Academic Integrity: Differences between Computing Assessments and Essays

Simon
University of Newcastle, Australia
simon@newcastle.edu.au

Beth Cook
University of Newcastle, Australia
beth.cook@newcastle.edu.au

Judy Sheard
Monash University, Australia
judy.sheard@monash.edu

Angela Carbone
Monash University, Australia
angela.carbone@monash.edu

Chris Johnson
Australian National University
chris.johnson@anu.edu.au

ABSTRACT
There appears to be a reasonably common understanding about plagiarism and collusion in essays and other assessment items written in prose text. However, most assessment items in computing are not based in prose. There are computer programs, databases, spreadsheets, and web designs, to name but a few. It is far from clear that the same sort of consensus about plagiarism and collusion applies when dealing with such assessment items; and indeed it is not clear that computing academics have the same core beliefs about originality of authorship as apply in the world of prose. We have conducted focus groups at three Australian universities to investigate what academics and students in computing think constitute breaches of academic integrity in non-text-based assessment items; how they regard such breaches; and how academics discourage such breaches, detect them, and deal with those that are found. We find a general belief that non-text-based computing assessments differ in this regard from text-based assessments, that the boundaries between acceptable and unacceptable practice are harder to define than they are for text assessments, and that there is a case for applying different standards to these two different types of assessment. We conclude by discussing what we can learn from these findings.

Categories and Subject Descriptors
K3.2 [**Computers and education**]: Computer and Information Science Education – *computer science education*

General Terms
Measurement

Keywords
Academic integrity, computing education, non-text-based assessment

Koli Calling '13, November 14-17 2013, Koli, Finland.

http://dx.doi.org/10.1145/2526968.2526971

1. INTRODUCTION
Plagiarism and collusion are two major manifestations of academic dishonesty. Plagiarism occurs when a student uses the work of others without appropriate acknowledgement. Collusion is somewhat similar, but is distinguished by the fact that the 'others' are typically the student's own colleagues: collusion is essentially the sharing of work among students, whether the submissions be based on the work of one student or on a collaborative effort. With plagiarism, work purporting to be that of the student or group has too much in common with work that is typically in the public domain. With collusion, work purporting to be that of the student or group has too much in common with the work of other students or groups, often in the same class. Just how much is too much tends to depend on the context [37]. It is also difficult to be clear about what constitutes collusion in circumstances where students are encouraged to work together except when doing assessable work [5].

There appears to be broad agreement [14, 31, 34] that:

- Plagiarism and collusion are not good for the student, because students who plagiarise or collude are failing to practice the academic skill of assimilating the ideas of others and using them as the basis of one's own ideas. Rather, they simply echo the ideas of others with no evidence of assimilation or even of understanding. These practices are also seen as diminishing the students' employment prospects: employers do not want to see their reputations put at risk by what they see as a form of irresponsible behaviour.
- Plagiarism and collusion are not fair on other students: when students choose to work hard for their marks – or, indeed, not to work hard and to accept a lower mark – they feel aggrieved when other students attain good marks for submitting somebody else's work.
- Plagiarism and collusion are not good for the institution, as graduates who have side-stepped the learning process may not perform well in the workplace, reflecting poorly on the institution and on the discipline.
- Plagiarism and collusion are not good for the education system as a whole because they suggest that the system is willing to produce graduates who have succeeded not by independent thought and analysis but by finding the work of others that has some bearing on the subject at hand and presenting it as their own.

The literature of academic integrity leans heavily towards plagiarism of prose text. This focus is sometimes explicit, but more often implicit. Books on avoiding plagiarism [5, 19, 29]

focus almost exclusively on text-based plagiarism, with references to such concepts as paper mills, literature, using other people's words, translating foreign articles, and so on. Suggested ways of avoiding plagiarism include learning to paraphrase and learning how to synthesise the words of multiple authors. Referencing guides explain how to place directly copied text into quotation marks and reference it appropriately. All of this is done with little apparent recognition that text is not the only form of work that can be plagiarised.

In relation to prose text, students are generally less likely than academics to recognise certain practices as plagiarism or collusion [4, 12, 33]. Moreover, many students remain confused about definitions of plagiarism and collusion and expectations regarding academic integrity [17, 26].

While there is much work on educating students to avoid plagiarism and produce work that is clearly their own, detection of similarity remains a cornerstone of practice in academic integrity [5, 9, 11, 39]. So long as there are some students who are willing to plagiarise, and so long as this is seen as inappropriate, some academics will apply techniques to help them determine whether the work they are assessing is plagiarised. Within the realm of text-based assessment there are many standard tools to assist in the detection of similarity, tools such as Turnitin and AcademicPlagiarism.

Yet the higher education system includes numerous academic disciplines in which prose text is not the principal medium of assessment. Art students and design students are required to produce images as part of their assessment; mathematics students are required to construct mathematical proofs and derivations; music students are required to write musical compositions; computing students are required to write computer programs and to construct databases and spreadsheets; architecture students are required to produce plans and drawings. These forms of assessment are all dramatically different from prose text. Furthermore, none of these forms of assessment are amenable to the similarity detection of the standard tools such as Turnitin. Turnitin cannot tell whether two computer programs, two databases, two images, or two musical compositions have too much in common to be considered as distinct pieces of work.

Furthermore, while there appears to be broad agreement on the nature and inappropriateness of text plagiarism, academics and students in the non-text-based areas do not necessarily regard the use of others' work in the same light as do academics and students in text-based areas. In a world where re-mixes, mash-ups, re-use of computer code and other combinations of existing work are increasingly accepted and valued as legitimate professional and creative practice, some authors question the value of insisting that the work produced by students must be substantially or entirely original [18, 21].

Further still, the types of academic dishonesty tend to vary from field to field. It is possible that design students lacking inspiration will tend to base their work on an image found somewhere in the public domain, whereas students struggling with computer programming might be likely to borrow and appropriate the work of their more capable colleagues, or to work in inappropriately large groups to produce a joint solution to the problem at hand.

2. BACKGROUND

Some types of assessment items in computing involve text, but of a type that is not amenable to the standard techniques and tools used for similarity detection in prose text. Computer programs, for example, generally have a large text component; but that component is written in a computer programming language, not in a natural language such as English, and the same similarity criteria do not apply. Formal structures such as those found in computer programs are not amenable to detection on the basis of the percentage of textual similarity [13, 14, 18]. Students quickly learn that two computer programs look quite different textually if they have different variable names, different comments, and different spacing, even though they are to all intents and purposes the same program. For this reason, students copying one another's computer programs tend to change the variable names, comments, and spacing, in the hope of evading detection [3, 23, 25]. However, the similarity between the programs is actually evidenced by their logical structure, and such copying tends to be readily detected.

Other items for assessment in computing include databases and spreadsheets. Although these items have some text within them, they cannot in any way be described as text documents, they cannot meaningfully be converted into reader-friendly text documents, and again the points of similarity between two such items are far more likely to be in their structure than in their textual content.

Because copying and collusion are rife among students of computer programming [3, 7, 10, 13, 15, 36, 37, 41], a large number of similarity detectors have been created [1, 3, 8, 23, 30, 40]. Unfortunately, many of these detectors work with programs in just one programming language, and programming is taught in many different programming languages. There are some similarity detectors for multiple programming languages [3, 32], but their adoption appears to be far less wide than that of, say, Turnitin in the realm of prose assessment [5]. In other areas of computing, such as spreadsheets and databases, we have found no similarity detectors. Similarities between submissions are detected by eye, if at all.

There are computing academics who do not check for plagiarism and others who pay no heed to inappropriately similar submissions even when they notice them [7, 10, 18]. Anecdotally, they suggest that as the students are likely to be working collaboratively when they graduate and find employment, it is not inappropriate to do the same when they are studying. Others fiercely seek out similar submissions, insisting that the mark given to an individual should be for work carried out by the individual. This range of diverse academic opinions and behaviours has not been explored in any systematic way. Computing degree programs are almost all professionally accredited, and relate their development of student behaviour to the computing industry; but the expectations of good professional practice in acknowledging the work of others are not reported, and the codes of ethics and professional practice in computing do not cover this.

We aim to find answers to a number of specific questions:

- How do academics and students perceive academic integrity in regard to non-text-based computing assessment items?
- Are there assessments for which academics and/or students think that every answer is unique, so copying is acceptable so long as one personalises the copy?

- Are there areas in which academics and/or students think that there is only one correct answer, so copying cannot be detected?
- What steps do academics and students take to ensure that academic standards are adhered to?
- What do academics do to detect similarities that might suggest academic misconduct?
- How serious is academic misconduct considered to be and how is it dealt with?
- To what extent do computing academics believe that university policies for academic integrity are adequate for non-text-based assessments?

3. RESEARCH APPROACH

We conducted focus groups of computing students and academics at three Australian universities in late 2012 and early 2013. Three focus groups were made up of 12 students studying information technology, business systems, commerce and educational technology. Three focus groups of academics comprised 18 teaching staff. Participation in the focus groups was voluntary, with participants responding to posters or email advertising the research or to direct approaches by researchers. There is no assurance that the perceptions of participants are representative of other students and academics at these or other institutions. The staff and student focus groups were held separately, and no teachers were present at the student focus groups, as their presence might have constrained the students' responses.

The purpose of the focus groups was to inform the design of a broad survey to be conducted subsequently. The focus groups were based on an indicative set of questions, with the facilitators encouraged to ask exploratory follow-up questions when the discussion suggested so doing.

The focus groups were recorded and transcribed, and the transcripts were corrected and shown to the participants in case they might feel there were any egregious errors in the transcriptions.

Researchers used a directed or deductive content analysis methodology, using the extant literature to identify key concepts that were used to develop initial coding categories [20, 28]. Further categories were developed from the data using an inductive process. This approach combines the benefits of 1) using insights from the literature to validate or challenge previous findings; and 2) maintaining the flexibility to incorporate new insights directly from the data.

The initial analysis included listening to the tapes and reading the transcripts. The transcripts were then coded using the initial categories. Data outside these categories were identified and subsequently examined to determine whether they suggested new categories. Finally, the categories were reduced to themes, and the responses listed under each theme were analysed for concordance or conflicting views.

4. FINDINGS

The analysis identified five major themes: the level of understanding of academic integrity issues; perceptions of the importance of academic integrity; the complexity of computing compared with text-based situations; the processes of detecting and dealing with breaches; and, at an emotional level, the impact on relationships within institutions. This section explores these themes in relation to the research questions.

Based on the premise that most universities strive to inform their students about the requirements of academic integrity, but that the information provided is highly skewed towards the written word, the focus groups began with a discussion on academic integrity in essays and similar assessments.

4.1 Perceptions of academic integrity in regard to essays

Discussion around plagiarism and collusion relating to essays revealed considerable variation in the level of understanding of participants. Staff and students agreed that taking other people's words or ideas without acknowledgement constituted plagiarism. This includes 'copy and paste' and paraphrasing without referencing. However, a number of students thought that they needed to reference only when using a direct quote, including one who stated:

> 'If you're adding your own ideas in it I think it makes it yours because you're not directly like using their ideas. You're writing it differently.'

Plagiarism was viewed as a serious matter. Although the students might not share the accepted understanding of what plagiarism is, they know that it's not a good practice.

> 'It defeats the purpose of being at university.'

> 'Well it's not proving your own ability so it is extremely important.'

Among the students, few were familiar with the concept of collusion. Once it was defined for them, they had difficulty imagining what collusion might mean in the context of an essay-type assessment. One group concluded that it was not really possible to collude in writing an essay unless the students used exactly the same words. They felt that it was acceptable to show an essay to a friend and get them to correct any mistakes, including correcting the ideas and pointing out that the student had misinterpreted the question.

> 'Cause even if they do help, like they tell you that your thing is wrong and you suddenly have a change of heart and you agree with them. When you rewrite it it's in your own words. It's not in their words. So it's not copying.'

4.2 Perceptions of academic integrity in regard to computing assessment items

Perceptions about plagiarism and collusion in computing were shaped by opinions about seriousness, levels of understanding, the additional complexity (including requirements that varied by assessment), and fuzzy boundaries.

While computing assessments are many and varied, including databases, spreadsheets, design diagrams, and more, most of the discussion in the focus groups clearly revolved around computer programming assessments.

Students and academics stated that sources need to be acknowledged.

> 'Someone else's code that you've got to acknowledge where it's come from, and both document that in the code and acknowledge it in any formal, prose-form documentation that they produce.' (Student)

'...if there is some or all of their code that is not authored by them it has to be acknowledged.' (Academic)

The focus groups revealed that both plagiarism and collusion are issues, and that in the area of computing, collusion may be more of an issue than plagiarism, a view that supports the findings of Culwin et al [10]. Students spoke extensively about assisting friends with code if they were stuck, pointing them in the right direction, and comparable practices. They also discussed using libraries and message boards as a normal practice in computing.

Plagiarism in computing

As with essays, there was general agreement that plagiarism is copying or using someone else's ideas or work without acknowledging it. Some participants made no clear distinction between plagiarism and collusion: for example, some specific actions mentioned as forms of plagiarism were paying someone else to write code and taking someone else's USB device and copying their work from it.

However there were some issues that participants thought led to differences between plagiarism in essays and plagiarism in computing. First, there was the tradition of learning from the community, in which programmers adapt and learn from the code of other programmers. Second, students felt uncertain about whether they are permitted to reuse code they have previously developed for another purpose. There are certainly some programming courses in which code reuse is accepted as good practice [13, 22]. Third, there was the difference that while a reference in an essay is obvious when reading the essay, there is no way to make a reference in a program obvious when running the program, and reference guides provide scant information on how to reference code [13]. Of course the reference can be included in the code, but that will seldom be seen by users of the program. One student commented:

'…there's no formal way or clear way to do that. We've talked about writing comments in code to say where these come from but that's certainly not a standard way of doing that.'

Differences such as this led to a general feeling that it is much more difficult to establish the boundaries than it is with text-based assessments.

The fact that most university policies and plagiarism modules are largely silent on non-text-based assessments such as computer code has contributed to a situation where students are unclear about what practices constitute plagiarism [22]. Students who participated in the focus groups stated that they had received little specific education or guidance in relation to computer code. Therefore, they relied on what some referred to as their own ethics for guidance:

'We're not saying what we do is right, we're saying that's how we do it. We run it based on our ethics'

'We're just going on ethics and they haven't expressly told us how much of our code we can't reuse from someone else or where we can't get it from'

Possible consequences of such approaches were evident in one of the student focus groups where there was a lengthy discussion about individual students' abilities to determine intuitively whether practices constituted plagiarism or collusion.

'And in this it always comes back to ethics. If I feel like I've done something that's my work and … I don't feel like I've plagiarised it, then the chance is I haven't and even if I have then it's probably, like I could probably prove that I haven't, sort of thing. Like because of the amount of steps that you go through to do work and like no program has like one version of a program.'

Among students there was a feeling that it was acceptable to use other people's ideas but not to copy code.

'If you start copying their code, how they got their idea to work, then I think you've crossed the boundary.'

'It's still for me fuzzy 'cause there's an emphasis on peer learning as well. It's hard to know where help stops and plagiarism starts, I think.'

However, opinions varied significantly between individual students. Some felt that they could only be said to have copied code if they copied whole programs or whole websites, and there was a great deal of discussion about the acceptability of taking smaller pieces of code, and about how much of one's own work needed to be added to a piece of code in order to consider it one's own.

'…I just Google and I implement and I feel like because I had to implement it and stuff like that and make it work within my code and add things to it, it becomes mine.'

'Plagiarising code: it's just copying the exact code and pasting it into your own work.'

Confirming previous research [7, 13, 15, 22], the focus groups revealed that the academics' perceptions of plagiarism were generally somewhat stricter and more consistent than the students' perceptions For example, one academic described as a myth the widely held view of students that it was not plagiarism if they changed the code by 10 or 20 percent.

Notwithstanding this, the academics conceded that the boundaries remained blurred.

'…there's a difference between using something as a reference and actually copying something outright… I think it's just very hard to define.'

While previous research has established that students are likely to have a more liberal view than academics of the acceptability of certain practices [13, 15, 36], at least one study found some scenarios that students identified as unacceptable whereas teachers agreed they were acceptable [38]. In a similar vein, one academic in our focus groups was concerned that students are sometimes mistaken in their belief that something is plagiarism.

'I've actually had the reverse experience of plagiarism, of trying to get students to understand that it is actually quite legitimate to use existing code and further develop that, rather than having to start from scratch… there is always some of this, that I can't use someone else's work because that would be plagiarism.'

Collusion in computing

The focus groups revealed that there are different perceptions of what constitutes collusion between the two groups and also within the groups. This was further complicated by the fact that what constitutes collusion varies from one assessment to another, based on the assessment specifications.

The issue for academics revolved around the educational objectives of the particular assignment, although it was generally agreed that students were probably not entirely clear

on this point and that academics could do a better job of explaining why they wanted students to complete the work individually.

While academics often permitted some level of collaboration, there was considerable variation in what was acceptable. Some academics said that students could discuss the assignment but could not collaborate on the development of code.

> 'I encourage students to collaborate and brainstorm right down to the point where they've discussed details about the assignment but they're writing their own code…if it's about helping a friend to work out how methods are used, then do it without the context of the assignment.'

One academic indicated that he would be prepared to allow students to work together on an individual assessment as long as they declared it.

> 'I was trying to think of a model where we didn't make it illegal for them to work together, because as far as I'm concerned if four of them work together and they turn in a good assignment, well, they've learnt something. It doesn't worry me.'

However, without further safeguards such as interviewing students and ensuring that they know how the code works and can explain how it was developed, this approach would not necessarily ensure that all students had contributed to the assignment and met the learning objectives.

Students generally had a more liberal view of the level of support that was permissible. In some instances this included obtaining assistance with coding from fellow students and posting code on message boards to seek assistance. These methods were seen as legitimate so long as the answers pointed them in the right direction rather than supplying the code.

> '…no one ever says "here's the code to make it work". They always say "Have a look at this, it's because IE does this or Chrome does this to display it".'

Some students adopted a more nuanced stance, distinguishing between the legitimacy of these two approaches.

> 'No, the difference is if you post your question online it will be their solution. And you're copying their solution. But when your friend is checking, he'll tell what's wrong and you have to change it yourself, so that's different.'

A small minority of students thought that collusion occurred only in extreme cases such as when two assignments were exactly the same. One student stated that in situations where students could not get code to work they had no alternative other than to get help from somewhere:

> 'Like if you're stuck on like code or something, and like you just don't know how to fix it, if you don't seek help you can't finish the assignment. You're going to have to get help.'

Student misconceptions about what constitutes plagiarism and collusion, combined with the difficulty that both academics and students mentioned in determining the boundaries of acceptable and unacceptable practices, point to the need to clearly define acceptable academic practice and educate students.

4.3 Is unreferenced copying legitimate so long as one personalises the copy?

In general, students were more likely to agree with this proposition than academics, although both groups acknowledged that the boundaries are blurred. Both academics and students expressed the view that it was permissible to copy commonplace or trivial pieces of code without referencing. Similarly, students sometimes felt that it was not necessary to reference some code if they made considerable changes to it.

> '…like more complex but like still generic… Even stuff like that I just Google and I implement and I don't like, I feel like because I had to implement it and stuff like that and make it work within my code and add things to it, it becomes mine. Whereas if it was like the variables were set and I just copied a chunk, pasted a chunk and then left it, it would be closer to plagiarism.'

To tease out where the boundaries are, students were presented with a scenario where a student copied methods from the web without referencing them. Students generally felt that this was an acceptable practice, and one student justified it in these terms.

> 'You have to change the code and make it fit in the context of your assignment so I wouldn't think it would be plagiarism.'

4.4 Is there sometimes only one correct answer, so copying cannot be detected?

Academics agreed that students do think they can get away with copying, particularly in relation to computer code. However, they had differing views on why students had this perception. The first position was that students think there is only one correct solution so all the assignments will be identical

> '…the ones who do the totally copying seem to assume that "there is only one right answer so nobody will know I've copied".'

An opposing view was that students think teachers simply run the code to see if it works, and are not aware that their work would be compared to that of other students.

> 'Well I wondered if they don't realise some of them that we look at their code. I think maybe they think we only run it. Because the cheating is so bad when we catch it, you know I'm insulted that they think I'm so stupid.'

A third view was that there was a combination of these two reasons.

> 'They'll at least know from the unit tiff, they think all we do is run the unit tiff and we never look at their code and they also think… a lot of them think that if we do happen to look at their code, well everybody is going to have the same code anyway.'

One group of computing students stated that on the rare occasions where the question would lead to everyone having similar or identical code, teachers would ensure that the task was completed in class under supervision so there was no real opportunity for students to collude or copy one another's code. They contrasted this with individualised assessments where it was unlikely that they would have similar code, so high levels of code similarity would alert academics to the possibility of copying.

'…with most of the open assessments like with make it up, there's infinite answers. So if they were so specific, and because there's infinite answers, if they were within a range, like, this range, you can tell.'

However the ability of teachers to pick up similarity in such situations was seen to be severely constrained if there were large classes.

'…if there's a large class of say 100 students doing one assessment, just an idea, how would they remember after marking 100 assessments, which ones are similar? And you can't just put it in the computer.'

4.5 What do academics and students do to ensure standards are adhered to?

Academics outlined numerous strategies they employ to ensure that students adhere to academic integrity standards. The major strategies involve education, monitoring work in class, and viewing work in progress, and are well aligned with recognised practice as summarised, for example, by Carroll [5].

In addition to the general information given to students by the university, such as academic integrity modules, some academics outlined methods they use to ensure that students understand what plagiarism and collusion are and the parameters of acceptable practice for the degree, the course, and specific assignments. Frequently, supplementary written information is provided in course outlines and/or specifications for each assessment. Some academics also provide specific courses, lectures or tutorials on academic integrity.

'I present my lecture in the first or second week. I make it very explicit about what they cannot do. Then if a student gets caught they know they've been caught.'

A common practice is the provision of specifications with each assignment so that students understand what is permissible for the particular assessment task in line with the learning objectives.

'a very specific indication in the assignment specification as to the expectation… that it is their own work, and that anything that's not their own work is clearly referenced as such.'

Academics also employ a number of other strategies to reduce opportunities for breaching academic integrity, strategies such as designing individualised assignments, viewing work in progress, and awarding marks for work in progress to encourage students to apply sufficient effort in this area.

'…a week after I give out an assignment they have to hand in some pseudocode.'

Similarly, monitoring progress in class made it easier for staff to keep track of individual students and reduce the likelihood of cheating.

'If you actually want students to generate code you don't send them away to do it, you have them sit down and generate code in front of you.'

Some academics permitted collaboration within the specification of assessment tasks, with safeguards to ensure that learning objectives are met. One approach permitted students to collaborate and hand in one assignment, and then each student was interviewed to ensure that they understood the work. Another academic outlined a similar approach:

'The way I do it is by producing assignments, the collaborative parts were small portions, so that even if they do collaborate then I have these lab tests or quizzes or other things and an exam where I can check that they've actually learnt what they are meant to learn.'

Students expressed the view that they received little appropriate instruction, either from academic integrity modules or similar tasks to be completed by all students, or from special lectures, tutorials, or assessment guidelines from computing academics. This perception was in stark contrast to the extensive discussion by academics detailing all their attempts to help students understand and adhere to academic integrity standards.

The discussion of the steps students themselves take to maintain academic integrity standards was succinct and centred upon referencing when they copied code and not sharing code with other students. Students also acknowledged that their opportunities to plagiarise or collude were reduced by the practices adopted by academics, as outlined above, such as being given unique assignment tasks; submitting work in progress as well as the final assessment item; working on assessments in class; being required to demonstrate how they developed their work; and attending interviews to explain how the code worked.

4.6 How do academics detect similarities that might suggest academic misconduct?

In line with much of the academic integrity literature [2, 6, 27, 29], the consensus among academics was that prevention is more important than detection, although it is necessary to take action when breaches occur.

'If it is difficult to detect plagiarism in these non-text things then the solution is not to find better ways of detecting it, but to avoid the problem.'

Academics indicated that it was more time-consuming and more difficult to detect plagiarism with non-text-based items such as computer code than was the case for text. While some academics said it was easy to detect copied code, others said it could be detected only 'with difficulty'. It was mentioned more than once that detection was easier for small student cohorts where one person is more likely to mark all of the assignments.

There are a number of strategies that can be used to detect breaches of academic integrity. In the first instance academics rely heavily on knowledge of the abilities of their individual students. Breaches were frequently identified when students submitted work that was better than expected.

'If I get an assignment that's better than I would expect from that student, then I would be looking at it closely.'

A second method involved the use of technological solutions, typically involving detection tools such as Google, Google Image, Plaggie, JPlag, MOSS, and others.

Some academics indicated that they used techniques such as looking for white space in the program that should not be there. Similarly, variable names that did not seem appropriate could be a pointer to copied code in which students had simply changed the names of variables in an attempt to avoid detection.

'I still use the detect tool because it looks at the words and comments as well, it looks at identifiers and it doesn't attempt to look at program structure. And it sticks out like a signature, particularly if they've taken stuff on the web and

they haven't understood it and they haven't changed the identifiers…it is a fairly subtle signature sort of tool.'

However, detection tools are less effective with computer code than with text due to the more constrained nature of computer code, and the consequent increased probability of matching segments in truly independent programs. Compared to the situation with text, using the percentage of similarity to detect plagiarism is not as straightforward in the case of code for many reasons, including the formal structure of the language [7, 8, 25] and the fact that students have similar levels of experience and are using the same textbook [25].

One student participant in a focus group recounted a situation where researchers asked students from different classes to submit their assignments and then analysed the code for similarities. There was a 90 per cent match between this particular student's code and that of another student even though the students had never met or associated in any way. While anecdotal, this raises the possibility of code detection tools suggesting that students have copied code even when they have worked independently.

In some instances staff indicated that other students provided information on suspected breaches by their fellow students.

'…we actually got emails from some students saying that other students were asking questions on some sort of cheat sites, and they were literally the questions that were being put to them on the assignment.'

Testing a student's knowledge of code (as mentioned in the previous section) is also a potential detection method as well as a reasonably effective deterrent.

4.7 How serious is academic misconduct and how is it dealt with?

Focus groups were asked how serious an issue plagiarism and collusion are. Due to some confusion over the wording, some focus groups interpreted this in a normative sense and others responded in relation to prevalence. Breaches of academic integrity were seen as a serious issue since 1) they involve the unethical practice of taking credit for work that was not completed by the individual student; 2) they have the potential to devalue qualifications from particular universities, or universities in general, if graduates enter the workforce without the necessary skills; and 3) they will eventually have negative consequences for the individuals involved since they have not developed the skills they need. Student comments included:

'…so people will think [name of university] standards aren't very high and … it devalues the value of our degree'

'…people will find out you're not that good anyway.'

Academics' assessments of prevalence varied and were very general. However, breaches of academic integrity were generally thought to greatly exceed the number of detected cases.

'I always have a guilty feeling that it's more prevalent that I'm aware of and so I feel like that I'm being a bit naïve and dumb. It's the tip of the iceberg here.'

In relation to the issue of how breaches of academic integrity are dealt with there was a broad difference of knowledge between students and academics. Students appeared to be genuinely concerned about the prospect of committing an inadvertent breach of academic integrity and were aware that the consequences could include expulsion from the university. One student stated that the reason for not cheating was fear of being caught and punished. However, students were not generally aware of situations where breaches had been detected or of the consequences for the students involved. While this could be because few instances of breaches were being detected, some students surmised that it was likely to be due to privacy protections at the university.

On the other hand, academics recounted specific instances of detecting breaches and how they were dealt with at their universities. Some academics indicated that students were first spoken to by the lecturer as part of the investigative process. If students admitted to cheating the consequences might include a warning or a zero mark for the assignment and/or a note in their personal file. The case was escalated if the student denied the breach and was unable to provide a satisfactory explanation.

'…it's only the ones where they're in dispute that tend to escalate to higher authority.'

In other universities there was no flexibility since university policies stipulated that academics must immediately report any suspicions to a nominated academic conduct officer.

The prevailing view of academics was that breaches are difficult to prove and that pursuing breaches was time-consuming and resource-intensive. Some stated that they received inadequate support from their university, both in terms of resources and because the university was sometimes too lenient with students.

'…you're not supported higher up in the system and that happens a lot at our institution.'

4.8 Is the university policy adequate for non-text-based assessment items?

Academics were unanimous in the view that university policies related predominantly to text-based situations and there is not enough information on what is expected in non-text-based situations. Academics identified the need for further development to incorporate non-text assessments.

'… I think when it comes to interpreting the policy with regard to non-text it doesn't really stack up… our students … are required to do an academic integrity module … and it really doesn't address at all things like images and computer programs and databases, and diagrams, and mathematical proofs, and all the rest of it.'

However, some academics also noted that if there was a rigid policy developed for non-text-based assessments it could leave them without the flexibility they now enjoy.

Due to the inadequacy of whole-of-institution approaches that concentrate on text, some other initiatives have been developed for non-text-based assessments and practice as mentioned previously: special lectures / tutorials for students to make sure they are aware of how academic integrity relates to non-text-based assessments; and specifications for each assignment outlining what can and cannot be done.

Academics stressed that there is a need for consistency in policies pertaining to non-text-based assessment items, but this is complicated by the fact that there are different requirements related to educational objectives between faculties, degrees, courses, and even within courses. Therefore, it is necessary to have very clear guidelines at all levels and to ensure that

students understand what is required of them and why – what the learning objectives are that these requirements contribute to.

> '…how do we inform students that this difference in approach is expected? That in one course we've got "you are expected to implement your own, bar very simple, almost year one things". But in another you are expected to make optimum reuse of existing things?'

Both academics and students expressed the view that there is a case for applying different standards for non-text-based computing assessments. One academic commented:

> 'I think that policy should give just general guidelines, it depends on the discipline. Like Computer Science and the Law school in terms of plagiarism are different.'

Similarly, students pointed to what they considered to be significant differences for non-text situations.

> '…when you're just getting chunks of code, you're not really copying, you're just getting concepts behind it and interpreting it to yourself.'

Moreover, some students felt that guidelines would need to vary by the type of non-text-based assessment involved, which would result in different guidelines for programming, web design, and images, for example. They also thought that establishing rules would be problematic due to the rapidly changing nature of computing.

> '…that's why it's just going to be so grey and it's got to be a process. It can't be a rule book.'

Another issue that emerged with both academics and students was the tension between academic expectations and the standards that apply in the workforce. One student noted that, in contrast to the academic situation,

> '…in the real world in IT stuff there's like exceptions and there's different rulings based on different things which makes them a bit more complex.'

One student expressed the more extreme view that there was nothing wrong with collusion because it was the normal way of working in professional world.

> 'I don't understand collusion myself because like basically collusion is like working together, right, and 99% of the world is made up of people working together to get things done. So in my eyes I'm not really for rules against.'

In the focus groups with academics there was a difference of opinion as to whether there should be two distinct standards (academic and commercial) or whether a common standard should be developed. Academics also debated whether students should be educated to understand commercial standards while at university in order to prepare them for the workforce. Comments from academics included:

> '…to respect real practice in our discipline and to make education match what people are going to find in the workplace later, which has to do with reuse, teamwork, and so on.'

> '…in their professional life….it is perfectly legitimate for them to go to Google and look on the web and look at other companies that have solved similar problems.'

5. OTHER ISSUES ARISING

Other issues that emerged from the data related predominantly to deficits in students' understanding of the importance of academic integrity. Academics discussed a number of issues that were almost entirely neglected in the student focus groups. First, academics stressed that one prominent reason for developing and implementing standards was to ensure that the learning objectives of the course were met. If students were conscious of the learning objectives of assignments they would be more likely to be engaged and honour the assessment specifications.

> 'So that they can engage with the kind of educational objectives, rather than misunderstanding them and then behaving in perverse ways because they think we're assessing them on basic learning stuff.'

While the majority view was that communicating the educational objectives of assessments could be improved, there was a perception that practices varied considerably between universities, and one academic thought that current practices at their university were adequate.

> 'At my university we have a legal obligation to do so. We have to. We have an academic integrity statement that we have to make at the beginning of the course, and before every assignment; and then on top of that we like to – it's not legally required, but we like to articulate the learning goals of the assignment as well, to make it clear for the students.'

A second area that academics identified was facilitating a greater understanding of why students should cite sources. This included principles of scholarship:

> '… the provenance of your work, why is this a reliable thing to use, what's the integrity of this, so there's traceability…

In addition, a number of academics mentioned that from a pragmatic point of view students should be made aware of the fact that they would obtain higher marks by citing their sources.

The third issue that academics raised related to cultural differences between international and domestic students. Cultural differences have been identified in studies of computing students' perceptions of plagiarism [22]. Some instances related by academics in the focus groups included international students working in groups despite being instructed to work alone or copying solutions when they were expected to develop their own solutions. In another instance, Chinese students justified copying solutions from a lecturer by saying:

> '… what are we supposed to do, because in China, in Asian culture … when there is a mentor, what are we taught, I should follow that. If the teacher says do it this way we should just follow the teacher.'

This suggests that there is a need to ensure that international students are educated to understand the requirements of studying in an Australian university.

Finally, the academic focus groups postulated reasons why students breach academic integrity guidelines. Major reasons included: looking for shortcuts; not being able to see the relevance of the work; high expectations of teachers; time pressures; and the myth that it is not plagiarism if they change it by 10 or 20 per cent. Previous studies involving computing

students have identified similar reasons for cheating [10, 13, 36, 37, 41].

6. CONCLUSION AND FURTHER WORK

The focus groups revealed that plagiarism and collusion are viewed as serious issues by academics and students, but many students lack a holistic appreciation of the importance of academic integrity. Perhaps as a consequence, some students revealed a fear of inadvertently breaching the academic integrity rules. These issues also impact on the relationship between students and the academics who police academic integrity.

The overwhelming view of the academics and students who participated in the focus groups was that there is a substantial difference between computing assessments and text-based assessments, and that the boundaries of acceptable and unacceptable practices are much more difficult to define for computing assessments than for essays.

The focus groups confirmed that academic integrity policies in these institutions remain heavily prose-based and guidelines for non-text-based assessments are underdeveloped. Both the academics and the students supported the view that there is a case for different standards of academic integrity for non-text-based assessments, but also stressed the need for consistency in policies pertaining to such assessment items.

Academics understood the need for clarity and consistency in guidelines. However, policy development is complicated by differences in perceptions of academic integrity, both between academics and students and within each group, as well as different requirements related to educational objectives between faculties, degrees, courses, and even within courses. Furthermore, if specific non-text-based policies were devised, they would need to be carefully incorporated into the existing policies, which relate mainly to prose-based items.

Initiatives that academics have developed to fill the current void and inform students of requirements for non-text-based assessments include delivering special lectures or tutorials on academic integrity as it relates to these assessments, and issuing detailed academic integrity guidelines in the specifications for each assignment. Despite these initiatives, student awareness remains low, echoing findings of knowledge or perception asymmetries from other research in computing [13, 15, 35]. The chasm between academics' attempts to inform and prevailing student understandings warrants further research to determine why these attempts to inform students are currently ineffective, and to develop improved strategies.

Previous research has emphasised the need for a holistic approach to academic integrity that incorporates education, clearly defined policies which are adhered to, and attempts to inculcate antipathy to breaches of academic integrity [14, 15, 16, 34]. Joyce et al [24: 195] state:

> 'we must ensure that we give equal emphasis to the necessary ingredients: assessment design, education of staff and students, detection tools, academic integrity policies, and disciplinary processes.'

Greening et al [16] advocate integrating ethics education into computing courses rather than teaching it in stand-alone units. Innovations such as these would allow a discussion of ethics in relation to the real situations likely to be faced by students in their academic and professional lives.

A possible direction for future research that would contribute to this holistic model involves developing communication strategies that enunciate why students are expected to conform to academic integrity requirements. An emphasis on the learning objectives of assessment tasks might engender a greater commitment from students to engage with and abide by the policy.

In relation to breaches of academic integrity, the majority of participants agreed that collusion was more prevalent than plagiarism amongst computing students.

Academics use a number of methods to detect plagiarism and collusion, including similarity detection software and inspecting code for similarities.

Policies for dealing with breaches varied between institutions. Some academics have a level of discretion when suspected breaches are detected, while others are required to escalate breaches immediately.

While these focus groups involved staff and students at three universities, we are now conducting a survey across a far larger number of institutions within Australia. This survey will ensure a more representative perspective, and will give us a feel for the prevalence of the beliefs that we have identified.

At the heart of this work is the question of whether the same academic integrity guidelines should apply to non-text-based computing assessment items as to essays. The focus groups suggest that there might be a case for different standards, and we shall look with interest to see whether the survey confirms this belief.

If the feeling is that the same guidelines should apply, but that this is not currently happening, we would hope to develop tools to help academics and students in computing understand and apply the academic integrity guidelines. On the other hand, if the feeling is that the guidelines were developed for essays and similar assessment items and are not appropriate for non-text-based assessment items, we would aim to develop more appropriate academic integrity guidelines and to argue for their adoption.

7. REFERENCES

[1] Ahtiainen, A, S Surakka, and M Rahikainen (2006). Plaggie: GNU-licensed source code plagiarism detection engine for Java exercises. Sixth Baltic Sea conference on Computing Education Research (Koli Calling 2006), Koli, Finland 141-142.

[2] Alam, LS (2004). Is plagiarism more prevalent in some forms of assessment than others? 21st ASCILITE Conference, 48-57.

[3] Arwin, C and SMM Tahaghoghi (2006). Plagiarism detection across programming languages. 29th Australasian Computer Science Conference, 277-286).

[4] Brimble, M and P Stevenson-Clarke (2005). Perceptions of the prevalence and seriousness of academic dishonesty in Australian universities. The Australian Educational Researcher 32(3): 19-44.

[5] Carroll, J (2002). A Handbook for Deterring Plagiarism in Higher Education. Oxford Centre for Staff and Learning Development, Oxford, UK.

[6] Christe, B (2003). Designing online courses to discourage dishonesty. Educause Quarterly 2003(4): 54-58.

[7] 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.
[8] Cosma, G and M Joy (2012). An approach to source-code plagiarism detection and investigation using latent semantic analysis. IEEE Transactions on Computers, 61(3): 379-394.
[9] Crisp, GT (2007). Staff attitudes to dealing with plagiarism issues: perspectives from one Australian university. International Journal for Educational Integrity 3(1): 3-15.
[10] Culwin, F, A MacLeod, and T Lancaster (2001). Source Code Plagiarism in UK HE Computing Schools: Issues, Attitudes and Tools. Joint Information Systems Committee.
[11] Curtis, GJ and R Popal (2011). An examination of factors related to plagiarism and a five-year follow-up of plagiarism at an Australian university. International Journal for Educational Integrity 7(1): 30-42.
[12] de Lambert, K, N Ellen, and L Taylor (2006). Chalkface challenges: a study of academic dishonesty amongst students in New Zealand tertiary institutions. Assessment & Evaluation in Higher Education 31(5): 485-503.
[13] Dennis, L (2004). Student attitudes to plagiarism and collusion within computer science. International Plagiarism Conference 2004, viewed 28 March 2011, available at: http://www.plagiarismadvice.org/research-papers.
[14] Dick, M, J Sheard, C Bareiss, J Carter, D Joyce, T Harding, and C Laxer (2003). Addressing student cheating: definitions and solutions. SIGCSE Bulletin 35(2): 172-184.
[15] Dick, M., J Sheard, and S Markham (2001). Is it okay to cheat? The views of postgraduate students. ACM SIGCSE Bulletin 33(3): 61-64.
[16] Greening, T, J Kay, and B Kummerfeld (2004). Integrating ethical content into computing curricula. Sixth Australasian Conference on Computing Education, 91-99.
[17] Gullifer, J and GA Tyson (2010). Exploring university students' perception of plagiarism: a focus group study. Studies in Higher Education 35(4): 463-481.
[18] Hamilton, M, SMM Tahaghoghi, and C Walker (2004). Educating students about plagiarism avoidance - A computer science perspective. International Conference on Computers in Education, 1275-1284.
[19] Harris, R (2001). The Plagiarism Handbook. Pyrczak Publishing, Los Angeles, USA.
[20] Hsieh, H-F and SE Shannon (2005). Three Approaches to Qualitative Content Analysis. Qualitative Health Research 15(9): 1277-1288.
[21] Johnson-Eilola, J, and SA Selber (2007). Plagiarism, originality, assemblage. Computers and Composition 24(4): 375-403.
[22] Joy, M, G Cosma, J Y-K Yau, and J Sinclair (2011). Source code plagiarism – a student perspective. IEEE Transactions on Education 54(1) 125-132.
[23] Joy, M, and M Luck (1999). Plagiarism in programming assignments. IEEE Transactions on Education 42(2): 129-133.
[24] Joyce, D (2007). Academic integrity and plagiarism: Australasian perspectives. Computer Science Education 17(3), 187-200.
[25] Mann, S, and Z Frew (2006). Similarity and originality in code: plagiarism and normal variation in student assignments. Eighth Australasian Computing Education Conference, 143-150.
[26] McCabe, D L, KD Butterfield, and LK Treviño (2006). Academic dishonesty in graduate business programs: prevalence, causes, and proposed action. Academy of Management Learning & Education 5(3): 294–305.
[27] McCabe, DL, LK Treviño, and KD Butterfield (2001). Cheating in academic institutions: a decade of research. Ethics & Behavior 11(3): 219-232.
[28] Moretti, F, L van Vliet, J Bensing, G Deledda, M Mazzi, M Rimondini, C Zinnermann and I Fletcher (2011). A standardized approach to qualitative content analysis of focus group discussions from different countries. Patient Education and Counselling 82: 420-428.
[29] Neville, C (2010). The Complete Guide to Referencing and Avoiding Plagiarism. Open University Press, Maidenhead, UK
[30] Ohno, A, and H Murao (2008). A quantification of students coding style utilizing HMM-based coding models for in-class source code plagiarism detection. Third International Conference on Innovative Computing, Information and Control, ICICIC '08, 553-556. doi: 10.1109/ICICIC.2008.614.
[31] Popyack, JL, N Herrmann, P Zoski, B Char, C Cera, and RN Lass (2003). Academic dishonesty in a high-tech environment. SIGCSE 2003, 357-358.
[32] 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.
[33] Razera, D, H Verhagen, TC Pargman, and R Ramberg (2010). Plagiarism awareness, perception and attitudes among students and teachers in Swedish higher education – a case study. Fourth International Plagiarism Conference, Newcastle upon Tyne, UK.
[34] Riedesel, CP, AL Clear, GW Cross, JM Hughes, Simon, and HM Walker (2012). Academic integrity policies in a computing education context. ITiCSE Working Groups 2012.
[35] Sheard, J and M Dick (2011). Computing student practices of cheating and plagiarism: a decade of change. ITiCSE'11, Darmstadt, Germany, 233-237.
[36] Sheard, J and M Dick (2012). Directions and dimensions in managing cheating and plagiarism of IT Students. Fourteenth Australasian Computing Education Conference (ACE2012), Melbourne, Australia, 177-185.
[37] Sheard, J, S Markham, and M Dick (2003). Investigating differences in cheating behaviours of IT undergraduate and graduate students: the maturity and motivation factors. Higher Education Research and Development 22(1): 91-108.
[38] Stepp, M and B Simon (2010). Introductory computing students' conceptions of illegal student-student collaboration. SIGCSE 2010, Milwaukee, USA, 295-299.
[39] Sutherland-Smith, W and R Carr (2005). Turnitin.com: teachers' perspectives of anti-plagiarism software in raising issues of educational integrity. Journal of University Teaching & Learning Practice 2(3), accessed at http://ro.uow.edu.au/jutlp/vol2/iss3/10.
[40] Vandeventer, J and B Barbour (2012). CodeWave: a real-time, collaborative IDE for enhanced learning in computer science. SIGCSE 2012, 75-80.
[41] Vogts, D (2009). Plagiarising of source code by novice programmers a "cry for help"? 2009 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists, 141-149.