Quantitative Peer Assessment: Can students be objective?

Nicole Herbert
School of Computing,
University of Tasmania
Private Bag 100, Tasmania, 7001, Australia
Nicole.Herbert@utas.edu.au

Abstract

Team work can have a positive impact on student learning and commitment, but it is challenging to determine a method of assessment that does not require lecturers to involve themselves intimately with each team. Team members are often the best source of meaningful data, and as a result, lecturers are including self and peer assessment. One method of peer assessment is to have team members quantify their own contribution and that of team members. Concerns have been raised in the literature about distribution patterns with this method of peer assessment. An online peer assessment system has been capturing data from a capstone project course for three years with over 24 teams and 100 students each year. This paper analyses the following questions: do students take the easy option of equal distribution to avoid conflict, are students honest about their own contribution, are females treated fairly and are international students unfairly discriminated against.

Keywords: peer assessment, teamwork, capstone project.

1 Introduction

Rewarding a student with an individual grade is challenging in a capstone project course due to the team and project structure – the work products vary between projects and an individual's contribution can be hard to identify. As stated by Wilkins and Lawhead (2000) many instructors shy away from team project situations, partly because of the challenge to devise a way to assign grades to individual team members. Gates, Delgado, Mondragon (2000) identified the importance of structuring individual accountability to ensure that all members of a team contribute to the project concluding that team members are often the best source of meaningful data.

Software Engineering Project is a 26-week capstone program divided into two consecutive 13-week units; the students get two grades. The students undertake real industry projects suggested by local businesses in teams of 4-5 students, occasionally 3. Students form their own teams and choose a project from the list of suitable projects (Clark 2005). Table 1 summarises the team and student configuration data from 2004, 2005, and 2006.

	2004	*2005*	*2006*
Total Teams	27	25	24
Teams of 5 students	21	19	20
Teams of 4 (or less) students	6	6	4
Teams with gender mix	11	10	16
Teams > 2 of both genders	2	2	7
Teams all-domestic students	20	14	10
Teams all-international students	3	5	4
Teams with domicile mix	4	6	10
Teams > 2 of each domicile	1	3	3
Total Students	129	118	115
Females (mixed gender teams)	13	17	24
Males (mixed gender teams)	41	30	52
Males (not mixed)	75	71	39
Domestics (mixed domicile)	11	13	22
Domestics (not mixed)	96	65	47
Internationals (mixed domicile)	8	16	26
Internationals (not mixed)	14	24	20

Table 1: Team and student data 2004 , 2005, 2006

The number of female students has increased steadily over the years. The number of international students doubled in 2005 as students from China, who had completed the first 1 – 1.5 years of their degree in China, came to Tasmania to complete the final 1.5 – 2 years of their degree. All international students are temporarily in Australia to study and the majority are Chinese (71%).

In Software Engineering Project students receive a grade that reflects their input into the project. One of the main sources of assessment data is peer assessment. A range of tools are used to allow students to indicate their own contribution and that of team members, all of these are described in detail in Clark, Davies and Skeers (2005):

- Timesheets – an online version of the timesheets described by Humphrey(1997).
- Individual Contribution Reports – a personal report detailing their contribution to a work product. Each student also has to indicate agreement or disagreement (including explanation) with the reports written by each of their team members.
- Self/Peer Evaluation Surveys – students rate each team member (including self) on a list of behavioural characteristics of good team work.
- Work Product Pay Packet – each team member gives a quantitative opinion of how much each team member contributed to a work product.

The timesheets are assessed weekly. The other peer assessments are conducted at the conclusion of a major work product (eg design report). Once the work product has been assessed a meeting is held with the lecturer to receive feedback on the work product and discuss the peer assessment data.

An analysis of the numbers input into the Work Product Pay Packet (WPPP) is the focus of this paper. The approach is similar to that described by Hayes, Lethbridge and Port (2003) and Kennedy (2005) and the numbers are used similar to the way Brown (1995) used the numbers from the Autorating system. Clark (2005) gives a detailed description of the evolution of the tool and how it is used in the assessment process to determine a final mark. In first semester 2004 each student had to distribute 20 marks amongst their team members, since second semester 2004 the students distributed a virtual $100, using a work product pay packet form, Figure 1. They are told to distribute the amount between team members based on the quality and quantity of work contributed by each individual. In 2004 the WPPPs were used 7 times and in 2005 they were used 10 times and in first semester 2006 they have been used 3 times. There is a pay packet associated with each of the design reports (3 a year) and at least one for each software release (2 a year). Pay packets are also used to quantify contribution to documentation and marketing work products.

Kaufmann (2000), Layton and Ohland (2001) and Hayes et al (2003) all used a process of having students quantify the contribution of team members, but they expressed some concerns about undesirable distribution patterns. Kaufmann (2000) raised concerns about team members forming a pact to give equal amounts to avoid conflict, and concerns about individuals inflating their own performance. Hayes et al (2003) raised concerns that team members may "gang up" on another member (form partial team pacts). Kaufmann (2000) and Layton and Ohland (2001) both raised concerns about gender and ethnicity bias. Kennedy (2005) asked whether peer assessment was worth the effort, concluding that marks awarded to individuals based on peer assessment differ only slightly from equal allocation.

This paper focuses on the data contained in the WPPPs in 2004, 2005 and first semester 2006, and analyses the following questions: do students take the easy option of equal distribution, are students honest about their own contribution, are females treated fairly and are international students unfairly discriminated against.

	Pay	Comments
Jim	30	Both Claire and I undertook the majority of the release 2 implementation while the others concentrated on the documentation
Claire	30	
Ben	15	
Max	10	
Colin	15	
Total	**$100**	

	Jim	Claire	Ben	Max	Colin	Pay
Jim	30	25	30	30	25	**140**
Claire	30	25	25	25	25	**130**
Ben	15	15	15	15	15	**75**
Max	10	15	15	10	15	**65**
Colin	15	20	15	20	20	**90**

Figure 1: Work Product Pay Packet and Work Product Pay Packet Team Summary

2 Do students take the easy option of equal distribution?

As shown in Table 2, in 2004 the students gave equal amounts to all their team members 50% of the possible times (number of students 129 x 7 tests). In first semester it was 59% of the time, whereas in second semester it was only 44%. In 2005 the students gave equal amounts 43% of the time. Again there was a drop from first to second semester, but only 8%. In first semester 2006 the students gave equal amounts 23% of the time - a dramatic reduction on the previous two years.

In 2004, females gave equal amounts 37% of the time – considerably less than males, but this was reversed in 2005 and then reversed again in 2006. This indicates that equal distribution is not influenced by gender.

In 2004, international students gave equal amounts 70% of the time considerably more than domestic students. In 2005 this was reduced by 10% and in 2006 this was reduced by nearly 50%. The data does indicate that international students are significantly more likely to distribute amounts equally than domestic students.

There are a number of reasons why students may distribute amounts equally:

1. They genuinely believe the contribution was equal.
2. They are lazy or doing the form in a hurry.
3. They are trying to disguise their own contribution – be it too much or too little.

		N	*Possible*	*% of Possible*
2004	Females	34	91	37
	Males	419	812	52
	Domestics	345	749	46
	Internationals	108	154	70
	Semester 1	228	387	59
	Semester 2	225	516	44
	Total	453	903	55
2005	Females	92	170	54
	Males	415	1010	41
	Domestics	269	780	35
	Internationals	238	400	60
	Semester 1	279	590	47
	Semester 2	229	590	39
	Total	508	1180	43
2006	Females	13	72	18
	Males	67	273	25
	Domestics	33	207	16
	Internationals	47	138	34
	Semester 1	80	345	23

Table 2: Individuals who gave equal amounts to all team members

4. They disagree with peer assessment and refuse "to do the lecturers work for them".
5. They have a pact and believe other team members will also distribute equally.

Students are able to write comments on the WPPPs and many students do confirm that they mean to give an equal distribution because they believe this was the contribution pattern. Reasons 2, 3 and 4 are fairly easy for the lecturer to identify after reading the Individual Contribution Reports, Timesheets and the WPPPs from other members. The offenders are quizzed at the meeting with the lecturer and tend not be repeat offenders. Though students doing it for reason 4 can be hard to dissuade.

Teams that have a pact to distribute equally are a concern. Often teams form a pact to all contribute equally (which is great) and agree to do the distribution equally – sadly the contribution is not equal but members still distribute equally. There are indications that when a team makes a pact that it is hard to break – with 4 teams in 2004 and 3 in 2005 hardly wavering at all. Though in 2004 on only 5 (out of 26) pay packets and in 2005 on only 4 (out of 34) pay packets was the lecturer able to discern a different contribution by each member based on the Individual Contribution Reports indicating that these team members did do an equal contribution most of the time.

As shown in Table 3, in 2004 an entire team gave equal amounts to all their team members 22% of the possible times (number of teams 27 x 7 tests each). In 2005 it was reduced to just 15 % of the time. In 2006 it was only 6% of the time. Teams consisting entirely of international students are most likely to distribute equally. In 2004, the all-international teams gave equal amounts 76% of the time. In 2005, it was reduced to 58% of the time – but a staggering eleven times more often than the all-domestic

teams. In first semester 2006, the all-international teams gave equal amounts only 25% of the time – but all-domestic teams have been reduced to 0%. In 2004 there

		N	*Possible*	*% of Possible*
2004	Teams of 5	24	161	15
	Teams of 4	18	28	64
	All-international	16	21	76
	All-domestic	26	140	19
	Mixed domicile	0	28	0
	Total	42	189	22
2005	Teams of 5	26	190	14
	Teams of 4	11	60	18
	All-international	29	50	58
	All-domestic	7	140	5
	Mixed domicile	1	60	2
	Total	37	250	15
2006	Teams of 5	4	60	7
	Teams of 4	0	12	0
	All-international	3	12	25
	All-domestic	0	30	0
	Mixed domicile	1	30	3
	Total	4	72	6

Table 3: Entire team gave equal amounts

were three teams that gave the same amount every time, two of them were entirely domestic teams, the other was entirely international. There was another international team that rewarded evenly 5 out of 7 times. In 2005 there were three international teams that gave the same amount 9, 8 and 7 times out of 10. The most an all domestic team rewarded evenly was 3 times in 2005.

The most interesting revelation from all years is that mixed domicile teams (teams with both international and domestic students) nearly never (once only in 2005 and 2006) gave equal amounts to everyone.

Table 3 shows that teams of 4 (or less) are more likely to distribute amounts equally than teams of 5. In both 2004 and 2005 only one team of 4 was entirely international students and none were in 2006. In 2004, teams of 4 gave equal amounts 64% of the time – four times more often than the teams of 5. In 2004 there was one international team of 4 and one domestic team of 4 that gave equal distribution every time. In 2005, there was a significant reduction in the number of times teams of 4 gave equal amounts, but there was one international team of 4 that gave equal amounts 8 times out of 10. In first semester 2006 teams of 4 never gave equal distribution.

Since the teams stay the same and continue to work on the same project it is interesting to consider the four factors that influence the students to not distribute amounts equally.

Firstly, in 2004 a change was made to the amount the students had to distribute between semesters – in first semester they had only 20 marks to distribute but in second semester they had a virtual $100. The reasons for this change are discussed in Clark et al (2005). This allowed students to be more discerning, a large number of students only differentiate by a $1 between team members. In 2004, 25 of the equal distributions by teams were done in first semester, and 17 in second semester when 4 of the 7 tests were conducted.

Secondly, students have been advised in course materials and in lectures not to make a pact, the following has been in the Project Manual (the unit bible) since 2004:

"It is tempting to have a pact with your team members to always give high ratings. You are advised ***not*** to do this. This encourages individuals not to do their share of the work and you will end up carrying them or submitting sub-quality work. You should respond based on ***your*** opinion of each person's contribution. You should find that if you are honest with each other you will all learn more and improve, as students are often in a better position to provide one another with meaningful feedback regarding both technical and interpersonal performance." (Herbert, Ollington 2006, pg 9).

When an entire team distributed amounts equally they were quizzed at a follow-up meeting with the lecturer as to whether this was really a true reflection of the work completed. In 2005 teams were advised at the follow-up meeting that this indicated that the team felt they all deserved the same mark for that work product but that they would all get the same lowest mark determined by the lecturer based on the Individual Contribution Reports

and timesheets. The impact of this is indicated by the number of teams that only reward evenly once. In 2006 from the very first follow-up meeting with the lecturer teams were given this warning when only a single member of the team distributed amounts equally. This appears to have had a dramatic effect in 2006. In 2004 there were five teams that rewarded evenly once in first semester only, and four other teams rewarded evenly 4 or less times with only one of these teams doing it again once in second semester. In 2005 there were seven teams that all rewarded evenly 4 or less times, with five teams doing it once in first semester only. In 2006 there were four different teams that all rewarded evenly once – two on the first pay packet and one each on the second and third pay packet.

Thirdly, students receive two formative marks during first semester that indicate the influence of the pay packets. The impact of this is also indicated by the number of teams that only reward evenly once. Also at the end of first semester students receive a grade and are thus made aware (in some cases painfully) of the impact of carrying other team members on their own overall results. The impact of this is indicated by the number of teams that reward evenly consistently in first semester but rarely in second semester (4 teams in 2004, and 2 in 2005).

Finally, in first semester students have to allocate work equally during all phases of the lifecycle (particularly design and implementation), but in second semester they are advised to allocate work based on students particular strengths while ensuring everyone does the same overall amount of work. If teams do this you would expect the pay packets to be less equal in second semester.

3 Are students honest about themselves?

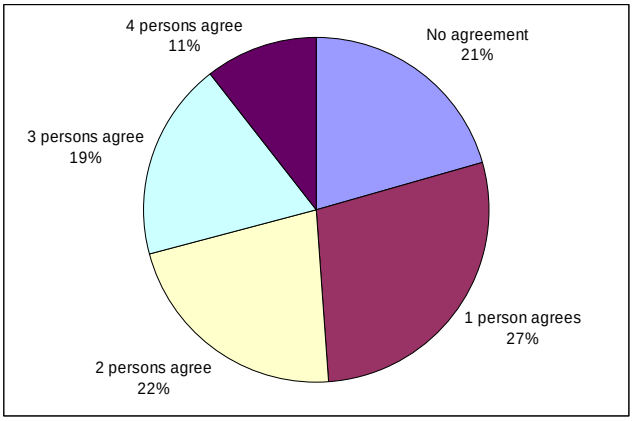
A major concern with having students quantify contributions is that they will exaggerate their own contribution. In 2004 students gave themselves the highest amount 191 times. In 134 cases they gave the same amount to at least one other member of their team, but not the entire team. From table 4 you can see that in both 2004 and 2005 students gave themselves the highest amount around 21% of the time and in 2006 this has increased to 26%. International students gave themselves the highest amount less often than the other categories, but the earlier section demonstrated that they are much more likely to give equal amounts to everyone.

Do students give themselves the highest amount when they shouldn't? In 2004 of the 191 cases where a student gave themselves the highest amount other team members gave them their highest amount 142 times, 74%. In 2005 203 of the 257 cases had agreement, 79%. In 2006, 77 of the 91 cases had agreement, 84%. Graph 1 shows that 79% of the time team members agreed that a member should have had the highest amount. It also shows that over 52% of cases had the agreement of at least two other team members (meaning more than half the team agreed on who should get the highest amount). Graph 2 shows how often the people who agreed gave the student the same, more or less money than they gave themselves. 21% of the time they weren't willing to give them as much money as students gave themselves. This analysis indicates that giving yourself the highest amount is justified in the majority of cases, but there are some students giving themselves the highest amount when the rest of the team does not agree.

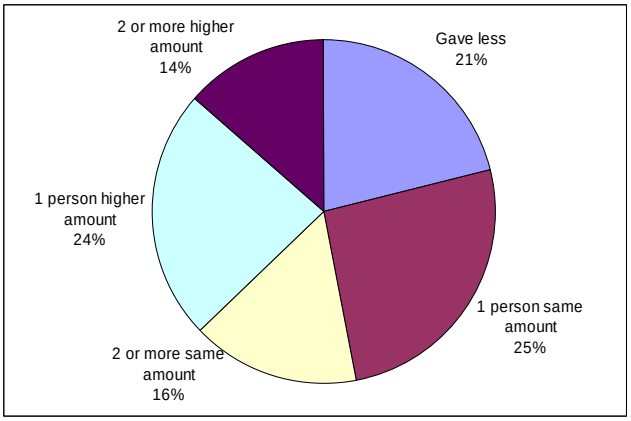
Do students give themselves the highest amount when they should? Over the three years, when at least two members agreed that another member should be given the highest amount in only 59% of these cases did the member agree. When at least three people agreed that another member should get the highest amount the member agreed only 70% of the time. So it seems some individuals do not give themselves the highest amount when they should.

		Possible	*Highest to themselves*	*Others equal high*	*% of possible*	*Lowest to themselves*	*Others equal low*	*% of possible*
2004	Females	91	2	23	27	9	7	18
	Males	812	55	111	20	29	39	8
	Domestics	749	51	117	22	29	42	9
	Internationals	154	6	17	15	9	4	8
	Total	903	57	134	21	38	46	9
2005	Females	170	11	20	18	4	28	19
	Males	1010	92	134	22	65	123	19
	Domestics	780	74	127	26	61	93	18
	Internationals	400	29	27	14	8	58	17
	Total	1180	103	154	22	69	151	19
2006	Females	72	0	15	21	10	15	35
	Males	273	33	43	28	23	33	21
	Domestics	207	29	36	31	24	28	25
	Internationals	138	4	22	19	9	20	21
	Total	345	33	58	26	33	48	23

Table 4: Individuals that gave themselves their highest/lowest amount (but NOT equal amount to all)



Graph 1: How many agreed a person should be given the highest amount



Graph 2: How often the agreer gave equal, more or less money than student gave themselves

As shown on the right side of table 4, in 2004 students gave themselves the lowest amount 84 times. In 46 cases they shared an equally low amount with at least one other member of their team, but not the entire team. From 2004 to 2005 there was a significant increase in the number of students who gave themselves the lowest amount. This change possibly occurred as a result in the change of amounts being distributed in 2004 (20 marks in first semester versus $100 in second semester) which meant in 2004 many more students distributed equal amounts. Strangely there was no change in the number of students who gave themselves the highest amount.

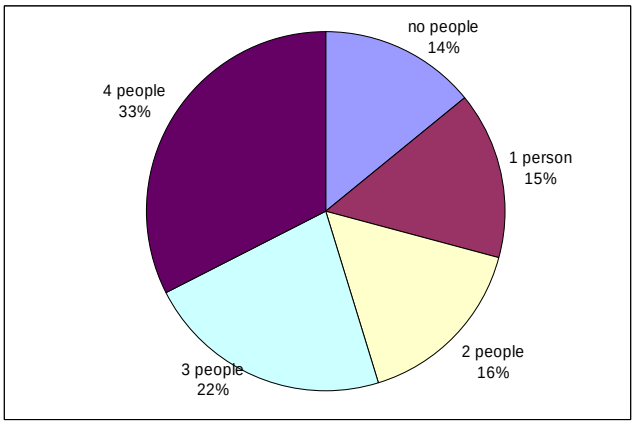
In 2004 and 2006 female students were more likely to give themselves the lowest amount, but this was not indicated in 2005. In 2005 there were two gender mixed teams where the majority of students were female (only one male). There has only been one other team where females have out numbered males and the ratio was 3:2.

Do students give themselves the lowest amount too often? In 2004 of the 84 cases where a student gave themselves the lowest amount other team members gave them their lowest amount 66 times, 78%. In 2005 175 of the 220 cases had agreement, 80%. In 2006, 60 of the 81 cases had agreement, 77%. This indicates that giving yourself the lowest amount is justified in the majority of cases, but there are some students being too hard on themselves.

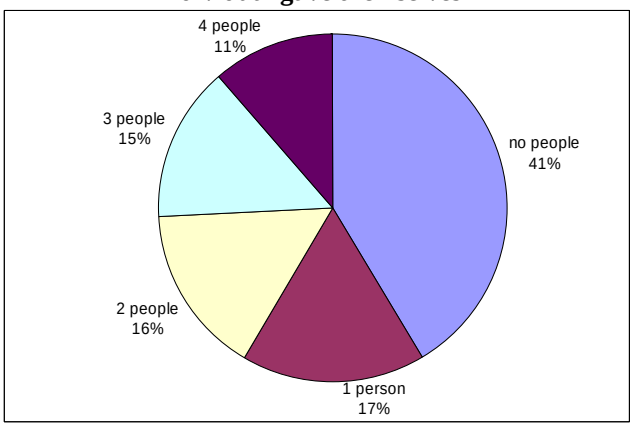
Do students give themselves the lowest amount often enough? Over the three years, when at least two members agreed that another member should be given the lowest amount in only 44% of these cases did the member agree. When at least three people agreed that another member should get the lowest amount the member agreed only 50% of the time. It seems many individuals do not give themselves the lowest amount when they should.

It is worth noting that over 71% of the time at least 2 people gave the same or more than the person gave themselves as shown in Graph 3, which is at least half the rest of the team. Again it is encouraging that 58% of the time at most 1 person in a team thinks a person should get less than they gave themselves, as shown in graph 4. These two graphs are indicating that in the majority of cases an individual is giving themselves an amount that the majority of the team agrees with.

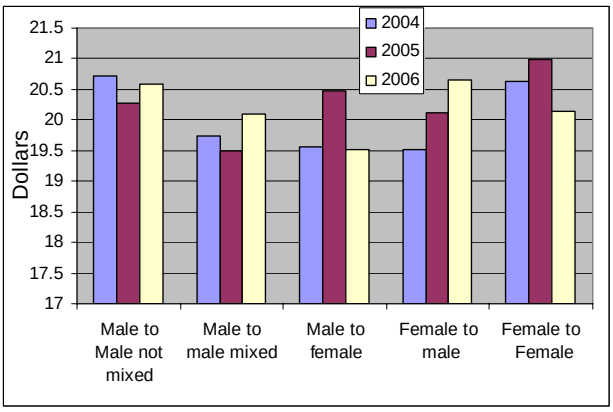
This evidence suggests that student quantification of contribution should not be used in isolation and that it is necessary to have other forms of evaluating the contribution of individuals to confirm the quantitative opinions of the students. Two other methods utilized in Software Engineering Project are timesheets filled in by an individual and individual contribution reports written by a team member about their contribution to a work product and agreed with by other team members, these are further described in Clark et al (2005).



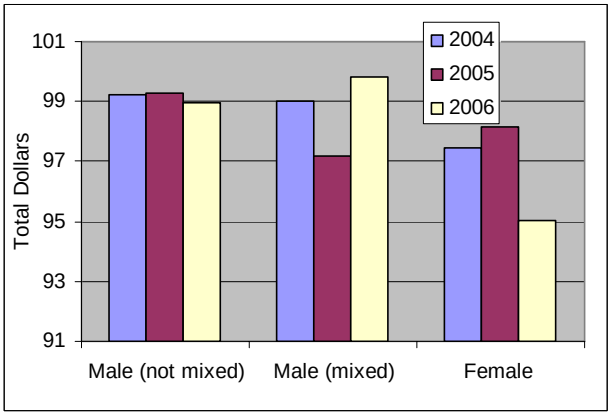
Graph 3: How many people gave same or more than individual gave themselves



Graph 4: How many people gave less than individual gave themselves.



Graph 5: Average individual amounts



Graph 6: Average total amounts

4 Are female students treated fairly?

Graph 5 shows the average amounts given to team members based on gender. Graph 5 indicates that males gave fellow males in a mixed gender team lower amounts than males do to fellow males that are in all male teams. Males in 2005 gave higher amounts to females than males, in 2004 and 2006 they gave slightly lower amounts to females. Male students gave female students lower amounts than female students gave to fellow female students. In 2004 and 2005 females gave higher amounts to females than males. In 2006 females gave higher amounts to males than they gave to females. In 2005 and 2006 females gave higher amounts to male students than male students gave each other. So in conclusion, the individual amounts are not showing any consistent gender bias.

Table 5 is showing the significance of differences in average individual allocations by genders for 2005 and 2006. All levels of significance are determined using a Mann-Whitney nonparametric test for significance between the distributions of two independent samples, with statistical significance defined by $p<0.05$. Table 5 is showing that all the differences in 2005 were not significant but that the difference between what males gave to females and what males gave males was significant in the first semester of 2006.

Graph 6 shows the average total amounts (the accumulation of all the WPPP amounts) and indicates that in 2004 and 2006 female students averaged less than their male students, but in 2005 females were averaging slightly higher than males in mixed teams.

Table 6 is showing the number of times the maximum and minimum total amounts were given to different groupings of students in the mixed teams. In 2005 and 2006 female students were given the minimum total amount more often than their male team members. In 2004 and 2006 female students were given the maximum total amount less often than their male team members.

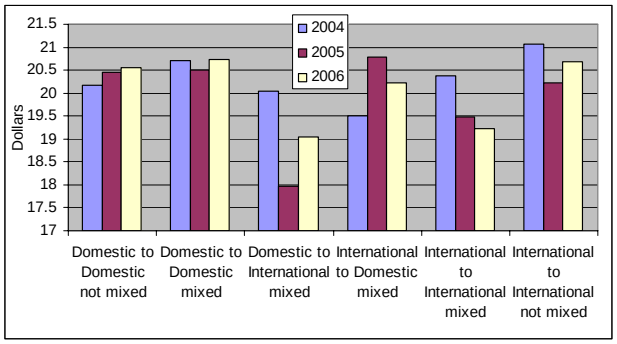
Since the distribution patterns have not been consistent over the three years it is not possible to conclude that there is a gender bias.

		2005			Semester 1, 2006		
		N	Average amounts	p	N	Average amounts	p
1	Male to Female	377	20.48	0.1515	219	19.53	**<.0001**
	Male to Male (mix)	932	19.5		534	20.10	
2	Male to Female	377	20.48	0.0735	219	19.53	0.0526
	Female to Female	160	20.98		126	20.15	
3	Female to Male	375	20.13	0.4483	219	20.66	0.0436
	Female to Female	160	20.98		126	20.15	
4	Female to Male	375	20.13	0.2483	219	20.66	0.4207
	Male to Male (mix)	932	19.5		534	20.10	

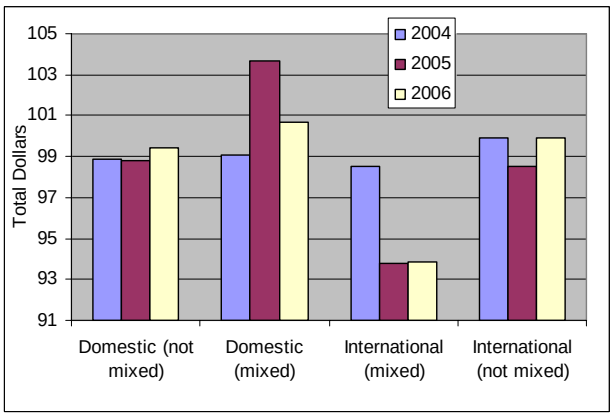
Table 5: Significance of differences in average individual allocations for 2005 and 2006

		N	Possible	Minimum	% of poss	Maximum	% of poss
2004	Females	13	91	33	36	30	33
	Males in mixed	41	287	117	41	123	43
2005	Females	17	170	89	52	84	49
	Males in mixed	30	300	138	46	125	42
2006	Females	24	72	30	42	19	26
	Males in mixed	52	156	38	24	54	35

Table 6: Minimum and Maximum total amount distribution



Graph 7: Average individual amounts



Graph 8: Average total amounts

5 Are international students unfairly discriminated against?

Graph 7 shows the average amounts given to team members based on domicile. Table 7 is showing the significance of differences in average individual allocations by domicile for 2005 and 2006. Graph 7 indicates that domestic students in both mixed and non-mixed teams average around the same amount for each other in all years, with domestics in mixed teams being fractionally higher. Domestic students gave international students lower amounts than they gave fellow domestic students and this difference was significant (Table 7, row 1). Domestic students gave international students lower amounts than international students gave each other in all years and the difference was significant (Table 7, row 3). International students gave domestic students significantly more than domestic students gave domestics students in 2005 (Table 7, row 2), interestingly international students gave domestic students less than domestics students gave each other in 2004 and 2006! Internationals students gave lower amounts to fellow international students rather than fellow domestic students in 2005 and 2006, the reverse happened in 2004, the difference in 2005 and 2006 was significant (Table 7, row 4). Internationals in mixed teams gave significantly lower amounts to each other than internationals in non-mixed teams (Table 7, row 5), though all-international teams distribute amounts equally often, whereas no (or nearly no) mixed teams did.

Graph 8 indicates that domestics in mixed teams regularly received average total amounts higher than domestics in non-mixed teams. Internationals in mixed teams consistently got lower amounts than internationals in non-mixed teams. Domestics in mixed teams got higher amounts than internationals in mixed teams in all years. There was no indication of a problem in 2004, but in 2004 the amount of equal distribution was the highest. 2005 data indicated a potential problem that is also being seen so far in 2006. In 2005 and 2006 the number of international students increased significantly on 2004.

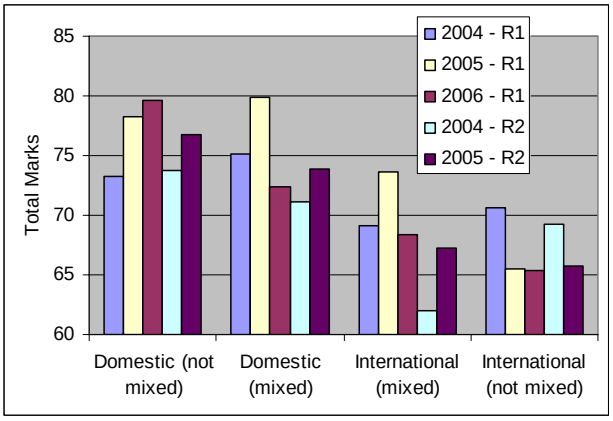
Table 8 indicates that in all years internationals in mixed teams were twice as likely to be given the minimum total amount. Likewise in all years the international students were less likely to be given the maximum amount.

		2005			Semester 1, 2006		
		N	Average amounts	p	N	Average amounts	p
1	Domestic to Domestic (mix) Domestic to International	220 290	20.51 17.98	**<.0001**	186 132	20.74 19.05	**<.0001**
2	International to Domestic Domestic to Domestic (mix)	290 220	20.78 20.51	**0.0143**	132 186	20.22 20.74	0.2676
3	International to International (mix) Domestic to International	320 290	19.47 17.98	**0.002**	246 132	19.21 19.05	**0.0375**
4	International to International (mix) International to Domestic	320 290	19.47 20.78	**<0.0001**	246 132	19.21 20.22	**0.0001**
5	International to International (mix) International to International (not mixed)	320 920	19.47 20.22	**<0.0001**	246 300	19.21 20.69	**0.0031**

Table 7: Significance of differences in average individual allocations for 2005 and 2006

		N	Possible	Minimum	% of poss	Maximum	% of poss
2004	Domestics in mixed	11	77	13	17	29	38
	Internationals in mixed	8	56	19	34	13	23
2005	Domestics in mixed	13	130	25	19	39	30
	Internationals in mixed	16	160	65	41	35	22
2006	Domestics in mixed	22	66	12	18	27	41
	Internationals in mixed	26	78	28	36	11	14

Table 8: Minimum and Maximum total amount distribution



Graph 9: Average Final Marks

So all this analysis indicates that there maybe an issue of domicile bias, but maybe the differences in allocations are justified. An analysis of the final marks is necessary to determine if there is unfair discrimination.

The entire software development lifecycle is repeated each semester. In the first semester teams complete release one (or a third of the project), in second semester they complete release two (the remaining two-thirds). Graph 9 shows the average final marks for release 1 in all years and for release 2 in 2004 and 2005.

In 2004 and 2005 domestic students in mixed or non-mixed domicile teams were achieving around the same results in each semester. In both years in first semester domestics in mixed teams got slightly higher results than domestics in non-mixed teams but this was reversed in both years in second semester. In 2006 domestics in non-mixed teams have got slightly higher marks than domestics in mixed teams.

Overall graph 9 indicates that international students whether in mixed or non-mixed teams are averaging around the same grade but less than domestic students. Though in second semester 2004 international students in mixed teams got significantly lower results. This result is partially explained by the fact that there were only 8 international students in mixed teams and 4 of them were in the one team and that team did not do well. In 2005 international students in mixed teams achieved significantly higher results in first semester to international students in non-mixed teams (8%), but in second semester the difference was only 1.5%, and table 9 shows this difference was not significant.

In all years international students in mixed teams consistently achieved lower average final marks that domestic students in mixed teams. Table 9 shows that the difference was significant in second semester 2005.

The final marks are calculated using the WPPPs and so they are not completely independent. The teams have to produce three design reports, two in first semester and one in second each worth 10% for the team. The teams developed a release of the software in each semester. The team marks for these work products are not influenced by peer assessment.

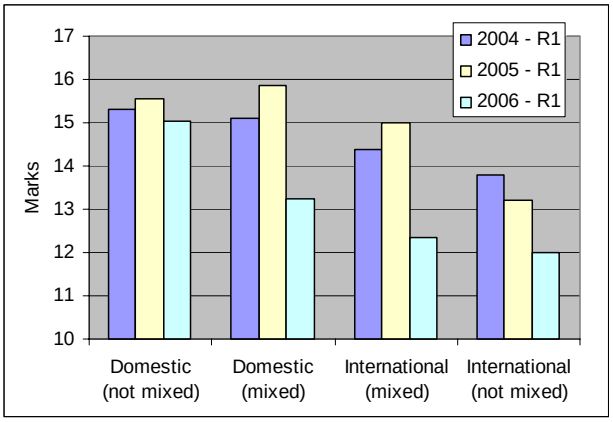
There has been no significant differences in marks for design reports for domestic student whether in a mixed or non-mixed team, in the majority of cases the differences have been less than 2% except for design report 2 in 2006 when the difference was 10%. An analysis does indicate though that internationals in mixed teams do achieve higher results than internationals in non-mixed teams for the design reports (over 20% for design report 1 in 2005 and 2006), and table 9 confirms that the difference was significant for design report 3 in 2005 where the difference was only 11%.

Graph 10 shows the team software results for release 1. Domestic students averaged around the same in both mixed and non-mixed teams though the difference in 2006 is larger than the other two years. International students in mixed teams averaged slightly higher than those in non-mixed teams. The difference between international students in mixed and non-mixed teams was significant in second semester 2005, as shown in table 9.

So the analysis indicates that there is no significant difference in the results of domestic students whether in a

		N	*Average Mark*	*p*
Domestic (mix) International (mix)	Semester 2 individual mark	13 16	73.85 67.31	**0.0314**
	Design report team mark (max 10)	13 16	9 8.13	0.102
	Software team mark (max 20)	13 16	15.77 15.13	0.0869
International (not mixed) International (mix)	Semester 2 individual mark	24 16	65.75 67.31	0.496
	Design report team mark	24 16	7 8.13	**0.0244**
	Software team mark	24 16	14.42 15.13	**0.0322**
Domestic (not mixed) Domestic (mix)	Semester 2 individual mark	65 13	76.75 73.85	0.1131
	Design report team mark	65 13	9.2 9	0.4325
	Software team mark	65 13	15.42 15.77	0.1814

Table 9: Significance of differences in 2005 semester 2 marks



Graph 10: Team software results for release 1

mixed or non-mixed team. This indicates that the results of the domestic students in mixed teams are not suffering and they gain many generic skills from the experience.

The analysis of the amounts given in WPPPs did indicate a significant difference in both average individual amounts and total average amounts given to international students in mixed teams compared to international students in non-mixed teams and domestic students in mixed teams, indicating that there is a possible bias against international students in mixed teams. But the fact that there is a significant difference in the marks given to internationals in mixed teams versus internationals in non-mixed teams for team items such as design reports and software, and no significant difference in the overall marks indicates the allocations are reflecting actual performance. It is important to remember that both domestic and international students in mixed teams were giving higher amounts to the domestic students.

6 Conclusion

This paper investigated distribution pattern concerns about quantitative peer assessment. Equal distribution is a concern, if it is done for the wrong reason. The analysis has shown that individual students will distribute amounts equally, particularly international students. In Software Engineering Project four factors were influential in reducing equal distribution such that it rarely happens on a team wide basis. Consideration needs to be given to the amount that is distributed, releasing grades that show the impact of equal distribution and explaining to students the interpretation of distributing equally. Lecturers are concerned that some students are not honest about their own contribution. The analysis has shown this to be true, students are likely to give themselves the highest amount both too often and not enough, likewise, students give themselves the lowest amount both too often and not enough. This evidence suggests that the tool should not be used in isolation and that it is necessary to have other forms of evaluating the contribution of individuals to confirm the quantitative opinions of the students. Gender bias is a real concern but there was no conclusive consistent evidence that there was a gender bias. There was concern that international students consistently received lower amounts in mixed teams (from both domestic students and fellow international students) but an analysis of the results data concluded that there was no domicile bias and that peer assessments were reflecting actual (or at least perceived) performance.

7 References

Brown, R. (1995): Autorating: Getting Individual Marks from Team Marks and Enhancing Teamwork. *Proceedings of Frontiers in Education Conference*.

Clark, N. (2005): Evaluating student teams developing unique industry projects. *Proceedings of the seventh Australasian Conference on Computing Education*. Newcastle, Australia

Clark, N., Davies, P., Skeers, R. (2005): Self and Peer Assessment in Software Engineering Project. *Proceedings of the seventh Australasian Conference on Computing Education*. Newcastle, Australia

Gates, A.Q., Delgado, N., Mondragon, O. (2000): A Structured Approach for Managing a Practical Software Engineering Course. *ASEE/IEEE Frontiers in Education Conference*, October, Kansas City. **1**:21-26

Hayes, J.H., Lethbridge, T.C., Port, D. (2003): Evaluating Individual Contribution Toward Group Software Engineering Projects. *Proceedings of International Conference on Software Engineering, Portland*, Oregon. 622-627

Herbert, N., Ollington, R. (2006), *Software Engineering Project: Project Manual version 9*, Accessed July 31, 2006, http://www.comp.utas.edu.au/units/kxa351//2006ProjectManualv9.pdf

Humphrey, W. (1997): *Introduction to the Personal Software Process*. Addison-Wesley.

Kaufmann, D.B., Felder, R.M., Fuller, H. (2000): Accounting for individual effort in cooperative learning teams. *Journal of Engineering Education*. **89** (2), 133-140.

Layton, R.A., Ohland, M.W. (2001): Peer Ratings Revisited: Focus on Teamwork, Not Ability. *Proceedings of American Society for Engineering Education Annual Conference*, Albuquerque.

Wilkins, D.E., Lawhead, P.B. (2000): Evaluating Individuals in Team Projects. *Proceedings of the 31st SIGCSE Technical Symposium on Computer Science Education*, Austin, Texas, 172-175