioning e-Commerce website as their project) and staff research interests means that staff are frequently supervising a number of projects outside of their area of expertise and are therefore ill-equipped to detect plagiarism of background literature. It has been suggested that increasing the number of deliverables for projects to minimise the amount of writing up done in a rush at the end would reduce the problem but this would be at the expense of increasing the marking load on members of staff who already feel overwhelmed by the amount of project supervision required from them. This is currently unpopular as a solution to the plagiarism problem.

This study was an attempt to gain some understanding of the student attitudes to plagiarism in order to see if there were any unexplored avenues that could be developed which would reduce the problem. It aimed to address several specific questions[1].

1. Did the students understand what constituted plagiarism or collusion? a frequent defence offered by culprits being that they did not understand their behaviour was wrong.

2. Is the incidence of plagiarism as high as staff fear.

3. Why do students plagiarize?

It was hoped that the answers to these questions would help clarify the way forward.

Methodology

Two questionnaires were distributed to students. The first questionnaire focused specifically on collusion in programming exercises and was distributed to first year undergraduates while the second looked at the issue of plagiarism in general and was distributed to second and third year undergraduates and MSc students. In both cases there were approximately 80 responses.

The questionnaires consisted primarily of a list of scenarios. Students were asked to indicate in each case whether they thought it fair for a student to lose marks for working in such a way (this caused some problems in interpreting the data – it wasn't clear that deduction of marks necessarily implied that cheating had taken place), whether they thought the scenario counted as plagiarism and whether they had ever behaved as outlined in the scenario themselves. Initial interest was focused on identifying students who considered that they had cheated rather than in attempting to define which scenarios counted as plagiarism. There were two reasons for this focus, firstly staff believed there was widespread and deliberate cheating and secondly plagiarism is notoriously hard to define and this avoided the need to classify individual scenarios as plagiarism or otherwise. The questionnaires also asked students to estimate the amount of plagiarism among their peer group and to choose the three reasons why they thought people plagiarised – the results of analysing this data are naturally suspect since students were not asked directly why they had plagiarised because of concerns that they would not be prepared to admit to cheating even on an anonymous questionnaire.

Lastly it should be noted when interpreting the data that the questionnaires were only distributed to students attending lectures and in each case attendance was less than half what it theoretically would be should every student registered for the module have attended. It can be assumed therefore that the sample is biased towards more conscientious and hard-working students who, it could be argued, are less prone to plagiarise. As a results estimates of the amount of plagiarism taking place arising from this data are likely to be conservative.

[1]

Results

Collusion Questionnaire

This questionnaire was distributed in an first year lecture (although a couple of second years were taking the module as an option). 80 students responded out of a possible 210 taking the module. The number attending the lecture was approximately 80.

The scenarios the students were asked to comment upon were as follows:

1. Two students discuss the coursework together and produce a joint design for the program (e.g. how information will be passed and when loops will be needed). They then each work separately to produce a final program.
2. Two students work together on large sections of a coursework e.g. certain methods. These are the same in both of their submissions though there are differences elsewhere.
3. A student copies another student's program with their knowledge and submits it as their own.
4. A student is told by another student how some feature of the language works (e.g. how to read files in Java). They use this within their own program which is otherwise entirely their own work.
5. Two students work together on a coursework and both submit the same program.
6. A student copies several methods from another student but makes a note as a comment in their program that they have done so.
7. A student copies another student's program with their knowledge making small changes, such as the naming of variables and submits it as their own.

Students were confused by the options presented when assessing the scenarios many thinking that they had to check one box (It is fair if the student lost marks, I think this counts as plagiarism, I have done this myself) and only one box. As a result, in the analysis, ticks in the first two boxes have been counted together simply as a measure that students are aware the activity is in some sense considered wrong. This problem was rectified in the plagiarism questionnaire by putting "yes/no" at the top of each box indicated that the students should fill them all in.

While it is difficult to agree a precise definition of collusion most staff members would consider the activities outlined in scenarios 2, 3, 5 and 7 as ones that should be penalised in individual work. These were classified as examples of collusion in the analysis. In cases 3, 5 and 7 over 90% of students indicated they considered it wrong in some fashion. However only 75% of students considered it wrong to share methods. These numbers are pleasingly high suggesting that the majority of student do agree with staff over definitions of collusion. However while 75% of students thought sharing methods was wrong in some sense 16% admitted to doing this and none of these indicated they thought it wrong in any way. This means approximately 1 in 6 computer science students have, at some point in their first year of study, committed a collusion offence without apparently being aware that they are doing anything wrong.

Only 10% of students actually confessed to knowingly participating in cheating behaviour which is considerably lower than the staff perceptions of collusion. As discussed the disagreement between staff and students over whether it is acceptable to share large portions of code goes some way to explaining this. Furthermore, since the majority of the respondents were only half way through their first year of study it may be that the problem increases during the course of their studies (backed up by an increase to 25% of students confessing to cheating in 2nd and 3rd year). Lastly the confusion over whether all boxes should be ticked or only one may have lowered the reporting rate.

Setting aside the question of deliberate cheating 24% of students admitted to scenarios 2, 3, 5 and 7 which were those classified as examples of collusion. This is much closer to the estimates produced in conversation with staff and is clearly a cause for concern.

Student estimates of the extent of the problem show that the general perception is that incidence of plagiarism is fairly but not very low this would tie in with an estimate of between 10% and 25% cheating.

The most popular reason given for collusion was inability to do the coursework (80%) followed by lack of time (70%) and inability to keep up (63%). Reasons which indicate contempt for the assessment process and learning in general such as getting higher marks (30%) and not needing to know the material (19%) ranked rather lower.

The actual ranking was:

1. They simply couldn't do the coursework on their own.
2. They started too late and ran out of time.
3. They couldn't keep up with the work.
4. They have to succeed. They got higher marks this way.
5. They didn't need to learn that material, just pass the module.
6. They felt the tutor didn't care, so why should they.
7. They didn't think it was wrong.
8. We are encouraged to work together.
9. They wanted to see if they could get away with it.

The reasons chosen were taken from [3] who reported they were from [2]. These suggestions missed out straightforward laziness which 6% of respondents identified in the "Other" option – this would rank it 9th in the above list.

An open-ended question asking for input revealed a lot of dissatisfaction with the methods of teaching programming in first year from "not enough support – some tutors aren't as good as others and make you feel thick." through "In a case where not all the material in a coursework are covered by the lecturer in lectures, what do you do." to "the lecturers assume that students already know things like a bit of programming." In all out of 80 responses 9 people chose spontaneously to complain about either lack of support for or the pace of programming modules. This is 10% of responses. This can be interpreted in two ways. 10 years ago a typical incoming computer science undergraduate already had some experience of programming (often as a hobbyist) and at least had an expectation that programming would form a significant part of their studies whereas now many incoming students have no experience of programming, or experience that is limited to interfaces that remove much of the complexity of the task. Similarly their expectations are based on their experience in using of computer programs (particularly web browsers) rather than the development of computer programs and they no longer necessarily consider the development of programs as an important part of their studies . This interpretation would suggest that the department needs to better motivate the programming modules and revise its assumptions about the ability of the students. The other interpretation is that the swing towards coursework at GCSE and A Level has paradoxically produced an expectation that a student will receive a high degree of detailed help (from teachers, peers or parents) when performing exercises and reduced their willingness to seek help for themselves from textbooks or even simple reflection upon the problems they encounter (lab demonstrators frequently complain that students put their hands up before taking the basic steps required to diagnose a problem for themselves). Under this interpretation greater attention needs to be placed on improving the study skills of incoming undergraduates and inculcating them into a culture where they are expected to take greater responsibility for their own learning.

Plagiarism Questionnaire

This questionnaire was distributed in a module open to second and third year undergraduates and MSc students. 80 responded out of a possible 280. The member of staff involved estimated there were approximately 140 students present in the lecture theatre when the questionnaire was distributed.

The scenarios presented to the students were based on an exercise in [5]:

1. A student composes a paragraph by taking short phrases of 10 to 15 words from a number of sources and putting them together, adding words of their own to make a coherent whole; all sources appear in a list of references at the end of the coursework.
2. A student copies a paragraph from a textbook or web page making small changes – e.g. replacing a few verbs, replacing an adjective with a synonym. The source appears in a list of references at the end of the coursework
3. A student paraphrases a paragraph with substantial changes in language and organisation; the new version also has changes in the amount of detail used. The source material is acknowledged in the text e.g. (Jones, 1999) and included in a list of references at the end of the coursework

4. A student composes a paragraph by using sentences of an original source but omitting one or two and putting one or two in a different order. The source appears in a list of references at the end of the coursework

5. A student quotes a paragraph by placing it in italic font and/or using quotation marks with the source cited in the text and included in a list of references at the end of the coursework.

6. A student copies a paragraph from a textbook or web page that they did not write this paragraph.

Scenarios 1, 2, 4 and 6 were classified as plagiarism for the purposes of analysis although staff might disagree on the need to take disciplinary action over individual scenarios. However all are to be discouraged in academic writing despite the referencing of sources at the end in several cases. 93% of students recognised that case 6 represented plagiarism and 95% thought students should lose marks for it. However, in all other cases students were less clear – in Case 2 and Case 4 only 49% and 56% respectively thought it was plagiarism with 59% and 64% thinking marks should be deducted. In case 1 only 22% thought it constituted plagiarism and only 28% thought marks should be deducted. These results suggest two things: the fact that in all cases more students considered it fair to deduct marks for the behaviour than considered it plagiarism suggests that a stress should be placed on not cheating over not plagiarising. It is also clear that we need to stress that even slight modifications of texts still count as plagiarism. Altogether 25% of respondents admitted to indulging in behaviour they thought would be considered wrong with 8% admitting to direct plagiarism (case 6). Worryingly 65% admitted to at least one of cases 1, 2, 4 and 6 again representing a mismatch in staff perceptions of misconduct and student perceptions.

Most students (36%) thought less that 10% of their peers plagiarised followed by between 10 and 25% (27%). This suggests that by 2nd and 3rd year the actual level of plagiarism has exceeded the student perception.

Moving on to the reasons students gave for plagiarism these were ranked as follows:

1. They started too late and ran out of time.

2. They simply couldn't do the coursework otherwise.

3. They didn't think it was wrong.

4. They have to succeed. They got higher marks this way.

5. They didn't need to learn that material, just pass the module.

6. They couldn't keep up with the work.

7. They wanted to see if they could get away with it.

8. They felt the tutor didn't care, so why should they.

9. They thought paraphrasing would be disrespectful.

These appear largely similar to the results for collusion although the top two reasons have switched and the confusion over definitions is represented by "They didn't think it was wrong" appearing third. Being unable to keep up has dropped down the list.

Discussion

The responses suggest that plagiarism is indeed a problem. 10% of the respondents to the first questionnaire and a quarter of respondents to the second admitted to activities they either counted as plagiarism or at least considered should be penalised. Of even more concern are the numbers when unintentional plagiarism is taken into account with 25% of first years and a staggering 65% of second and third years indulging in academically suspect behaviour.

Over 90% of students understood that the most extreme cases of direct quotation without attribution counted as plagiarism. This suggests that students need more detailed explanations about activities short of this that might count as plagiarism, possibly using similar scenarios to those used in the questionnaire. An examination of the department's handbook, distributed to all students on a yearly basis, which contains the definition of plagiarism officially used

by the department revealed a simplistic definition which failed to explicate many of the subtleties revealed by the questionnaire. The department is currently engaged in revising this document as a result of these findings.

It was also clear that in most cases more students would think it fair for marks to be deducted than would label a scenario plagiarism. This suggests that the technical term "plagiarism" may, if anything, obscure the issue when students are warned against cheating.

The most common cause of plagiarism was seen to be lack of time and ability rather than laziness or contempt for the assessment process. Some of this is clearly due to poor time management on the part of students and a tendency for the less able students to be the ones least likely to seek out the various forms of support on offer. It is not immediately obvious that either of these pressures can be reduced by simple measures since they arise from a complex interaction of student and staff expectations, and the difficulties involved in managing coursework deadlines across a number of modules some of which are optional and some of which are available to more than one year of study. However it was valuable to learn that the majority of students did not feel plagiarism occurred because people simply wanted to gain marks without having any real desire to learn.

Related Work

A number of surveys have been conducted into levels of plagiarism in UK institutions (even more into plagiarism in American Institutions), frequently revealing very high levels of cheating.

Franklyn-Stokes and Newstead [4] looked at cheating in UK institutions and received depressing results: 72% students allowed coursework to be copied, 66% paraphrased without acknowledgement, 54% plagiarised from a text. This is based on self-reporting by the students.

This suggests several things. Firstly if these results are representative and the answers on our questionnaires are broadly correct then computer science has considerably less of a problem with collusion and plagiarism than other disciplines. It should be noted that since the questionnaires were distributed in a lecture it isn't unreasonable to assume that the sample may have included students less likely to plagiarise. Obviously if the cases of "unintentional plagiarism" are taken into account the story is more coherent and its impossible to easily determine whether the students in Franklyn-Stokes and Newstead's study had a better understanding of the definition of plagiarism than those in this study. It is hard to determine whether the perception of plagiarism is higher within computer science than other subjects – it is certainly the case that the programming exercise (one of the primary focuses for concern within the discipline) is peculiarly well suited to the use of electronic detection measures and that staff are in general highly computer literate and so eager to use such tools where they are available. On an anecdotal level no computer scientist the author has spoken to has been surprised by the level of plagiarism reported by this research while some staff members in other disciplines have expressed surprise and concern that plagiarism could be this widespread and, within their own discipline, relatively undetected.

Conclusions

The research reported in this paper set out to answer four questions

1. Did the students understand what constituted plagiarism or collusion?
2. Is the incidence of plagiarism as high as staff fear.
3. Why do students plagiarize?

The study revealed that there was some confusion among students over what might count as plagiarism when only sections of material were considered or when small amounts of paraphrasing were taken into account. The incidence of plagiarism did appear to be as high as staff feared especially when unintentional plagiarism was taken into account.

The reported reasons for plagiarism were primarily lack of time and ability. A more detailed study would be required before any strong conclusions could be drawn from this for on the one hand it could be caused by staff setting unreasonable deadlines for coursework and failing to deliver module content in a clear and engaging fashion on the other hand the problems could be the result of poor time management on the part of students and an unwillingness to seek out information for themselves or make appropriate use of the support facilities already in place. However what was clear was that disinterest in learning per se was not the main cause of cheating.

The most revealing finding of the study was a clarification of the statement "Students do not understand what plagiarism is". The use of scenarios within the questionnaires helped to make it clearer where the students' confusions lie.

References

[1] J. Carrol and J. Appleton. Plagiarism: A good practice guid. Technical report, Oxford Brookes University, 2001. Report for JISC.

[2] G. Chester. Plagiarism detection and prevention: final report ont he jisc electronic plagiarism detection project. Technical report, JISC, 2001.

[3] M. Davidson. Plagiarism. Technical report, University of Nottingham, Staff Development Unit, 2003. Notes for a half day workshop on plagiarism.

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

[5] Swales and Feale. *Academic Writing for Graduate Students.* University of Michigan, 1993.