Peer-assessment in Group Projects: Is It Worth It?

Geoffrey J. Kennedy
School of Information Technologies
The University of Sydney
NSW, 2006
gkennedy@it.usyd.edu.au9

Abstract

The incorporation of group projects into university computing courses is wide spread. However, there is evidence that students undertaking such projects express concern about the way in which marks awarded for outcomes produced by the group collaboratively are allocated to individuals in of the group. For example, since some group members contribute more or less than others, students feel that awarding the same mark to all members of a group is unfair. To address this problem some tertiary educators employ a strategy called 'peer assessment' as a way of determining how group marks are to be distributed among individuals.

Peer assessment in this context refers to any of a variety of approaches where students are required to assess other members of a group on their relative contribution to a project. This paper describes one method of using peer assessment for this purpose and presents a detailed analysis of data collected from a class of 90 students undertaking a system development project at an Australian university. The paper raises the question whether procedures like the one described in the case study are valid and explores the impact of such procedures on the purported objectives of the course.

Finally, it addresses the question "Is it worth it?" Does the effort expended on administering peer assessment result in significantly better assessment?[1]

1 Group Work in University Computing Courses

A recent survey of computing courses within a computer science department in an Australian university (White *et al.,* 2004) showed that the practice of having students carry out work in small groups, say from three to six, was quite common. Examples ranged from cases where the group task was a relatively minor one within a more conventionally taught course to those where the course consisted solely of students undertaking a single group 'project'.

Looking at recent literature on the subject it is apparent that this university is far from unique in this respect (see Boud, 1999, Cheng and Warren, 2000 or McInnis and Devlin, 2002). Considering the extent to which project teams are utilised in the computing industry as well as other areas of science, it is no surprise that universities should be keen to help students develop the team skills needed that mode of operation.

In this paper we will first look briefly at the options for assessing collaborative work in tertiary courses and then examine more closely the use of peer assessment as a mechanism for arriving at grades for individual group members. Finally some conclusions will be drawn regarding the application of peer assessment to the allocation of group marks.

2 Assessing Group Work

'Assessment' in this paper refers to any process where a 'mark', supposedly being some reflection of quality, is awarded for some piece of work undertaken by students. In cases where the work is carried out collaboratively by a group of students and a single mark is awarded for the work, this is referred to as 'group work' and the mark is called a 'group mark'.

There are two aspects of group work that might be assessed, (i) the *product*, or (ii) the *process*.

The *product* includes any outputs such as the results of an investigation or experiment, a research report, a working program or prototype system, an oral presentation involving a variety of media, a management plan or budget, or an eye-catching web-site.

The term *process* refers to any of a range of processes involved in the conduct of collaborative work. It includes such behavioural elements as the way in which the group goes about its task, the formulation of a project plan, the way in which a methodology is arrived at and the management of the project throughout its duration. It also takes in the management of group dynamics, the 'effectiveness' of group members in terms of their participation in group decisions, approach to allocating tasks, accomplishment of allocated tasks, interaction with group members, and other aspects of collaborative work.

In general, tertiary teachers are confident about being able to assess the products of group work (Race, 1995). Usually they determine a mark by comparing the product with established standards using some predetermined criteria.

The assessment of 'process' is rather more difficult, but the instructor may wish to assess aspects of *process,* for example in cases where the objectives of the course are expressed in such terms as "to experience", "to have the opportunity to" or "to work effectively in a group".

Either or both aspects may be considered important and therefore the subject of assessment, however in the author's experience, where group work forms part or all of a course in computing it is invariably some *product* that is assessed, and for the remainder of this paper that will be assumed to be the case.

3 Allocating Group Marks to Individuals

Let us assume that a mark has been awarded for some product of group work. Most often then it is necessary for each member of the group to receive an individual mark which counts towards their final grade for the course. There appear to be two obvious ways of arriving at this distribution of the group mark (see for example Gatfield, 1999, Reynolds and Trehan, 2000 or Utting, 1998):

(i) all members of the group receive the same mark, or

(ii) the mark is distributed according to contribution of each individual to the project.

The first approach has the merit of simplicity. However, it makes an assumption that all members of the group have contributed equally to the project, and this is often not the case. As a result, students sometimes complain that the practice is unfair (see McInnis and Devlin, 2002 and White *et al.*, 2004)

The second approach would be a simple matter if only we could be confident of determining the relative contributions of each group member. Many academics believe that the 'peer assessment' can provide a mechanism for doing just this. In the next section one method of using peer assessment for this purpose is examined.

4 A Typical Application of Peer Assessment

Peer assessment refers to any of a variety of approaches where group members are required to evaluate other members of the group on their relative contribution to a project. The group mark awarded for the collaborative product can then be distributed accordingly.

Various methods of peer assessment are described in the literature for example Cheng and Warren (200), Gatfield (1999) or Utting (1998). One typical approach, used for the past several years in an information systems course at an Australian university, was investigated by the author[1]. The approach involved the following procedure:

> Given *n* members in a group, each member is required to provide an assessment of the contribution of the other *n-1* members by allocating each a score out of 100, such that the marks awarded total *100* * (*n-1*).
>
> Should the score awarded for any individual exceed 110, or be less than 90, the student is required to provide a short justification. Where students fail to submit an allocation, it is assumed that all members are awarded a score of 100.
>
> The final apportionment of the group mark for the project is then made according to the average score obtained by each of the group members from their peers.

Table 1 shows the scores awarded by one student (Id = 18), Jason Bournes, assessing the contributions of each of his eight colleagues. The names have of course been changed to preserve anonymity. Each member of the group provides eight scores in a similar fashion.

Team 2: Peer assessment by Jason Bournes		
Id	**Name**	**Score**
11	Land, Arthur	115
12	Thomas, Matthew	100
13	Karas, Muhammad	100
14	Long, Nathan	100
15	Heng, Jin	95
16	Coutts, Boris	95
17	Hunt, Carmel	90
18	Bournes, Jason	-
19	Reid, Jane	105

Table 1: Peer assessments by student Id = 18

In the class studied there were 90 students arranged in 10 teams consisting of from six to ten members. The actual scores awarded by students were collected, checked for consistency and completeness, and stored in a database. Comments, where provided as requested, were transcribed into a document together with the assessors' names.

Table 2 (next page) shows a matrix of the peer assessments of the whole group of nine students. The row values represent the asessments by each group member of the contribution of his/her peers. The blank cells result from the fact that group members do not assess themselves.

Student Jin Heng (Id = 15) chose not to provide scores and so by default all members were awarded 100. Student 17, Carmel Hunt, gave equal scores to all members of the group explicitly.

The columns of the table give the scores awarded to each group member by other members of the group. The averages of each column are then used to apportion group marks. For example, Jason (Id = 18) received 97, 100, 100, 100, 100, 95, 100, -, and 96.5. The mean of these scores is 98.6 so he will receive 98.6% of any mark which is awarded to the group for a group output.

1. The author is indebted to Paul Bowen of the School of Business, University of Queensland for providing the source data.

Team 2											
StuId	11	12	13	14	15	16	17	18	19	Total	Sdev
11		97	97	97	97	97	97	97	121	800	8.5
12	110		100	100	90	100	90	100	110	800	7.6
13	110	100		100	100	90	80	100	120	800	12.0
14	112	100	100		95	95	90	100	108	800	7.2
15	100	100	100	100		100	100	100	100	800	0.0
16	120	95	95	100	100		95	95	100	800	8.5
17	100	100	100	100	100	100		100	100	800	0.0
18	115	100	100	100	95	95	90		105	800	7.6
19	125.6	96.5	96.5	96.5	91.5	96.5	100.5	96.5		800	10.6
Total	892.6	788.5	788.5	793.5	768.5	773.5	742.5	788.5	864	7200	
Avg	111.6	98.6	98.6	99.2	96.1	96.7	92.8	98.6	108.0		
Sdev	8.9	2.1	2.1	1.5	3.9	3.5	6.8	2.1	8.6		

Table 2: Matrix of peer assessments for Team 2

Table 3 illustrates how individual marks can be calculated using the average scores shown in Table 2. It assumes that a group mark of 18/20 has been awarded. It should be noted that, using this 'constant sum game' approach described here, it is possible for a student to score over 100% if both their contribution and the group mark are very high. This effect can be seen in Table 3 where student 11 scores 20.1 out of a possible 20.

Student Id	**Column totals**	**Column averages**	**Group Mark**	**Individual Marks**
11	892.6	111.6	18	20.1
12	788.5	98.6	18	17.7
13	788.5	98.6	18	17.7
14	793.5	99.2	18	17.9
15	768.5	96.1	18	17.3
16	773.5	96.7	18	17.4
17	742.5	92.8	18	16.7
18	788.5	98.6	18	17.7
19	864.0	108.0	18	19.4
Totals	7200.0	900.0	162.0	162.0

Table 3: Calculation of individual marks

5 Analysis of Peer Assessment Data

The purpose of this analysis is to understand better what the peer assessment scores are actually measuring and to assess their validity for the purpose of distributing group marks.

For this 'constant sum game' the total of each of the rows of Table 2 is 800, that is *100 ** (*n-1*). Standard deviations of the rows can be thought of the variance in the contribution of group members as perceived by one particular group member. If all group members were consciously attempting to measure the same quantity then we should expect the variance in each row to be about the same. Clearly this is not the case. As mentioned, student 17 felt that there was no variance, and student 15 indicated the same by not submitting any assessment. Students 13 and 19 perceived considerable differences in contribution. This raises the question "*Were all the students trying to assess the same quantity?*"

The relatively high standard deviations of rows 18 and 19, suggest that these two students tried to discriminate between the contributions of different group members. However, they differ in opinion as to how much more student 11 contributed than the others, and they disagree entirely about the contribution of student 17, and to a lesser extent 15.

The column averages shown in Table 2 are taken in this method of using peer assessment to be a measure of the contribution of each group member.

If all members of the group were trying to estimate the same quantity, we should expect standard deviations to be about the same for each column, reflecting simply the variability in estimates of the same quantity by different assessors. In fact the column standard deviations vary widely, suggesting instead that each assessor has rather a different perception of the 'contribution' of his/her peers.

Table 4 (next page) shows the peer assessment matrix for Team 6. The anomalies noted in the case of Team 2 are even more exaggerated. The row standard deviations range from 0 to 53.8, showing considerable disagreement about whether group members contributed equally or not. The column standard deviations also vary widely, suggesting that group members had differed significantly in their estimates of the contribution specific peers.

Student 51 was apparently a poor contributor, receiving an average score of 47.3. But individual assessments of this student by group members differ widely, with scores ranging from 9 to 100. The column standard deviation was 40.9. Students 52 and 53, achieved very high 'contribution factors' averaging 137%, but individual assessments by the group were again extremely variable, ranging from 100 to 190.

Peer assessments for the other nine groups show similar trends with high variability. Table 5 shows the distribution of column standard deviations for all 90 students. It can be seen that assessments of team member contributions typi-

Team 6												
StuId	46	47	48	49	50	51	52	53	54	55	Total	Sdev
46		100	100	90	60	10	170	170	100	100	900	49.5
47	110		100	105	70	50	130	130	105	100	900	25.9
48	100	100		100	100	100	100	100	100	100	900	0.0
49	93.8	99.8	97.8		94.8	94.8	107.8	109.8	99.8	101.8	900	5.7
50	100	100	100	100		100	100	100	100	100	900	0.0
51	100	100	100	100	100		100	100	100	100	900	0.0
52	119.3	113.9	92.2	108.4	48.8	10.8		184.3	113.9	108.4	900	48.3
53	118.6	118.6	75.4	107.8	53.9	10.8	183.2		118.6	113.2	900	48.6
54	100	80	60	85	70	40	190	190		85	900	53.8
55	140	81	108	90	63	9	153	148	108		900	46.0
Total	981.6	893.2	833.4	886.2	660.5	425.4	1234	1232	945.2	908.4	9000	
Avg	109.1	99.3	92.6	98.5	73.4	47.3	137.1	136.9	105.0	100.9		
Sdev	14.6	12.7	15.2	8.4	19.9	40.9	37.6	37.3	7.1	7.6		

Table 4: Matrix of peer assessments for Team 6

cally have standard deviations of from 10 to 20%. This high variability among the assessments of different members of a group suggests either significant differences in their ability as assessors, or that they are not at all clear about what they are supposed to be assessing.

Figure 1 (next page) shows the distribution of scores recorded for the whole class, compared with a normal distribution with the same mean (100) and standard deviation (21.2). The scores can be seen to be tightly grouped around the mean. Whilst individual peer assessments ranged from 10 to 190 the distribution is actually very narrow. Either most students are perceived to have contributed equally, or most student assessors are reluctant to discriminate among their peers.

Standard Deviation	Frequency
0 - 4	1
5 - 9	15
10 - 14	32
15 - 19	17
20 - 24	14
25 - 29	3
30 - 34	5
35+	3

Table 5: Column standard deviations

The final 'contribution factors' used to determine the allocation of group marks to individuals, are the averages of the scores awarded to each individual by other members of the group. As a result, attempts at discrimination by student assessors tend to be smoothed out. In Figure 1 the distribution of these average scores is shown as the bottom dotted line: 40% of values lie between 95 and 105, 60% lie between 90 and 110 and 72% lie within one standard deviation (13.7) of the mean (100).

The analysis of peer assessments given above prompts two crucial questions: "*What quantity were students trying to assess*?" and "*What possible means could they have used to estimate it?* These questions need to be answered if peer assessment as described here is to be a valid procedure.

6 The Response of Students to a Peer Assessment Requirement

Students react in different ways to the requirement to provide assessments of their peers. In this section comments of student assessors in the case study are examined to gain insight into the attitudes of students to the use of peer assessment and to find evidence of any effect that its use has had on the group work and group processes themselves.

Students reluctant to judge peers

Some students are reluctant to mark their colleagues down, even if they appear to have contributed less than others. One student who gave equal marks to all other group members commented: "*Although some members of the group did put in more work that others, I don't feel they deserve more marks, especially at the expense of others".* These egalitarian feelings were not reciprocated by her peers and she was awarded scores ranging from 49 to 100.

White *et al.* (2004) also reported some students who preferred not to judge their colleagues: *"Others in the group felt bad about allocating marks (lower) to members who contributed much less".* One student noted a more pragmatic motive for students' reluctance to award low scores: *"Some students give everyone favourable marks because they hope to receive favourable marks from those students".*

Students keen to discriminate

Other students in the case study class were quite prepared to mark down those who had not done a fair share. Margaret criticised the contribution of one of her colleagues and awarded a score of 40: *"Thomas did nothing at the meetings and indicated that he just wanted to 'pass'*". Andrew

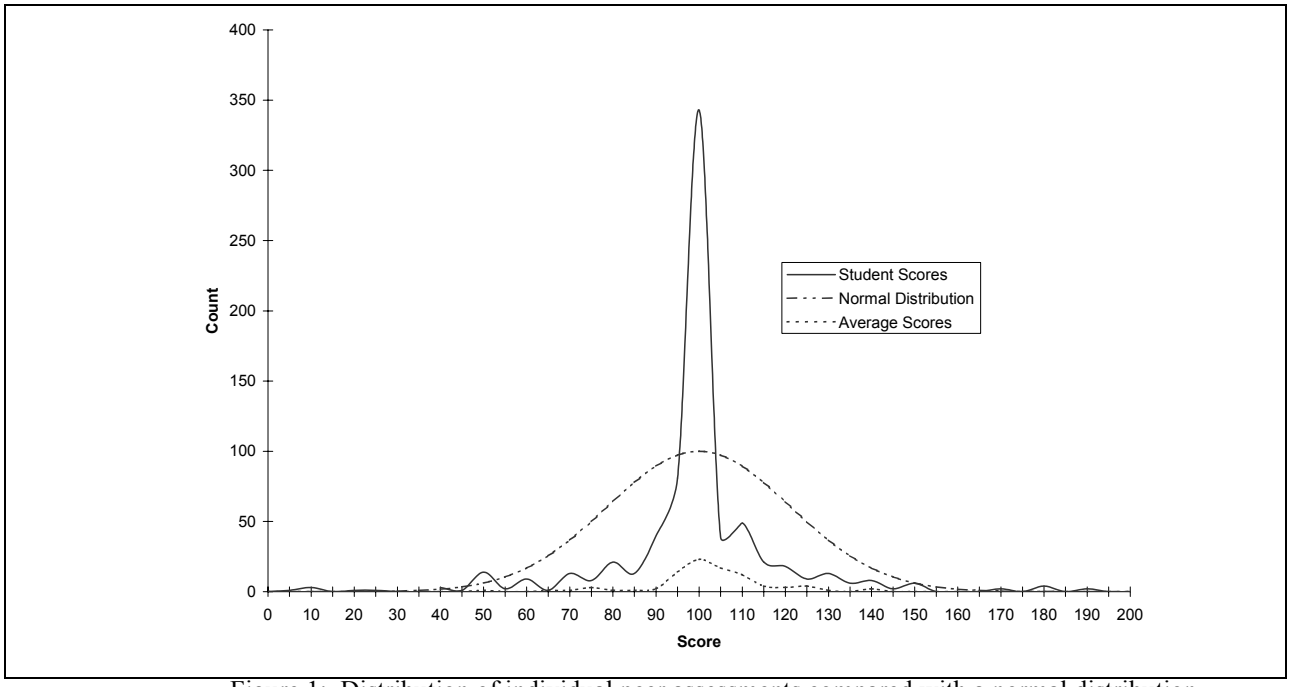

Figure 1: Distribution of individual peer assessments compared with a normal distribution

also noted Thomas's failure to contribute and awarded him a score of only 10%.

Keith was a very capable, high achiever. He provided quite detailed comments about his colleagues and awarded a wide range of scores (shown in parenthesis):

> Geoffrey (130) "*Committed and indispensable member of the team*".
>
> Robin (75) "*I had to redo quite a few data flow diagrams for him. It is unpleasant because I should be doing the forms and sql triggers*".
>
> Sharon (75) "*Too much work commitments and thus can't fulfil her promises*".
>
> Terry (50) "*The group functions well with or without him. [He] contributed a total of 7 hours? ... I would be happier to redistribute his 50 points to the rest of the group*". This opinion was shared by others.

Terry was clearly a poor contributor. None the less, he agreed that Keith (140) and Geoffrey (145) had done a good job, and that Sharon (75) and Robin (75) had not. Despite the stated requirement, he provided no comment to justify such high and low scores he awarded.

David also perceived great differences in the contributions of his group. He awarded marks ranging from 10 to 170 (though he got the total wrong so his marks had then to be scaled by 0.9 to give a total of 800).

Dysfunctional effects of peer assessment

There was evidence of considerable tension within groups resulting from the requirement that students assess one another. One phenomenon noted was the tendency for particular students to dominate the group and manipulate tasks to their own advantage.

One student commented on this dysfunctional effect of dominant students: "*These same people were domineering, and 'took over' the project. Ideas of other members of the group were disregarded. This was a group project and some members seemed to forget this. Surely this is not effective team work?*"

Margaret, felt the same way when she commented: "*Keith and Andrew (190) ... wouldn't really let anyone else help them ...*".

Kristin also noted the detrimental effect that this could have: "*Everyone did similar amounts of work except Gordon who did significantly more than anyone else ... Cheryl was a little less involved towards the end of the assignment... Cheryl* [changed her lecture times, so she] *didn't really know what was happening, ... tasks were not delegated to her*".

Cheryl saw things in a very different light. She felt under utilised and found it impossible to contribute fully. She commented, "*... Gordon did most of the programming... Although I appreciate the amount of effort he put in, I strongly view it as being unfair to the rest of the members. I was frustrated not really given a chance to work on the system myself".*

Weaker group members, especially those lacking programming skills, can be marginalised by being allocated tedious and unglamorous jobs, such as system testing or writing documentation, while the dominant students, often assuming the role of 'team leader', take on the high profile tasks like programming, which they enjoy doing and are good at. A probable victim of this sort of discrimination was Zeta, of whom Cheryl commented: *"As for Zeta, ... she not only failed to contribute in the Sample Data but she didn't really do a good job of testing the system*".

It becomes apparent that by judiciously manipulating the allocation of tasks, dominant students can significantly enhance their own grades at the expense of their less aggressive colleagues. If peer assessment can have the effect of causing group dysfunction and reducing coopera-

tion then it the question must be asked, should it be employed as an assessment strategy in course for which the development of team skills is seen as a prime outcome?

Student perceptions of fairness

In the study by White *et al*, (2004) referred to above, considerable pains were taken to explain the criteria to be used in the peer assessment. Even so, some students doubted that the process was consistent or fair: *"While one or two members of the group would mark objectively according to the criteria, some people prefer to award marks to friends as a goodwill even though they placed little support. The process was then somewhat unfair"*.

Another comment reflected similar concern: *"Even though one or couple of students didn't contribute as much in assessments, they were still given the full marks for consideration of that student. I don't think that's fair unless the student had a problem"*.

One other student expressed doubts about effectiveness of peer assessment: *"The process is still not very effective – realistically, everyone always just awards full marks because it is easier and less hassle."*

7 Problems with Peer Assessment

In this paper one common method of using peer assessment for the allocation of marks in group work has been examined, but any such approach depends ultimately upon group members providing assessments of their peers' contributions to the collaborative work. All methods therefore suffer from the same set of problems. Some of these problems are now listed.

(a) Students' judgments of each other are wildly inconsistent, as evidenced by the large column variances observed in Table 4. Not only are students unskilled in evaluating the contribution of their peers, but in most cases have no way of knowing how much time or effort their colleagues might have expended. Their 'assessment' of the contribution of the other group members is more a reflection of personal characteristics than any kind of quantitative evaluation.

(b) Peer assessment can impact badly on weaker students limiting their opportunities to participate. They are often ignored or allocated only less important tasks 'where they can do little damage', which gives them little chance to contribute equally to the group effort. More appropriate team management would permit these group members to become more involved and to contribute to the extent of their ability.

(c) The requirement to provide peer assessment can result in significant personal conflict between group members and reduce true cooperation and teamwork. Many who advocate peer assessment see it as a mechanism to coerce students into participating, however, it would be more useful for groups to learn how to manage non-contributors in ways that could apply equally to the work place.

(d) The collection and collation of data for the purpose of determining contribution factors is time consuming. The data collected for the 90 students in the case study took many hours to collate, correct and record. If peer evaluation does little or nothing to improve the validity of assessment then it is difficult to justify.

(e) Peer assessment does little to improve discrimination between student grades. Evidence suggests that students feel under pressure to award 'equal' marks to their colleagues for friendship's sake or so that they will receive reciprocal consideration. In the case study 44% of all students were awarded scores of 100 and 65% were between 90 and 110. When these individual assessments are further averaged to yield 'contribution factors' any attempt at discrimination is even further reduced.

(f) The task of assessment is the responsibility of the instructor. Students ought not to be placed in a position where they can influence their own grades or those of their peers. Evidence suggests that dominant students are able to manipulate task allocation to enhance their grades at the expense of colleagues.

8 Conclusions

The evidence provided by the case study presented in this paper raises a number of issues relating to the use of peer assessment as a mechanism for distributing group marks to individual team members in group projects. Though the use of peer assessment in courses involving group projects has been claimed by many to help solve the problem of unequal contribution by group members, the evidence suggests that the approach is of dubious worth in this connection and introduces a whole new set of inequities and problems.

1. The evidence provided raises serious doubts as to the validity or reliability of student assessments of the contributions of their peers. To improve the validity of the approach, considerable time and effort needs to be spent instructing students in assessment techniques and they need to be provided with appropriate data on which to base their judgments.
2. The effort required to collect and process the large amounts of data involved in peer assessment adds considerably to the burden of course administration.
3. The evidence suggests that marks awarded to individuals based on peer assessment in most cases differ only marginally from equal allocation. This appears to be a result of the general reluctance of students to be judgmental of their peers and of the averaging process employed to determine contribution factors. In any case, academics themselves should retain the responsibility for assessment.
4. Contrary to claims that the use of peer assessment provides some inducement for students to contribute to the team effort, evidence provided here suggests that the requirement for students to assess their peers, in many cases results in counter-productive tensions

among group members. A more enlightened approach would be to encourage students to invoke more constructive strategies that may help to motivate uncooperative colleagues and gain their cooperation. Since the acquisition of 'people' skills is always claimed as a desirable outcome from courses involving group work, students should be encouraged to practice using such skills.

As to the question "Is it worth it?", the evidence presented here, concerning the use of peer assessment as a basis for allocating marks in courses involving group work, suggests that perhaps it is not. Indeed, it may be that employment of such a strategy for assessment actually detracts from the attainment of course objectives.

9 References

Boud, D., R. Cohen, et al. (1999): Peer Learning and Assessment, *Assessment and Evaluation in Higher Education* **24**(4): 413-426.

Cheng, W. and M. Warren (2000): Making a Difference: using peers to assess individual students' contributions to a group project, *Teaching in Higher Education* **5**(2): 243-255.

Gatfield, T. (1999): Examining Student Satisfaction with Group Projects and Peer Assessment", *Assessment & Evaluation in Higher Education*, **24** (4) (December): 365.

McInnis, J.R. and Devlin, M. (2002): Assessing Learning in Australian Universities: Assessing group work, http://www.cshe.unimelb.edu.au/assessinglearning/docs/Group.pdf Accessed 10 Nov 2004.

Race, P. (1995): The Art of Assessing, *New Academic*, Autumn 1995, 3-5 and Spring 1996, 3-6

Reynolds, M. and K. Trehan (2000): Assessment: a critical perspective, *Studies in Higher Education* **25**(3): 267 - 278.

Utting, I.A. (1998): "Negotiated Assessment Criteria and Peer Assessment in Software Engineering Group Project Work: A Case Study", Research Report, Computing Laboratory, University of Kent at Canterbury, UK

White, F., Lloyd, H., Kennedy, G. and Stewart, C. (2004): Effective Management and Assessment of Group Work, Final Report: Teaching Improvement Fund 2003, Faculty of Science, University of Sydney.