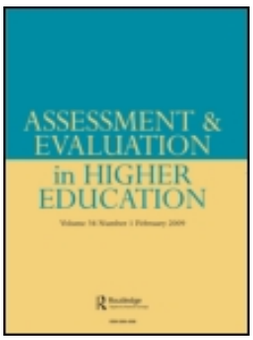

Routledge
Taylor & Francis Group

Assessment & Evaluation in Higher Education

ISSN: 0260-2938 (Print) 1469-297X (Online) Journal homepage: https://www.tandfonline.com/loi/caeh20

Source-code plagiarism in universities: a comparative study of student perspectives in China and the UK

Dongyang Zhang, Mike Joy, Georgina Cosma, Russell Boyatt, Jane Sinclair & Jane Yau

To cite this article: Dongyang Zhang, Mike Joy, Georgina Cosma, Russell Boyatt, Jane Sinclair & Jane Yau (2014) Source-code plagiarism in universities: a comparative study of student perspectives in China and the UK, Assessment & Evaluation in Higher Education, 39:6, 743-758, DOI: 10.1080/02602938.2013.870122

To link to this article: https://doi.org/10.1080/02602938.2013.870122

Assessment & Evaluation in Higher Education, 2014
Vol. 39, No. 6, 743–758, http://dx.doi.org/10.1080/02602938.2013.870122

Routledge
Taylor & Francis Group

Source-code plagiarism in universities: a comparative study of student perspectives in China and the UK

Dongyang Zhang[a], Mike Joy[b]*, Georgina Cosma[c], Russell Boyatt[b], Jane Sinclair[b] and Jane Yau[d]

[a]*College of Information and Management Science, Henan Agricultural University, Zhengzhou, P.R. China;* [b]*Department of Computer Science, University of Warwick, Coventry, UK;* [c]*School of Science and Technology, Nottingham Trent University, Nottingham, UK;* [d]*Department of Computer Science, Malmö University, Malmö, Sweden*

There has been much research and discussion relating to variations in plagiaristic activity observed in students from different demographic backgrounds. Differences in behaviour have been noted in many studies, although the underlying reasons are still a matter of debate. Existing work focuses mainly on textual plagiarism, and most often derives results by studying (small) groups of overseas students studying in a Western context. This study investigates understanding of source-code plagiarism (i.e. plagiarism of computer programmes) amongst university students in China. The survey instrument was a Chinese translation of a survey previously administered in English in the UK. This paper reports the results of the exploratory survey conducted in China, and compares these results to those from a parallel survey conducted in the UK. The results show that there is a significant difference in understanding between the respondents from the two surveys, and suggest topics which a future and more comprehensive study may focus on.

Keywords: source-code; plagiarism; UK; China

Introduction

Plagiarism in the academic community is regarded as malpractice and much work has been conducted on how to detect plagiarism, and how to prevent plagiarism by educating students that it is an unacceptable behaviour (Bradley 2011; Twomey, White, and Sagendorf 2009). A large volume of pedagogic material is available on university websites and other websites aimed at instructing students about plagiarism avoidance. Detection tools and services such as Turnitin (submit.ac.uk) are available to assist in the process of detecting plagiarism in student coursework consisting of essays and dissertations.

In the computing disciplines, much coursework consists of computer programmes, and the source-code for these can be plagiarised in a similar way to the contents of an essay. Tools exist to assist in plagiarism detection, such as MOSS (Bowyer and Hall 1999), JPlag (Prechelt, Malpohl, and Philippsen 2002) and Sherlock (Joy and Luck 1999), but such tools are relatively immature, since these do not detect plagiarism sources originating from the Internet.

*Corresponding author. Email: m.s.joy@warwick.ac.uk

The incidence of (textual) plagiarism and the motives behind it have often been viewed from a demographic and, in particular, cultural perspective. The IPPHEAE project (Impact of Policies for Plagiarism in Higher Education Across Europe: ipp-heae.eu) is currently comparing and evaluating the policies and procedures for detecting and preventing plagiarism, within the countries of the European Union, and it is hoped that the further substantial data collected from both staff and students will shed light on the European dimension.

Many studies have investigated plagiaristic activity amongst international students in Western universities. For example, Bamford and Sergiou (2005) obtained data from 35 international students at London Metropolitan University, over half of whom admitted to having plagiarised. Similarly, Bista (2011) studied international students at a southern US university, while Qi (2008) reports a study amongst Chinese first-year undergraduates in another US university. The general position often presented is of a higher rate of plagiarism detected amongst international students (Park 2003). Indeed, the figures presented are often very stark: for example, Qi (2008) refers to their own and five previous studies in which the majority of Chinese students at foreign institutions surveyed had plagiarised.

The reasons underlying these findings are somewhat less clear. Hayes and Introna (2005), in a study involving around 100 postgraduate students from a variety of countries, including the UK, China and Greece, noted a number of possible factors. These include proficiency in written English, different expectations relating to the examination process (in some countries, for example, assessment is almost completely by written examination), and differing understanding of the purpose of language. The latter is a theme explored in depth by Pennycook (1996), who argues that there are different types and levels of repetition and memorisation employed by Chinese learners, so that 'a student's "ownership" over a text may have different causes and different effects' (222). This is echoed in a more recent study by Hu and Lei (2012, 818), who note that 'memorization, repetition, and imitation of authoritative texts are valued as legitimate learning strategies'.

Yakovchuk (2008), in a study of non-native university students in the UK, further identifies a variety of reasons for plagiarising given by those students, including content problems (lacking individual ideas, background knowledge or subject understanding), language problems (poor writing skills and difficulties in expressing ideas) and laziness. Of these, the only factor specific to international students seems to be that relating to language. Gullifer and Tyson (2013), in a recent study at an Australian university, also suggest that students often misunderstand institutional policies on plagiarism, and note that many do not even read those policies even when instructed to do so.

Many authors (e.g. Leask 2006; Sowden 2005) have referred to cultural influences which might lead to plagiarism. However, Liu (2005) cautions against inferring that the incidence of plagiarism within certain identifiable groups is symptomatic of cultural conditioning, and calls for further 'extensive empirical research' (239). Liu questions assumptions about Chinese cultural influence, arguing that plagiarism is not viewed as acceptable. He also notes that some differences (such as a tradition of memorising and quoting accepted authorities) do exist, but are not in themselves explanations for plagiarism since attribution would still be expected. Many studies have been conducted in a context where the 'cultural outsiders' are also linguistic outsiders, and hence it is difficult to separate the two. Stone (2008), in a detailed analysis of Chinese perspectives on intellectual property, scruti-

nises the reasons for intellectual property rights for the written word not having been developed, and argues that ascribing these to the 'Confucian tradition' is 'not persuasive' (230).

In the context of source-code plagiarism, the dialogue is much less mature. Instances of industrial plagiarism in China have been reported (Eaton 2009; Patel 2005), but data evidencing the incidence of student source-code plagiarism within certain groups are almost non-existent. The few studies on student perceptions of source-code plagiarism mostly lack analysis of student cultural or national background (Chuda et al. 2012; Mann and Frew 2006), although Sraka and Kaučič (2009) report a small study conducted in Slovenia. In general, source-code plagiarism is seen to be just as much of a problem as textual plagiarism; indeed, there is reason to believe that it may give rise to greater confusion. Mann and Frew (2006) found that students viewed a proportion of between 60 and 90% similarity in code to be acceptable. Chuda et al. (2012) report that students who have a good understanding of textual plagiarism were uncertain as to what constitutes plagiarism in source-code. Joy et al. (2011) discovered widespread misunderstandings amongst students in the UK concerning certain aspects of source-code plagiarism.

It would be difficult, if not impossible, to perform an accurate empirical study to ascertain the comparative frequency of occurrence of source-code plagiarism between student groups of different nationalities. The detection tools are imprecise, and it is likely that many instances of source-code plagiarism go undetected or unreported. Furthermore, access to data would likely be denied, both for ethical reasons, and by restrictions imposed by institutional authorities.

An alternative approach, such as that adopted by Bamford and Sergiou (2005), is to gather data direct from students and to ascertain their perceptions about source-code plagiarism. There is a considerable amount of literature on plagiarism which states that students with high ethical views who claim to understand about plagiarism are able to avoid it. However, a recent study by Risquez, O'Dwyer, and Ledwith (2011) contradicts this assumption, and suggests that experiential activities are required to reinforce students' understanding. This cautions us that a student's perception may not be an accurate indicator of his/her understanding, but exploring this further is beyond the scope of this paper.

The purpose of this paper is twofold. First, we report the results of an exploratory study conducted with students in several universities in China, highlighting areas of uncertainty and consistent misunderstanding concerning source-code plagiarism. Secondly, we perform a comparison with the results of an equivalent study delivered in the UK (Joy et al. 2011) and consider the implications for future work.

Methodology

This paper follows on from work by Joy et al. (2011), in which a questionnaire was answered by 770 students in 18 institutions in the UK, and an equivalent methodological approach to that followed in the current research was taken. Although 770 responses to the UK questionnaire were received, 615 of the respondents answered 'I am a UK student' in response to the question 'which of the following most accurately describes your background?' and the analysis here is restricted to that subset of 615 responses in order to exclude Chinese and other non-UK students from that sample. The first part of the questionnaire consisted of demographic questions, and the second part consisted of 15 'scenarios' describing actions related to reusing and

referencing source-code. For each scenario, the participants were asked to select one response from the following Likert scale, to indicate whether or not they thought the scenario was an example of plagiarism:

(1) Yes, definitely.
(2) I think it is, but I am not completely sure.
(3) I don't know.
(4) I think it is not, but I am not completely sure.
(5) No, definitely not.

Prior to responding to the scenarios, the participants were reminded of what plagiarism is, and were given the Wikipedia definition: 'the practice of claiming or implying original authorship of (or incorporating material from) someone else's written or creative work, in whole or in part, into one's own without adequate acknowledgement' (the equivalent Chinese translation from Wikipedia was used in the Chinese survey).

Each of the 15 scenarios was scrutinised by at least four academics, experienced in plagiarism detection and university disciplinary processes, to ensure that each scenario was very clearly plagiarism or not plagiarism – there were no scenarios which were 'grey areas'. For example, the following scenario was used, and depicts an activity which is clearly plagiarism, and the correct response is 'Yes, definitely'.

> Andy is required to submit a Java applet programme for his assignment. He remembers reading about a similar applet programme from a textbook he's been using. He goes to find this, and uses this code and then submits it as his own programme without noting that he obtained the code somewhere else. Is this plagiarism?

The responses were then coded, with +1.0 for a correct answer (which will always be either a or e) and −1.0 for an incorrect answer (a or e), +0.5 or −0.5 for a (b or d) answer, and 0 for (c). This approach allowed a clear description of student responses to be presented, since the score for each question indicates how close the respondent is to the correct answer, although it is recognised that a Likert scale is not an interval scale. The original questionnaire was then translated into Chinese by a native speaker and checked by a second; the Chinese translation was then retranslated back to English to ensure the meaning had not been altered.

The questionnaire was then published as an online survey. A total of 159 responses were gathered from students at 30 universities throughout China, located in Beijing (2), Shanghai (2), Tianjin (1), and the provinces of Guangdong (2), Hebei (5), Henan (5), Hubei (1), Hunan (2), Jiangsu (2), Liaoning (1), Shaanxi (2), Shandong (1), Sichuan (3) and Zhejiang (1) (the number of institutions in each location is included in parentheses).

Data analysis

The number of respondents ($n = 159$) is sufficient for testing the hypothesis that there is a difference in the perceptions of plagiarism by Chinese students as compared to UK/European students. However, further detailed analysis is precluded by the sample size, and by the voluntary nature of participation in the survey. Indeed, given the size of China, one would expect regional differences in responses, and

perhaps also differences relating to the type of institution (research-led, regional, etc.). The respondents come from provinces throughout China, and cover different types of university, including regional teaching universities and nationally renowned research-led institutions. Thirteen institutions are included in the 'Project 211' list of 112 nationally important key universities, six in the 'Project 985' list of 39 top-ranking universities and three in the recently launched 'Project 2011', which aims to promote universities' innovation ability and includes 14 collaborative innovation centres formed by 23 universities and a few research centres. (Ministry of Education 2006, 2013a, 2013b). Although it is difficult to claim that the sample is a representative cross-section of students in China, it nevertheless covers a wide range of geographical locations and institutional types, and is sufficient for the exploratory nature of this project.

It is worth noting that, of the 159 Chinese respondents, only 70 claimed to have been informed about plagiarism by their institution, and 114 stated that they understood what plagiarism meant. As noted above, all respondents were then provided with a standard definition of plagiarism.

Overall scores

Each answer was allocated a mark in the range of {−1.0, −0.5, 0, +0.5, +1.0}. A total score was calculated for each student by summing the marks for all their answers. The total score scale ranged from −15 to +15, −15 was allocated if all the responses were incorrect and +15 was allocated if all responses were correct. The use of an interval scale was to be able to clearly depict the student responses, since the point allocated to each response gives an indication of how close the student was to the correct answer.

The majority of students (61.6%, $n = 98$) received a total score in the range between +0.5 and +5.0, 23.9% ($n = 38$) received a total score in the range of +5.5 to +10.5 (10.5 being the highest score achieved by students), and 14.5% ($n = 23$) of students achieved a total score in the range of −2.0 to 0.0. Figure 1 shows the frequencies of total scores.

Topic 1: self-plagiarism

The fundamental issue with self-plagiarism is that previous work must be acknowledged. Two scenarios were presented on the topic of self-plagiarism, describing different ways in which a student copied and integrated source-code they had authored as part of a programming assignment for academic credit. Scenario 1a describes a case where the student *has* acknowledged (as a source-code 'comment') that he/she copied source-code from his/her own previous work – this scenario does not constitute plagiarism – and scenario 1b describes a case where the student *has not* acknowledged that he/she copied source-code from his/her own previous work – this scenario does constitute plagiarism.

The results for question 1a revealed that 66.7% ($n = 106$) of respondents answered correctly that plagiarism had definitely not taken place. Also, 20.1% ($n = 32$) responded 'I think it isn't, but I am not completely sure', whereas the remaining respondents (i.e. 13.2%, $n = 21$) answered incorrectly. Concerning scenario 1b, which describes a case where a student copied from one of their previous source-code assignments without acknowledgement, only 3.8% ($n = 6$) answered correctly

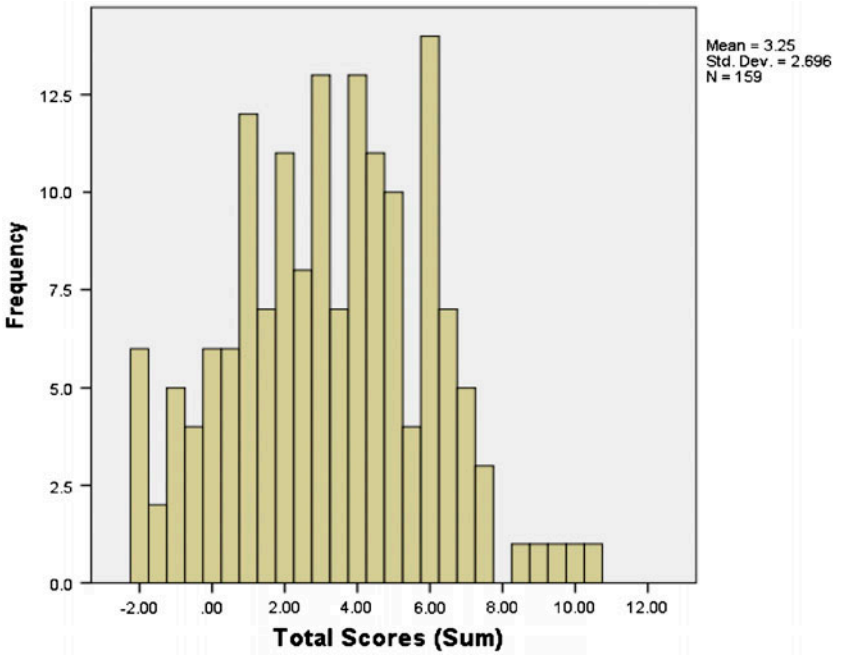

Figure 1. Histogram of total scores.

that plagiarism had definitely taken place and 9.4% ($n = 15$) of the respondents were unsure. The vast majority, 86.8% ($n = 138$), answered incorrectly.

The results of scenarios 1a and 1b suggest that students perceive copying their own work as not a plagiaristic activity, regardless of whether acknowledgement is included in the code.

Topic 2: copying code from books and other sources

Scenario 2a describes a case of plagiarism where a student copied source-code from a book without acknowledging this. Surprisingly, only 25.8% ($n = 41$) of the respondents believed that this action constitutes plagiarism, and only 28.3% ($n = 45$) thought it was plagiarism but were not completely sure. The remaining 45.9% ($n = 73$) of the respondents answered incorrectly.

Scenario 2d describes a case where a student read some books for gaining inspiration and then produced their own source-code, not acknowledging any of the sources they used to gain the inspiration. Since no copying was involved, this scenario does not describe a plagiaristic activity. 34.6% ($n = 55$) of respondents were certain that this scenario was not plagiarism, whereas 32.3% ($n = 53$) thought that this scenario did not describe plagiarism but were uncertain. The remaining 32.1% ($n = 51$) answered incorrectly.

Both scenarios, 2a and 2d, involve cases where the student did not provide references, but the main distinction between the two is that the first scenario, 2a, describes a clear act of plagiarism about copying source-code from a book, whereas the second scenario, 2d, describes a case where the student consults books for

inspiration and does not copy any source-code. Although it would have been appropriate for the student to add a note about the books they used for inspiration, this scenario does not describe a plagiarist activity. The results show that the majority of students understand that reading for inspiration does not constitute plagiarism, but are unclear whether copying without referencing constitutes plagiarism.

Scenario 2e describes a case where students had performed a group project and had ensured that all source-code reuse was acknowledged. The results for this scenario revealed that the majority, 64.2% ($n = 102$), of the respondents have a clear understanding that referencing work produced as part of a group does not constitute plagiarism, and a further 23.3% ($n = 37$) were unsure. In conclusion, the majority of respondents have a clear understanding that copying work and referencing it does not constitute plagiarism.

Scenario 2c describes a similar scenario to 2e, where a group of students appropriately referenced all the codes they have used from books and other sources and all other reused material. 57.9% ($n = 92$) of students responded correctly that this scenario does not constitute plagiarism and 32.7% ($n = 52$) of students think that the scenario is not plagiarism but are not completely sure. The remaining 9.4% ($n = 15$) of the respondents answered incorrectly.

Scenario 2b describes an act of plagiarism where a student has used code from a book but has provided an incorrect reference. Since they have taken code from a book and not referenced the particular book, then this act is clearly plagiarism, whether or not the incorrect reference was deliberate or negligent. Unexpectedly, only 10.7% ($n = 17$) answered this question correctly, and 19.5% ($n = 31$) thought that the scenario described plagiarism but were not completely sure. The remaining 69.8% ($n = 111$) of the respondents provided an incorrect response. These results verify the hypothesis that students do not clearly understand that taking code without providing correct references constitutes plagiarism.

Topic 3: copying from another student

Scenario 3a describes a case where a student copies code from their fellow student, modifies it and submits it as their own work. This is a clear case of plagiarism, even though the code has been modified, since the source has not been referenced (this is analogous to paraphrasing in an essay without citing the source). The majority of students seem to understand this. 57.2% ($n = 91$) of respondents answered correctly that this scenario constitutes plagiarism, and 26.4% ($n = 42$) responded that they think this scenario constitutes plagiarism but are not completely sure. The remaining 16.4% ($n = 26$) answered incorrectly.

Scenario 3c describes a case where a student finds another student's uncollected printout which contains the solution to the assignment they are working on, and copies this solution and submits it. This is another clear case of plagiarism, since the student is submitting some other student's work. 43.4% ($n = 69$) of respondents answered correctly and 27.0% ($n = 43$) thought that this scenario described a case of plagiarism but were not completely sure, whereas the remaining 29.6% ($n = 47$) of respondents answered incorrectly.

Scenario 3b describes a slightly different scenario where a student pays someone to create their assignment. This is also a case of plagiarism, since a student copied and submitted someone else's work, but it is *also* a case of cheating, since the student paid that someone to produce the code. Surprisingly, only 16.4% ($n = 26$)

considered this scenario to be plagiarism, and 17.0% ($n = 27$) thought that this was plagiarism but were not completely sure. The remaining 66.6% ($n = 106$) answered incorrectly.

Topic 4: inappropriate collaboration

Scenario 4a describes a scenario where two students collaborate and work together on an individual assignment and submit similar solutions. This is referred to as 'inappropriate collaboration', since the students are expected to work on the assignment alone and collaborating with other students has been forbidden by the instructor. 8.2% ($n = 13$) correctly responded that this scenario constitutes plagiarism, and 14.5% ($n = 23$) thought that this scenario is plagiarism but are not completely sure. Surprisingly, the majority, 77.3% ($n = 123$), of the respondents did not consider this scenario to be a case of plagiarism.

Scenario 4b describes a scenario where two students working in two different groups, which are assigned the same coursework, exchange parts of their work. Clearly, this is a case of plagiarism, since each student submits work that is in part not their own, and the other student's contribution is not acknowledged. A minority, 13.8% ($n = 22$), of the respondents considered this scenario to be plagiarism, and 22.6% ($n = 36$) thought that this scenario may be plagiarism but were not completely sure. The majority of respondents, 63.6% ($n = 101$), incorrectly responded that this scenario did not constitute plagiarism.

A supplementary question had been added to the China survey, and the corresponding scenario (4c) describes a case where a student posts a programming question on an online forum, receives a solution and copies that solution into their assignment, and is clearly a case of plagiarism, since the student submitted someone else's work. This question was added following suggestions by Chinese colleagues that this activity was commonplace in China. Importantly, only 20.1% ($n = 32$) of students considered this scenario to be a definite plagiarism, whereas 37.7% ($n = 60$) considered this scenario to be plagiarism but were not completely sure. The remaining 42.1% ($n = 67$) of students responded incorrectly that this scenario does not constitute plagiarism.

Clearly, these results indicate that students do not consider inappropriate collaboration by exchanging work as plagiarism.

Topic 5: converting to another programming language

In scenario 5a, a student takes Java source-code from a book, converts it to C++ and incorporates it into their assignment without referencing the origin of the Java code. The languages Java and C++ are similar, the functionality of the programs are the same and the act of conversion is analogous to minor paraphrasing in an essay; so, this is clearly plagiarism. 15.1% ($n = 24$) agreed, and a further 17.0% ($n = 27$) thought that this was plagiarism but were uncertain. The remaining 67.9% ($n = 108$) of respondents incorrectly responded that this scenario did not constitute plagiarism.

Scenario 5b describes a case where the student has acknowledged (as a 'comment') that they copied source-code from their own previous work and converted it to another similar programming language. Scenario 5b is not plagiarism. The majority of respondents, 60.4% ($n = 96$), correctly responded that this scenario was not plagiarism, whereas 23.9% ($n = 38$) responded that this scenario was not plagiarism

but were not completely sure. The remaining 15.7% ($n = 25$) incorrectly replied that the scenario described a plagiaristic activity.

The conclusion here is that students do not understand that converting source-code from one programming language to another constitutes plagiarism.

Topic 6: falsification as opposed to plagiarism

Topic 6 describes a case where a student's program displays output different to what is required by the assignment specification, and the student intentionally modifies the printouts of the program's output to make it appear as if the program displays the correct output. So, the student's program does not work, and the student cheats by adding extra program statements to make it look as if it does. This scenario describes a case of falsification instead of plagiarism, since the student does not copy someone else's work. A total of 45.9% ($n = 73$) of students were certain that this is not plagiarism, and 5.7% ($n = 9$) of respondents answered incorrectly that this scenario constitutes plagiarism. The majority of students were uncertain (48.4% ($n = 77$)). The low percentage of correct answers might simply suggest an imprecise understanding of the word 'plagiarism', in which students infer it also covers cheating not involving copying.

Comparative study

The two data-sets – the UK and China responses – were then compared to test the hypothesis that *there is a significant difference in understanding between the respondents from the two surveys* and to identify which of the topics addressed by the surveys suggested the most significant differences.

Statistical analysis tests

The UK and China student responses were statistically analysed using SPSS software. The aim of the analysis was to determine significant differences between the two independent groups on each scenario (i.e. one group consisting of students studying in the UK and the other group consisting of students studying in China), and determine topics of common understanding and misunderstanding.

A total of 615 UK student responses was compared to a total of 159 China student responses. The Mann-Whitney U test was performed to compare the differences between the responses of UK and China students (Figure 2). In conjunction with the Mann-Whitney U test, the *effect size measure* provides the strength of the relationship between two groups. Cohen (1988) proposed rules of thumb for interpreting effect sizes: a 'small' effect size is a 'Cohen's *d*' value of 0.20, a 'medium' effect size is 0.50 and a 'large' effect size is 0.80.

Significant differences on scenarios with large effect, medium to large, and medium effect sizes indicate the areas for which there may be significant and prominent differences in the responses of UK and China students. Significant differences with small or minimal effect sizes indicate areas with small differences in the mean ranks of responses between the two groups. Small effect size differences (and sometimes medium effect size, based on the *r* value) suggest common understandings or misunderstandings across topics between the two groups.

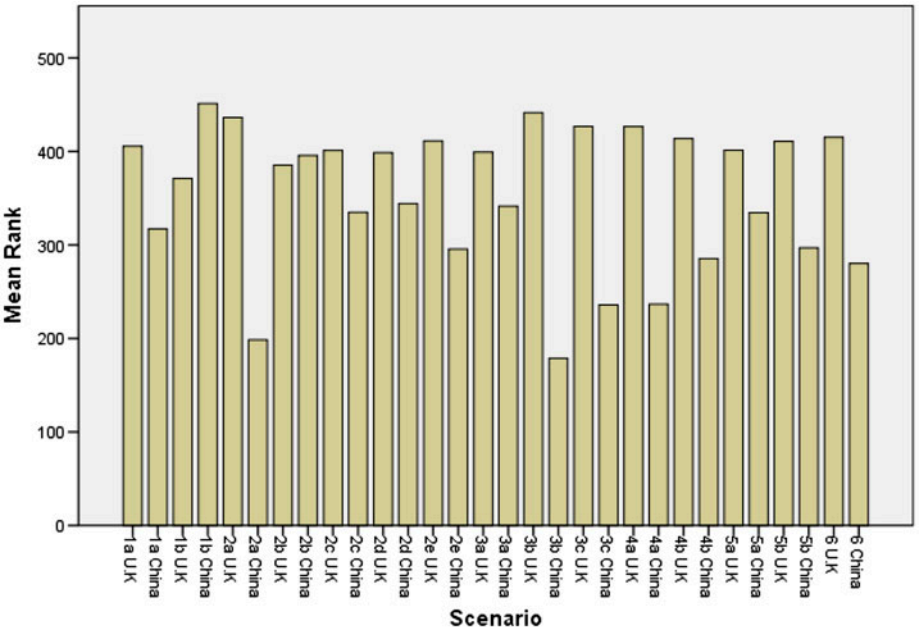

Figure 2. Results of the Mann-Whitney U test.

Topics that UK students understand but China students misunderstand

Scenario 3b describes a scenario where a student pays someone to create his assignment. Although this scenario describes an act of cheating, it also constitutes plagiarism since a student submitted someone else's work as his own. There was a significant difference in the mean ranks of UK and China students on scenario 3b, $U = 15718.5$, $p = 0.00$, $r = -0.57$, which is considered a medium effect size. This suggests that there was a variation across the responses of the two groups, and a total of 80.8% of UK students answered correctly, whereas only 16.4% of China students answered correctly.

Scenario 3c describes a clear case of plagiarism in which a student finds another student's uncollected printout of the assignment's solution and copies this solution and submits it. 43.4% of China and 80.8% of UK respondents answered correctly. For scenario 2a, which described a case where a student copied code from a book without providing acknowledgement, only 25.8% of the responses from China were correct compared to the 82.1% of the UK responses. The results for 3c revealed significant differences of medium effect size, $U = 24{,}790$, $p = 0.00$, $r = -0.51$, as did the results for scenario 2a, $U = 18836.5$, $p = 0.00$, $r = -0.54$.

Common topics of understanding between UK and China students

Self-plagiarism is a topic for which the responses of UK and China students were similar, but with significant differences of small effect size. Scenario 1a (acknowledged copying from a previous assignment) elicited UK student responses which were 89.1% correct, and China responses which were 66.7% correct, with significant but only small effect size differences between the two groups, $U = 37{,}713$, $p = 0.00$, $r = -0.25$.

For scenarios 2c and 2e, which clearly state that the students have provided appropriate acknowledgement to code that is not their own, the results suggest that there is common understanding between the two groups of students that these scenarios do not constitute plagiarism, with significant but only small effect size differences between the two groups, $U = 40520.5$, $p = 0.00$, $r = -0.15$ for scenario 2c and $U = 34{,}282$, $p = 0.00$, $r = -0.37$ for scenario 2e. Some 76.3% of UK respondents and 57.9% of China respondents answered correctly on scenario 2c, and 93.8% of UK respondents and 64.2% of China respondents answered correctly for scenario 2e.

For scenario 3a, in which a student copies code from a fellow student, modifies it and submits it as their own work, a significant difference of small effect size in responses between the two groups was noted, $U = 41572.5$, $p = 0.00$, $r = -0.13$. The majority of UK and China students seem to understand that this is a clear case of plagiarism, with 70.6% of UK respondents and 57.2% of China respondents answering correctly.

Topic 5 is concerned with converting code to another programming language, and in scenario 5b, a case is described where a student takes Visual Basic source-code, converts it into Java and incorporates it into their assignment but provides a note of this. Scenario 5b is not plagiarism, and 89.4% of UK respondents and 60.4% of China respondents answered correctly. Therefore, there exists common understanding that translating code and providing acknowledgement of this fact is not regarded as plagiarism, with significant but only small effect size differences between the two groups, $U = 34{,}500$, $p = 0.00$, $r = -0.32$.

Finally, scenario 6 describes a case of falsification rather than plagiarism, where a student modifies the program's output to make it seem as it is working as specified in the assignment specification. A total of 80.8% of the UK students, and 45.9% of China students answered correctly (score = 1.0). The differences in the responses between the two groups was significant but of small effect size with $U = 31{,}836$, $p = 0.00$, $r = -0.32$.

Common topics of misunderstanding between UK and China students

For the second self-plagiarism scenario (1b), in which copying occurred from a previous source-code assignment without appropriate acknowledgment, the responses were very different: the majority of both UK and China students incorrectly responded that this scenario does not constitute plagiarism. Only 3.4% of UK responses and 3.8% of China student responses were correct, $U = 38{,}775$, $p = 0.00$, $r = -0.20$.

Results for scenarios 2b and 2d (copying from books) revealed common misunderstandings between the two groups. For scenario 2b (incorrect reference), the difference was not significant and of minimal effect size, $U = 47{,}592$, $p = 0.59$, $r = -0.02$, with only 11.2% of UK students and only 10.7% of China students answering correctly. Scenario 2d (where a student read some books for gaining inspiration but did not copy) received correct responses from a total of 47.8% of the UK respondents and 34.6% of the China respondents. The differences between the two groups were significant but of small effect size, indicating that both groups concerned, are uncertain about using resources for gaining inspiration, $U = 42015.5$, $p = 0.00$, $r = -0.10$.

Scenarios 4a and 4b are both considered as plagiarism. For scenario 4a, which describes a case where two students collaborate on an individual assignment and submit similar solutions, only 34.1% of UK respondents and 8.2% of China respondents answered correctly. Similarly, for scenario 4b, where two students working in two different groups are assigned the same coursework and exchange parts of their work, only 32.4% of UK respondents and 13.8% of China respondents answered correctly. These results strongly suggest that UK and China students do not consider inappropriate collaboration as plagiarism. For both scenarios concerned, the differences between the two groups were significant, with medium and small effect sizes, respectively, $U = 24903.5$, $p = 0.00$, $r = -0.35$ for scenario 4a, and $U = 32{,}672$, $p = 0.00$, $r = -0.24$ for scenario 4b.

Scenario 5a describes a clear case of plagiarism, where a student takes Java source-code from a book, converts it to C++ and incorporates it into her assignment without referencing the origin of the Java code. Only 24.2% of UK students and only 15.1% of China students answered correctly. The difference was significant and of small effect size, strongly indicating the common misunderstanding between the two groups, $U = 40457.5$, $p = 0.00$, $r = -0.12$. The conclusion here is that students do not understand that converting source-code from one programming language to another constitutes plagiarism.

Discussion and further work

This paper describes the results of a study, on the topic of source-code plagiarism, which gathered the perceptions of China students and compared these perceptions to those gathered from UK students (Joy et al. 2011). For the purposes of conducting a comparative study, the questionnaire distributed to UK students was translated to the Chinese language and was then distributed to students studying in China. Importantly, China was the home country and Chinese was the native language of all respondents from China. The results of the comparative study revealed areas of understanding and misunderstanding between the two groups, and these are summarised in Table 1. Table 1 shows the results for each group (i.e. UK and China) based on the responses of the majority of students (i.e. $n > 50\%$) who answered correctly (score = 1.0) on each scenario.

In conclusion, the topics of reuse of the student's own source-code without acknowledgement, translation of source-code without acknowledging the original source and incorrect referencing appear to be issues of concern both in the UK and China. In addition, some issues such as getting code from another person (for example, by paying them, or through an online forum), use of code from a book without acknowledgement and collaboration on individual assignments appear not to be fully understood by China participants.

These results indicate specific areas of student misunderstanding about plagiarism, but more generally they suggest that both UK and Chinese students are unsure of the boundary between legitimate academic activity and unacceptable plagiarism. This paper does not address *why* students misunderstand, and further studies will be required to elicit the reasons, but we may conjecture that the approach to instructing students about academic integrity is deficient in some institutions.

In the Chinese context, there is a clear lack of advice given to students, since a majority of the students participating in this study claim not to have had any instruction, and this would need to be addressed. Possible approaches might include

Table 1. Areas of understanding and misunderstanding.

Scenario	Scenario topic	Scenario summary	Plagiarism and not plagiarism
Topics of common understanding			
1a	Topic 1: self-plagiarism and source-code reuse	Copying from own previous work *and* acknowledging this as a source-code 'comment'	Not plagiarism
2e	Topic 2: copying code from books and other sources	Reusing source-code and acknowledging it	Not plagiarism
3a	Topic 3: copying code from another student	Copying source-code from a fellow student and with consent modifying it and submitting	Plagiarism
5b	Topic 5: converting code to another programming language	Copying source-code from own previous work and converting it to another similar programming language and acknowledging this	Not plagiarism
2c	Topic 2: copying code from books and other sources	Appropriately referencing all the source-code and other material reused from books and other sources	Not plagiarism
6	Topic 6: falsification (as opposed to plagiarism)	Modifying a programme's output to make it seem as it is working as specified in the assignment specification	Not plagiarism
Topics of common misunderstanding			
1b	Topic 1: self-plagiarism and source-code reuse	Copying source-code from own previous work without acknowledging this	Plagiarism
2b	Topic 2: copying code from books and other sources	Reusing source-code from a book and providing an incorrect reference	Plagiarism
2d	Topic 2: copying code from books and other sources	Gaining inspiration from books and then producing source-code without copying	Not plagiarism
4a	Topic 4: inappropriate collaboration	Inappropriately collaborating on individual assignments and submitting similar solutions	Plagiarism
4b	Topic 4: inappropriate collaboration	Members of different groups working on the same coursework exchange parts of their work	Plagiarism
5a	Topic 5: converting code to another programming language	Reusing Java source-code from a book, converting it to C++ and incorporating it into an assignment without referencing the origin of the Java code	Plagiarism
Topics which UK students understand and students from China do not understand			
2a	Topic 2: copying code from books and other sources	Copying source-code from a book without acknowledging this	Plagiarism
3b	Topic 3: copying code from another student	Paying someone to create an assignment	Plagiarism
3c	Topic 3: copying code from another student	Copying and submitting another student's programming assignment without consent	Plagiarism

richer support materials (especially tailored to the computing disciplines), or a more fundamental ethics-led approach to student writing (and coding) skills, as suggested by Heitman and Litewka (2011).

Acknowledgements

This work was supported by the Higher Education Reform Research Project of Henan Agricultural University (Xiao Zheng Jiao [2012] No. 3).

Notes on contributors

Dongyang Zhang received her PhD from Reading University, UK. She is an associate professor in the College of Information and Management Science, Henan Agricultural University, China. She has industrial and academic experience in both UK and China. Her research interests include industrial sustainability and sustainable development of higher education.

Mike Joy received the PhD degree in computer science from the University of East Anglia. He is currently an associate professor at the University of Warwick. His research interests focus on educational technology and computer science education.

Georgina Cosma received a PhD Degree in computer science from the University of Warwick, and a BSc computer science degree from Coventry University. She is currently a senior lecturer at Nottingham Trent University. Her research interests include source-code similarity detection, information retrieval, artificial neural networks and other machine learning algorithms.

Russell Boyatt is a research associate in the University of Warwick's Computer Science Department where he is also completing his PhD. His research interests include educational technology, software development and formal methods and he is currently working on the EU Lifelong Learning Project 'MALog', developing learning resources for mathematical logic. His work on educational technology includes online assessment, learning environments and source code plagiarism detection.

Jane Sinclair received her PhD in Computer Science at the Open University, UK. She is an associate professor in the Department of Computer Science at the University of Warwick. Her interests include educational technology, computer security and formal methods. Her published research relating to education covers a variety of topics including plagiarism detection, the student perspective and the use of learning objects.

Jane Yau received her PhD degree in Computer Science from the University of Warwick. She is currently a postdoctoral researcher in Computer Science at Malmö University, Sweden. She has published around 30 scientific articles primarily in the area of mobile learning.

References

Bamford, J., and K. Sergiou. 2005. "International Students and Plagiarism: An Analysis of the Reasons for Plagiarism among International Foundation Students." *Investigations in University Teaching and Learning* 2 (2). Accessed April 16, 2013. www.londonmet.ac.uk/library/b10884_3.pdf

Bista, K. 2011. "Academic Dishonesty among International Students in Higher Education." *To Improve the Academy: Resources for Faculty, Instructional, and Organizational Development* 30: 159–172.

Bowyer, K., and L. Hall. 1999. "Experience Using MOSS to Detect Cheating on Programming Assignments." *Proceedings of 29th Annual Frontiers in Education Conference (FIE99)*, 18–22, San Juan, Puerto Rico.

Bradley, C. 2011. *Plagiarism Education and Prevention: A Subject-driven Case-based Approach*. Oxford: Chandos Publishing.
Chuda, D., P. Navrat, B. Kovacova, and P. Humay. 2012. "The Issue of (Software) Plagiarism: A Student View." *IEEE Transactions on Education* 55 (1): 22–28.
Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum.
Eaton N. 2009. "Plurk Has no Plans to Sue Microsoft – Yet." The Microsoft Blog. Accessed April 16, 2013. http://blog.seattlepi.com/microsoft/2009/12/15/plurk-has-no-plans-to-sue-microsoft-and150-yet/
Gullifer, J. M., and G. A. Tyson. 2013. "Who has Read the Policy on Plagiarism? Unpacking Students' Understanding of Plagiarism." *Studies in Higher Education Ahead-of-print*. doi:10.1080/03075079.2013.777412.
Hayes, N., and L. D. Introna. 2005. "Cultural Values, Plagiarism, and Fairness: When Plagiarism Gets in the Way of Learning." *Ethics and Behavior* 15 (3): 213–231. doi:10.1207/s15327019eb1503_2.
Heitman, E., and S. Litewka. 2011. "International Perspectives on Plagiarism and Considerations for Teaching International Trainees." *Urologic Oncology: Seminars and Original Investigations* 29 (1): 104–108.
Hu, G., and J. Lei. 2012. "Investigating Chinese University Students' Knowledge of and Attitudes toward Plagiarism from an Integrated Perspective." *Language Learning* 62 (3): 813–850.
Joy, M. S., G. Cosma, J. Y.-K. Yau, and J. E. Sinclair. 2011. "Source-code Plagiarism – A Student Perspective." *IEEE Transactions on Education* 54 (1): 125–132.
Joy, M. S., and M. Luck. 1999. "Plagiarism in Programming Assignments." *IEEE Transactions on Education* 42 (2): 129–133.
Leask, B. 2006. "Plagiarism, Cultural Diversity and Metaphor – Implications for Academic Staff Development." *Assessment & Evaluation in Higher Education* 31 (2): 183–199.
Liu, D. 2005. "Plagiarism in ESOL Students: Is Cultural Conditioning Truly the Major Culprit?" *ELT Journal* 59 (3): 234–241. doi:10.1093/elt/cci043.
Mann, S., and Z. Frew. 2006. "Similarity and Originality in Code: Plagiarism and Normal Variation in Student Assignments." *Proceedings of the 8th Australasian Conference on Computing Education* (*ACE '06*), Hobart. 143–150.
Ministry of Education. 2006. "985 Project" School List. Accessed October 26, 2013. http://www.moe.gov.cn/publicfiles/business/htmlfiles/moe/s238/201112/xxgk_128833.html (in Chinese).
Ministry of Education. 2013a. "Project 211" List of Schools. Accessed October 26, 2013. http://www.moe.edu.cn/publicfiles/business/htmlfiles/moe/moe_94/201002/82762.html (in Chinese).
Ministry of Education. 2013b. "Notice of Ministry of Education and Ministry of Finance's announcement of the result of Collaborative Innovation Center 2012." *Educational Technology Letter* 26. Accessed October 27, 2013. http://www.moe.gov.cn/publicfiles/business/htmlfiles/moe/s7062/201306/xxgk_152878.html (in Chinese).
Park, C. 2003. "In Other (People's) Words: Plagiarism by University Students – Literature and Lessons." *Assessment & Evaluation in Higher Education* 28 (5): 471–488.
Patel, N. 2005. "Open Source and China: Inverting Copyright?" *Wisconsin International Law Journal* 23 (1): 781–805.
Pennycook, A. 1996. "Borrowing Others' Words: Text, Ownership, Memory, and Plagiarism." *TESOL Quarterly* 30 (2): 201–230.
Prechelt, L., G. Malpohl, and M. Philippsen. 2002. "Finding Plagiarisms among a Set of Programs with JPlag." *Journal of Universal Computer Science* 8 (11): 1016–1038.
Qi, S. 2008. "A Descriptive Study of Plagiarism in the Writing Assignments of First-year Chinese Students in the American English Language Institute at Stephen F. Austin State University." MEd thesis, Stephen F. Austin State University, ProQuest, Ann Arbor, MI.
Risquez, A., M. O'Dwyer, and A. Ledwith. 2011. "'Thou Shalt not Plagiarise': From Self-Reported Views to Recognition and Avoidance of Plagiarism." *Assessment & Evaluation in Higher Education* 38 (1): 34–43.
Sowden, C. 2005. "Plagiarism and the Culture of Multilingual Students in Higher Education Abroad." *ELT Journal* 59 (3): 226–233.

Sraka, D., and B. Kaučič. 2009. "Source-code Plagiarism." *Proceedings of the 31st International Conference on Information Technology Interfaces (ITI 2009)*, 461–466. Dubrovnik, Croatia. doi:10.1109/ITI.2009.5196127.

Stone, C. R. 2008. "What Plagiarism was not: Some Preliminary Observations on Classical China Attitudes toward What the West Calls Intellectual Property." *Marquette Law Review* 92 (1): 199–230.

Twomey, T., H. White, and K. Sagendorf. eds. 2009. *Pedagogy, not Policing: Positive Approaches to Academic Integrity at the University*. Syracuse, NY: Graduate School Press of Syracuse University.

Yakovchuk, N. 2008. "Towards Understanding the Root Causes of Plagiarism among Non-native Speaker Students." In *Current Developments in English for Academic, Specific and Occupational Purposes*, edited by M. Krzanowski, 39–53. Reading: Garnet.