Assessment & Evaluation in Higher Education
Vol. 29, No. 6, December 2004

Carfax Publishing
Taylor & Francis Group

Assessing contributions to group assignments

Lucy Johnston* and Lynden Miles
University of Canterbury, New Zealand

We report the use of a combination of self- and peer-assessment in an undergraduate social psychology laboratory course. Students worked in small groups on a self-directed empirical project that they each wrote up independently as a laboratory report. Marks for the written assignment were moderated by a contribution index measure based on the self- and peer-assessment measures. Our analyses indicated that: (i) students took the peer-assessment process seriously, clearly differentiating between group members on the contributions questionnaires; (ii) students show a self-bias, rating their own contribution to the group task higher than that of other group members; (iii) for a large majority of students the contribution index resulted in very little moderation of the final assignment marks; (iv) there was a strong correlation between the contribution index and the overall assignment score. Implications for the assessment of group work are considered.

Introduction

Group work is frequently included in tertiary education courses, across a wide variety of subject domains, including biology (Garvin *et al.*, 1995), business studies (Freeman, 1995), civil engineering (Rafiq & Fullerton, 1996) and computing and information systems (Lejk *et al.*, 1996). Group projects have a number of learning benefits (Jacques, 1984; Michaelsen, 1992; Freeman, 1995), including exposing students to other points of view from which they can learn and permitting the development of more comprehensive assignments than is possible for individual-based projects (Mello, 1993). Group projects also provide an opportunity for the development of inter-personal and teamwork skills, such as communication, leadership, planning and time management, skills that are highly sought by employers taking on graduates (Harvey & Green, 1994). In addition, students themselves respond positively to the opportunity that group-based projects provide for teamwork and for active learning (Garvin *et al.*, 1995).

Despite the potential benefits of group projects, they do raise problems for student evaluation and assessment. Converting the effectiveness of a given student's contribution to a group into a numeric grade is a complicated task. In encouraging teamwork it is essential that students feel confident that they will be rewarded fairly

*Corresponding author. Department of Psychology, University of Canterbury, Private Bag 4800, Christchurch, New Zealand. Email: l.johnston@psyc.canterbury.ac.nz

ISSN 0260-2938 (print)/ISSN 1469-297X (online)/04/060751-18

DOI: 10.1080/0260293042000227272

for their contributions and that any 'free-riders' will not benefit unduly from the efforts of others. The extent to which students should be rewarded for the process of the project versus the outcome of the project is, however, a contentious issue. Debate continues as to whether it is appropriate, in the tertiary sector, for a student to obtain higher marks than other group members simply as a result of greater than average effort to the group endeavour (Lejk *et al.*, 1996). The usual practice in higher education is for students to be graded solely on the quality of a submitted piece of work without consideration of the effort or input into that product. Indeed, in most cases the grader will be unaware of the amount of time and effort that a student has devoted to an assignment. For individual-based projects, however, both the input and output are determined by the individual. In this case, rewarding the student only for the output of their endeavours is, then, a reasonable means of assessment. For group assignments, however, the link between individual inputs and the output is not so clear; a 'free-rider', for example, might receive a high grade despite having very little input into an otherwise good group. Accordingly, in group assignments, the allocation of some marks to students based on their contribution to the group project rather than just on the final outcome of that project is not unusual. How to allocate such marks has, however, been a topic of considerable debate and analysis.

It is often difficult for a lecturer or tutor to adequately assess the contribution of individuals to a group project since much of the work occurs outside formal teaching sessions. Accordingly, grading of the contribution of individuals to a group assignment has increasingly been handed over to those with the most relevant information, i.e. the co-workers. Such peer-assessment has taken many forms (see Lejk *et al.*, 1996, for a survey of such methods), but the essential components are the same; group members evaluate each other and this evaluation is incorporated into the individual student's assessment, either as an addition to the marks awarded for the submitted assignment (Earl, 1986; Falchikov, 1988, 1991; Gibbs, 1992; Habeshaw *et al.*, 1993: Hindle, 1993) or as a moderator of those marks (Goldfinch & Raeside, 1990; Conway *et al.*, 1993; Goldfinch, 1994). In some instances student peer-assessment has included evaluation of the quality of the work produced as well as the contribution of students to that work. Although there is some evidence that students reliably reproduce staff marks (Orpen, 1992), some concern has been raised about the ability of students to assess the quality of work (Burke, 1969; Beard & Hartley, 1984; Rowntree, 1987), with student scores tending to cluster around the mean score with relatively little variance (Hughes & Large, 1993; Freeman, 1995). In our study the quality of the written assignment was assessed by the laboratory tutors and lecturers, with the peer-assessment related only to the group processes.

Peer-assessment has largely been applied to situations where the work group submits a single project (Lejk *et al.*, 1996). The project is awarded a mark, which is then moderated for each group member as a function of the peer-assessment. Group work does not always, however, culminate in the production of a single piece of work. In some cases each individual group member may submit a piece of work based on their group project and hence each student is assessed individually after

learning collaboratively (Hufford, 1991). This is the case for our laboratory assignment. Students worked together on a group project but each wrote a research report that was graded by either a laboratory tutor or lecturer prior to adjustment by peer evaluation factors. One aim of the present study was to investigate the use of peer-assessment in such a context. Individual assessment after group work may provide some protection for students against 'free-riders' but does not negate the utility of peer-assessment, since the basis for the individual reports is in the group work and so may be enhanced or impeded by high or low contributing group members, respectively.

A second aim of this project was to consider the question of whether or not self-assessment should be included in peer-assessment. Methods of peer-assessment differ according to whether they include self-assessment (see for example Goldfinch, 1994) or not (see for example Conway *et al.*, 1993). Some researchers have argued for the inclusion of self-assessment (for example Goldfinch, 1994) and others for its omission (Falchinov, 1991; Lejk & Wyvill, 2001). However, as Lejk and Wyvill (2001) noted, useful analysis of the self-assessment component of peer-assessment is infrequent. We did include self-evaluations, but our analysis also considered the consequences of omitting self-assessments on the grades obtained by the students. Goldfinch (1994) argued for the inclusion of self-assessment, noting that in some cases, where only peer but not self-assessment was allowed, one group member may be more generous in his or her evaluations than the rest of the group and accordingly penalize himself or herself. By including self-assessment, overly generous students would presumably inflate their own grade as much as they do that of their peers. However, where self- and peer-assessments have differed it appears as if students actually tended to assess themselves lower than did their peers (Krause & Popovich, 1996), especially the more able students (Lejk & Wyvill, 2001). Inclusion of self-assessment may then disadvantage more able students, which may be a reason not to include self-assessment (Lejk & Wyvill, 2001). Falchikov (1991) also argued against the inclusion of self-assessments, simply because she found a much lower agreement between self- and peer-ratings than between peer-ratings, although her very low sample size ($n = 4$) must be noted. Analysis of the impact of the inclusion and exclusion of self-assessment in the present study will add further to this literature.

The present paper presents details of the employment of a peer-assessment method in an advanced undergraduate social psychology laboratory class. Students worked in small groups on an empirical project which each group member then had to write up individually. After submitting their assignment each student was required to complete a contributions questionnaire for themselves and for each of the other members of their group, i.e. self- and peer-assessment. The contributions questionnaire was completed secretly, to facilitate greater discrimination by students in their peer-assessments (Lejk & Wyvill, 2001). A ratio of mean individual contribution ratings to mean rating for the work group was used to moderate the grade given to the individual's written assignment. Accordingly, the mark awarded to the written assignment could be either increased or decreased as a result of an individual's relative contribution to a group project.

Method

Participants

Sixty-one (48 female, 13 male) undergraduate students completed the group-based assignment and contributions questionnaires. Groups consisted of 3 ($n = 2$), 4 ($n = 10$) or 5 ($n = 3$) students. Five of the groups comprised female students only and 7 both male and female students. The groups were self-selected by the students. The project ran in the second semester of a full-year course and so it is likely that group selection was based on friendship and on previous interactions in the first semester laboratory sessions.

Nature of the assignment

The aim of the assignment was to give students experience in conducting social psychological research. Students worked in small groups designing and conducting a piece of social psychological research. Three topic areas, and key references, were introduced by the course lecturers and students were to select one of these areas in which to develop their own project, within the parameters provided by the lecturers concerned. Prior approval for these research projects had been obtained from the University of Canterbury Human Ethics Committee by the course coordinator. The topics available were: (i) investigating the relationship between empathy and prosocial behaviour; (ii) testing the Ideal Standards Model (Fletcher *et al.*, 1999; Simpson *et al.*, 2001) in mate selection; (iii) investigating the nature of impression formation from faces.

Groups designed their own study, collected data and analysed their results. Each student was then required to individually write up their group's study. Each group worked with a laboratory tutor (a graduate student in social psychology), meeting formally at 2 week intervals throughout the duration of the project (10 weeks). In the final week of the semester each group was required to give a brief (15 min) presentation to the class on their project. The entire assignment contributed 20% of the marks for the full year undergraduate paper in social psychology.

Contributions questionnaire

Students were informed, before they divided into work groups and began the project, that the marks assigned to their written assignment would be moderated by self- and peer-assessment of the contribution of each member to the group project. Full details of the moderation procedures were given to students both orally and in a written handout. After the reports have been handed in, each student rated the contribution of each member of their work group, including themselves, on a number of task functions and group management skills (Johnson & Johnson, 1975; Falchikov, 1988, 1991): generation of research ideas; contribution to background literature searches and reviews; contribution to research design and methodology; contribution to data collection; contribution to data analysis and interpretation; contribution to feedback presentation; contribution to project management (e.g.

communication with the laboratory assistant; organizing group meetings; preparing materials). For each of the tasks a score of between 3 and –1 was given (3, major contribution; 2, some contribution; 1, minor contribution; 0, no contribution; –1, a hindrance to the group). As previous research has demonstrated greater agreement in peer assessment on general impressions of an individual than on specific dimensions (Thornton & Byham, 1982; Klimoski & Brickman, 1987; Schectman, 1992), a total rating across the seven dimensions was calculated for each target. Accordingly, the maximum possible score for any individual was 21 and the minimum –7. For each group member their average rating out of 21 was calculated and then the average rating for each work group calculated. For each group member a ratio of their rating to the group average was calculated (contribution index) and used as a multiplier on the mark for the written report. If an individual contributed more than an average amount for their group, the ratio would exceed 1.0 and the mark for their written report increase. If an individual contributed less than an average amount for their group, the ratio would be < 1.0 and the mark for their written report decrease. It was acknowledged that it would be possible, using this ratio measure, for a student to score $> 100\%$ for this assignment and, if this occurred, the student would be credited with $> 100\%$ for this component of the course (i.e. they would receive more than the 'maximum' number of marks assigned to this component of course evaluation).

Results and discussion

Our analysis will be presented in four parts: (i) descriptive analysis of the contributions ratings; (ii) consideration of the relationship between the contributions ratings and performance on the written assignment; (iii) strategies; (iv) consideration of alternative ways of using the contributions ratings to calculate marks for the assignment.

Contributions ratings

Each student rated themselves and the other members of their work group. From these ratings a number of measures were calculated: ratings of their own contribution to the group (rating of self); the mean rating given by other members of the group to each individual (others ratings of self); the mean rating given by each individual to the other group members (ratings of others in group); the mean rating for each individual from themselves and other group members (own contribution); the mean rating for all members of the group (mean group contribution); the contribution index (own contribution/mean group contribution). The mean, median and range for each of these measures in shown in Table 1. The distribution of the contribution index measure is also shown in Figure 1.

A repeated measures ANOVA (rating of self; others rating of self; rating of others in group) revealed a significant effect [$F(2,120) = 3.49$, $P < 0.05$]. Ratings of the self were significantly higher (Fisher LSH, $P < 0.05$) than others rating of the self and rating of others in group, the latter two not differing from one another

Table 1. Contribution ratings

	Mean	Median	Minimum	Maximum
Rating of self	18.47	19.00	8.00	21.00
Others ratings of self	17.09	18.25	3.5	21.00
Ratings of others in group	17.00	18.00	7.33	21.00
Own contribution	17.36	18.00	8.50	21.00
Mean group contribution	17.32	17.85	13.00	20.75
Contribution index	1.002	1.008	0.562	1.256

(Means = 18.47 versus 17.09 and 17.00). Individuals saw themselves as contributing more to the group project than other members of their work groups. There was, however, no difference between own contribution (mean of self- and peer-ratings of self) and the mean group contribution (Ms = 17.36 versus 17.32), suggesting that once self and other perceptions of contribution were combined there was little difference in the scores of group members within a given workgroup. This is reflected in the mean and median values for the contribution index being very close to 1.0 (mean = 1.002, median = 1.008).

Additional analyses considering the impact of sex of student, size of group and type of group (mixed sex or all female) on the contribution measures revealed few significant effects. There were no significant effects of sex of student. There was an

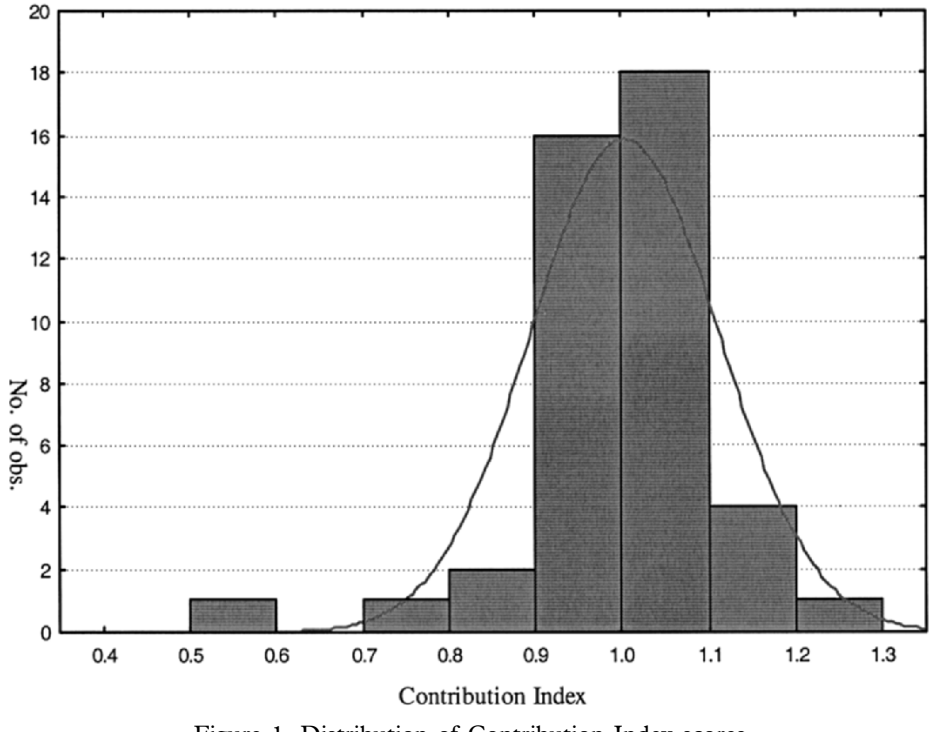

Figure 1. Distribution of Contribution Index scores

effect of group size on the ratings of self only, with participants rating their own contribution lower if they were in a group of 5 rather than a group of either 3 or 4 students [$F(2,58) = 6.07$, $P < 0.01$, Ms = 18.33, 19.30 and 16.47). There were no significant differences between the mean scores of those in mixed sex work groups and those in female only work groups. Considering only the female students, there were significant differences of group type (all female versus mixed sex) in the scores on the others ratings of the self [$F(1,46) = 4.60$, $P < 0.05$, Ms = 15.94 versus 18.21] and the ratings of others in the group [$F(1,46) = 4.34$, $P < 0.05$, Ms = 15.62 versus 18.02] with females both being rated higher by others and rating other group members higher when in mixed sex groups. Considering only students in the mixed groups there were no significant differences between male and female students on any of the contribution measures.

Pearson correlation coefficients were calculated between the ratings of the individual by themselves and by other group members. There were no significant correlations between ratings of the self and ratings of the self by others, i.e. there was no correlation between the ratings given to the self and given to the self by other group members [$r(61) = 0.159$, ns], indicating a difference in the perception of an individual by themselves and by other group members. This result parallels the results from the between-groups analysis reported above, which demonstrated a difference between self-ratings and others' ratings of the self and is consistent with the findings of Falchikov (1991). Similarly, there was no correlation between ratings of one's own contribution and that of other group members [$r(61) = 0.155$, ns], again suggesting a differentiation of own contribution from that of other group members. This again parallels the effect seen in the between-groups analysis, which showed individuals to rate their contribution higher than that of other members of their work group. Together these findings suggest that individuals are somewhat distorted in their view of their own contribution to the group project. This is also reflected in the absence of a significant correlation between ratings of the self and the self contribution index [$r(61) = -0.097$, ns].

Within each work group we calculated Kendall's measure of concordance, a measure of agreement in the ranking of items across judges and ranges (0, no agreement; 1, perfect agreement). Within each group we converted the contribution ratings provided by each group member to ranks, with higher ranks given to those who were rated higher on the contributions questionnaire. For each group we calculated Kendall's coefficient of concordance, once including self-assessment and once excluding self-assessment. Across the 15 groups the mean Kendall's coefficient of concordance was $W = 0.423$ (range 0.042–0.933) when self-assessment was included and $W = 0.584$ (range 0.204–1.00) when self-assessment was excluded [this difference was significant, $F(1,14) = 17.95$, $P < 0.001$]. There was relatively strong agreement within groups as to who were the greater and lesser contributors to the group project. Taking consensus as a reasonable indicator of accuracy, this finding would also suggest that students were indeed able to accurately evaluate the contribution of group members. Consistent with the previous findings of self-bias, agreement was greater when self-assessments were excluded. There was no effect of the sex of the group (all female versus mixed) on Kendall's coefficient of concord-

Table 2. Assignment marks before and after adjustment by the contribution index

	Mean	Median	Minimum	Maximum
Pre-adjustment				
Of 20	13.09	13.65	5.7	16.68
As a percentage	65.47	68.25	28.5	83.39
Post-adjustment				
Of 20	13.22	13.41	3.2	19.79
As a percentage	66.21	67.04	16.02	98.93
Change				
Of 20	0.132	0.138	− 4.42	4.04
As a percentage	0.741	0.689	− 22.12	20.18

ance. There was, however, an effect of group size with higher concordance for smaller groups. This effect was significant when self-assessment was excluded [$F(2,12) = 5.77$, $P < 0.05$, Ms = 1.0 for the groups with 3 members, 0.56 for the groups with 4 members and 0.43 for the groups with 5 members) and marginally significant ($P < 0.07$) when self-assessments were included (Ms = 0.8, 0.374 and 0.365). This finding may simply reflect the fact that the smaller a group is the more salient the contribution of each group member is or that with fewer judges there is less potential for variance of ratings.

Contribution ratings and performance

The marks awarded for the written assignment, before and after adjustment by the contribution index, are shown in Table 2. There were no effects of group type (mixed or all female) or group size on assignment marks. There was, however, a significant effect of sex on the assignment marks. Female students received higher marks before adjustment [$F(1,59) = 11.90$, $P < 0.01$, Ms = 67.73 versus 57.14%], a difference that was enhanced after adjustment for group contribution [$F(1,59) = 9.17$, $P < 0.01$, Ms = 69.07 versus 55.66%]. However, given the low number of male students ($n = 13$), this finding should be interpreted with caution.

There was no significant difference between the assignment marks before and after adjustment by the contribution index. Consistent with a mean contribution index very close to 1.0, there was very little mean change in the marks awarded (a median increase of 0.689%). There was, however, a wide range of changes to marks, from a decrease of 22.12% to an increase of 20.18%, as shown in Figure 2. Where there was a large discrepancy between the contribution of an individual and that of the work group this was reflected in a large impact on the assigned marks; the correlation between the contribution index and the change in percentage points was highly significant [$r(61) = 0.959$, $P < 0.001$].

Correlations between the contribution ratings and assignment marks are shown in Table 3. There was no relationship between ratings of self and assignment marks, either before or after adjustment [$r(61) = 0.132$, ns; $r(61) = 0.059$, ns]. Consistent with the previous findings, this suggests that individuals do not accurately evaluate

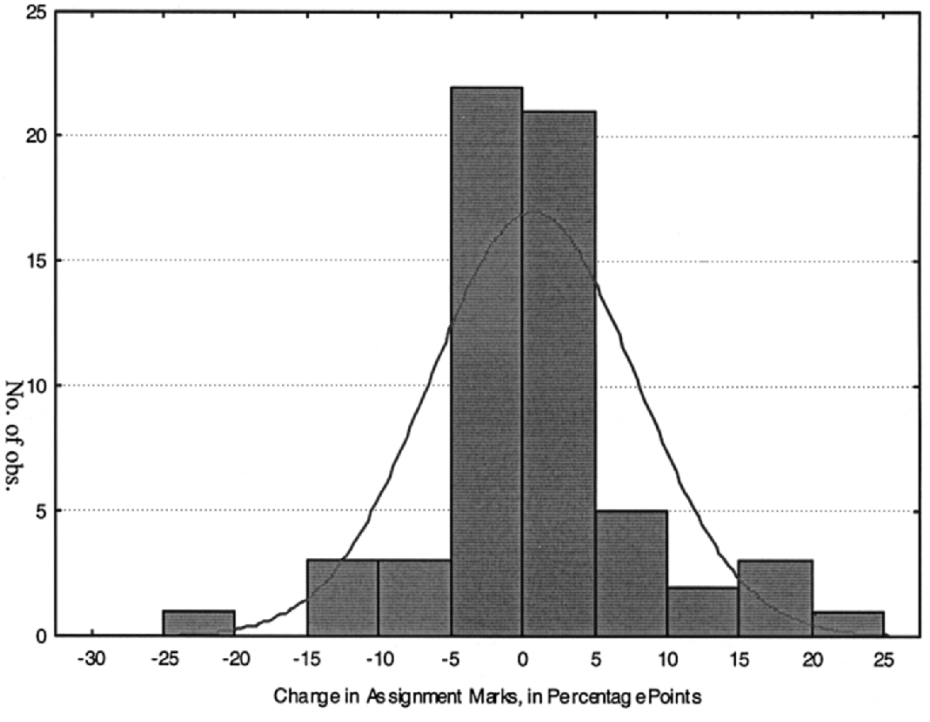

Figure 2. Distribution of change in assignment marks (as percentage points)

their own contributions. There was, however, a significant relationship between other's ratings of the individual and marks [$r(61) = 0.382$, $P < 0.01$; $r(61) = 0.534$, $P < 0.001$] and between the contribution index and marks [$r(61) = 0.484$, $P < 0.001$; $r(61) = 0.797$, $P < 0.001$]. These correlations indicate that those who contributed most to the group project also received higher marks for the project. It is also of interest that there is a significant difference ($P = 0.003$) in the correlation coefficients between the contribution index and marks before and after adjustment for contribution. The correlation is significantly stronger after adjustment of the marks. A strong contribution to the group effort does enhance overall performance on the assignment.

It has been suggested that self-assessment varies as a function of ability level, with higher ability individuals under-rating themselves and lower ability individuals

Table 3. Correlations between contribution measures and assignment marks

	Mark before adjustment	Mark after adjustment
Rating of self	0.132	0.059
Others rating of self	0.382[a]	0.534[b]
Contribution index	0.484[b]	0.797[b]

[a] $P < 0.01$.
[b] $P < 0.001$.

over-rating themselves (Lejk & Wyvill, 2001). To investigate this possibility amongst our participants we used a quartile split on the unadjusted marks to categorize participants according to ability level. A discrepancy score was calculated for each participant between their self-assessment and their assessment by peers; positive scores represented higher self- than peer-ratings and negative scores the converse. A single factor ANOVA (quartiles lowest/low medium/high medium/highest) on the discrepancy scores revealed a significant effect [$F(3,59) = 4.25$, $P < 0.01$]. *Post hoc* tests (Tukey HSD, $P < 0.05$) revealed a significant difference between those in the highest and lowest quartiles, but contrary to predictions, those in the highest quartile over-rated themselves whilst those in the lowest quartile under-rated themselves (Ms = 1.04 versus –0.76). Discrepancy scores for those in the middle quartiles feel in between the highest and lowest quartiles (Ms = 0.58 and –0.44).

Strategies

Somewhat surprisingly there appeared to be no instances of group collaboration over the ratings given. For no group were all group members given the same ratings by all members. Indeed, only 16 (26%) students failed to differentiate between group members and gave the same ratings to all members of their group, including themselves. A further 2 (3%) students rated all the other group members the same, but themselves differently. We interpret these data as indicative of the students taking the contributions rating task seriously. This is consistent with the relatively high mean values on the Kendall's concordance measure reported above. In discussing possible group strategies we are also mindful of the earlier analyses, which indicated a differentiation between the ratings given to the self and those given to the self by other group members; there was a tendency for participants to rate their own contribution as greater than that of other group members. This self bias may in itself have been sufficient to prevent group strategies developing. The contributions questionnaires were also filled in individually and were confidential such that it is possible that even if groups had discussed group strategies these were not adhered to in the actual ratings. There were indeed two work groups in which there may have been a conflict between prior discussions that all group members were to be given maximum marks and recognition of differential contributions—in these groups there was almost blanket ratings of all group members (e.g. in one group of 4, 2 group members gave all 4 members the maximum rating of 21 while 2 gave 2 members ratings of 21 and 2 ratings of 20).

Individuals could also have adopted strategies in their ratings in order to increase the likelihood of obtaining a high personal contribution index. This may be particularly pertinent to individuals who have contributed more or less than the average group member. Somebody who has made above average contributions to the group may attempt to emphasize this by rating their own contribution very highly but that of others lowly, rating themselves higher than others rate them and rating others lower than others do. Such a rating strategy would benefit the high contributing individual if the other group members similarly recognized their high level of contribution. Ironically, somebody who has contributed little to the group may show

a similar pattern of ratings in order to mask their lower contribution. Do our data show any evidence of such patterns? We looked at the students with the three highest (1.26, 1.25 and 1.23) and the three lowest (0.562, 0.646 and 0.796) contribution indices. The three with the highest indices did indeed rate themselves as the highest contributing group member, but in each case at least one other group member also did so. For the three lowest contribution indices, one individual also rated themselves as the lowest contributing group member. The other two individuals gave all the group members, including themselves, the highest contribution rating (the maximum ratings of 21), suggesting that they were indeed attempting to reduce the impact of their low contribution to the group (rated as 5 and 2 in one instance and 5, 4 and 4 in the other), but did not employ the optimal strategy to do so, only over-rating their own contribution and not denigrating the contribution of other group members. It is possible that students would feel more comfortable in inflating their own contribution than devaluing that of others. It is noteworthy that two of the work groups contained both one of the highest and one of the lowest scoring students (1.26 and 0.562 and 1.25 and 0.646) on the contributions index, suggesting that in each of these groups one group member may have compensated for a 'free-rider'.

Alternative approaches

The analyses above report detail regarding how the peer- and self-assessments were calculated and used in the grading of our laboratory assignment. Our data allow us to consider the possible impact of using the contributions ratings in different ways. We consider two such alternatives; removing the self-assessments and using the contribution index to allocate additional marks rather than as a moderator of the marks awarded for the written assignment.

There was no correlation between self- and peer-assessments, with self-assessments being higher, on average, than peer-assessments. Similarly, there was no correlation between self-assessments and marks awarded for the assignment, but a significant correlation between peer-assessments and marks awarded. This raises the question of whether inclusion of self-assessment leads to a less valid mark than if only peer-assessment were used (Falchikov, 1991; Lejk *et al.*, 1996). We calculated the contribution index for each student omitting the self-assessment ratings (contribution index = mean peer-assessment/mean group contribution). The mean peer-assessment ratings were the 'other ratings of self' scores considered above. The mean group score had to be recalculated, however, using only the peer-assessments. Differences between this revised index and that including self-assessments were small for the mean (1.00 versus 1.002) and median (1.02 versus 1.008) scores but somewhat larger for the minimum (0.27 versus 0.56) and maximum (1.42 versus 1.26) scores. There was, however, a very strong correlation between the two contribution index measures [$r(61) = 0.945$, $P < 0.05$]. The correlations between each of the indices and the assignment scores are also very similar, both pre-adjustment [index 1, $r(61) = 0.484$, $P < 0.05$; index 2, $r(61) = 0.431$, $P < 0.05$] and post-adjustment [index 1, $r(61) = 0.792$, $P < 0.05$; index 2, $r(61) = 0.727$, $P < 0.05$].

Table 4. Alternative calculations of the impact of the contribution index

Measure (contribution index of 1.0, equivalent to *x* marks/5)	Mean (out of 5 marks)	Median (out of 5 marks)	Minimum (out of 5 marks)	Maximum (out of 5 marks)	Mean difference between this measure (out of 20 marks) and that based on the contribution index ratio
3.25/5	3.26	3.28	1.83	4.08	− 0.148
4/5	4.01	4.03	2.25	5.02	0.603
5/5	5.01	5.04	2.81	6.28	1.61

The change to the assignment scores using this revised contribution index was small (mean–0.858%, median –0.912%) but with a wide range (from –30.57 to + 45.67%). There was no significant difference between the change to the assignment scores (Ms = 0.689 and –0.858) or in the final assignment scores using the two different contribution index measures. Further, the impact on the assignment scores using the two measures of contribution index correlated highly [$r(61) = 0.932$, $P < 0.05$], as did the two final assignment scores [$r(61) = 0.964$, $P < 0.05$].

In our assessment we used the ratings on the contributions questionnaire as a multiplier of the students' assignment marks and in doing so integrated the marks allocated to the process and the product of the group assignment. An alternative approach is to assign completely separate marks to these two aspects of the group project. For example, fewer possible marks might be allocated to the written assignment and additional marks attributed solely based on the contribution ratings. This would keep the total potential marks for the assignment constant but totally separate the marks awarded for process and for output. We considered one possible such approach, calculating grades on the basis of the marks awarded for the written assignment contributing 15% of the total course marks and those awarded based on the contributions rating contributing an additional 5% of the course marks. In using this method a decision must be made as to the number of marks out of 5 to allocate to an average contribution to the group (a contribution index of 1.0). We computed three measures, equating a contribution index of 1.0 to a mark of 3.25/5 which was equivalent to the mean percentage (65%) for the marked assignments prior to adjustment; equating a contribution index of 1.0 to a mark of 5/5, which is equivalent to assuming that average contribution was the maximum possible; equating a contribution index of 1.0 to a mark of 4/5 (80%). The impact of these different measures can be seen in Table 4.

Assuming that the average contribution was equivalent to a maximal contribution (contribution index of 1.0 = 5/5 marks) resulted in a mean allocation of > 5 marks for this component of the assessment and an average increase in marks of 1.61/20 (8.05%) for the assignment. Having the mean contribution being awarded maximum marks is somewhat unrealistic. Using the mean mark (65%) for the written assignments as the benchmark for the contributions marks (i.e. a contribution index of 1.0 = 65% or 3.35/5) is somewhat more realistic. This resulted in a mean mark (3.26/5) very close to the mean mark for the written assignments (9.75/15) and a total mark (out of 20) very close to that calculated above using the contribution index ratio. The mean difference in total marks between these two methods of calculation was –0.148/20 (0.74%). Equating a contribution index of 1.0 to a mark of 4/5 produced a slight increase in overall marks (0.603/20, 3.02%) compared with the index measure, but a mark of 4/5 (or 80%) is also unrealistic as a mean mark for an assignment at this level and therefore results in a greater relative weighting of the contributions component of the mark than the written assignment.

General discussion

We have discussed the nature and consequences of employing a self- and peer-as

sessment component to the awarding of marks for a group-based project in an advanced undergraduate social psychology laboratory course. We now highlight a number of key points involved in this method of assessment and consider the general usefulness of peer- and self-assessment.

On average, the contribution index measure employed in our laboratory had little impact on the marks awarded to students for this assignment, a mean increase of 0.7%. This lack of impact does not, in our view, however, indicate that inclusion of the peer evaluation was not worthwhile. Our aim in using peer-assessment was primarily to engage students in their work groups, to have them take responsibility for their learning (Michaelsen, 1992; Rafiq & Fullerton, 1996) and to minimize the likelihood of 'free-riders'. An average contribution index very close to 1.0 is consistent with achieving these aims as it indicates that most students were perceived to be 'pulling their weight' in their work groups. This interpretation is supported by evidence that students were taking the task seriously and differentiating between group members rather than simply giving 'blanket' ratings. There was a relatively high agreement on the rankings of the contributions of group members within each work group, suggesting that students did find it possible to evaluate relative contributions to the project by different group members. Further, the range of contribution index scores and the impact on assignment marks suggests that 'free-riders' were being penalized for their lack of contribution to the group effort and those who made above-average contributions to the group were being rewarded for so doing.

There were some noteworthy differences in our results and those of previous studies. Students in our sample rated their own contribution more highly than did their peers, in contrast to previous evidence that individuals tend to under-rate themselves (Krause & Popovich, 1996). Further, in our sample there was no effect of ability level on the tendency to over- or under-rate oneself (Lejk & Wyvill, 2001). The nature of the assignment in which we employed peer-assessment was somewhat different from that of most previous studies (Lejk *et al.*, 1996) with each student within each work group submitting a unique written assignment based on the group project which was marked by a laboratory tutor or lecturer and then moderated as a function of the peer-assessment regarding matters of process. It is possible that including an independently submitted written assignment and/or having the peer-assessment component confined to matters of process (Burke, 1969; Beard & Hartley, 1984; Rowntree, 1987) may explain the discrepancies between the present research and past studies, but further research is required before any firm conclusions can be drawn.

However, one concern raised by our analysis is the possibility of students receiving double rewards, or punishments, for their assignment through the grading and the contribution index measure. There was a strong correlation between the pre-adjustment marks and the contribution index measure, suggesting that those who contributed most were already receiving high marks whilst those who were not contributing were receiving low marks. It would appear that, despite it being a group project, it was difficult to achieve high marks on the written assignment if one did not also contribute to the group endeavours. Low marks and low contribution index scores could also, of course, indicate low motivation and engagement in the course.

Given the importance of both process and product in effective group work, the possibility of being twice rewarded is perhaps not of concern. Our analyses also considered the possibility of using the contributions index to award separate 'contribution marks' for the assignment rather than being used to moderate the awarded marks. Although one could argue that this separation of product and process marks is appropriate, it is noteworthy that the two measures have very similar impacts on students' marks for the assignment.

Whether self-assessment should be included within peer-assessment has been a controversial issue, with some researchers arguing that self-assessment should not be included (see for example Lejk & Wyvill, 2001) and others that it should (see for example Goldfinch, 1994). In our laboratory we used a contribution index calculated from both self- and peer-assessments of contribution to the assignment but, given previous questioning of whether self-assessment was a valid measure (Falchikov, 1991; Lejk *et al.*, 1996; Lejk & Wyvill, 2001), we also included analysis of the relative impact of a contribution index calculated solely on peer assessment. As expected (Falchikov, 1991; Lejk *et al.*, 1996), there was little relationship between self- and peer-assessments of the individual. Further, only peer-assessments, and not self-assessments, correlated significantly with both the contribution index and marks awarded. However, the impact on the assignment marks of the two contribution indices (including and omitting self-assessment) was very similar. The main impact of the exclusion of the self-ratings from the contribution index was to increase the spread of the distribution of that index, with the minimum index lower and the maximum index higher when self-ratings were omitted compared with when they were included. In turn, this resulted in a wider range of adjustments to the assignment marks, although the mean and median adjustments remained very similar.

Clearly the level of agreement between self- and peer-ratings is lower than that amongst peer-ratings (Falchikov, 1991; Goldfinch, 1994; Krause & Popovich, 1996; Lejk & Wyvill, 2001). Whether these findings should be taken as indicating that it would be better to omit than include self-ratings in peer-assessment is unclear. Self-assessment is an important skill to learn and omission of self-ratings from peer assessment may reduce the extent of such self-reflection. The key question is whether the learning benefits of self-reflection on one's own contributions toward group-based learning outweigh any small differences in the impact on the assignment grades. For most of the students in our research, omitting the self-ratings from the contribution index would have had a minimal impact on their final assignment grade, although doing so would have been advantageous to the top students and disadvantageous to the bottom students. In our opinion, the benefits of self-assessment in terms of the promotion of critical, reflexive thinking (Somervell, 1993) are sufficiently strong such that we will include self-assessments in future evaluations. Future research might consider investigating whether self- and peer-ratings within this assessment process might best be used in different ways. Peer-ratings might be best used to moderate grades, as described in this paper, whilst self-ratings might best be used in a developmental manner, teaching self-reflection and critical evaluation skills.

Conclusion

Peer-assessment allows for the contribution of individuals to a group project to be evaluated, and enhances the learning experience of students. Peer-assessment has been shown to promote independent, reflective, critical learners (Somervell, 1993), to enhance motivation for participation amongst students (Michaelsen, 1992) and to encourage students to take responsibility for their learning (Rafiq & Fullerton, 1996). Feedback from peer-assessment also allows students to learn about their own effectiveness in a group setting. In sum, our project supports the positive contribution of peer-assessment within student group-based assignments. While, for the majority of our students their final result for the assignment was not substantially moderated by the contribution index, the range of index scores indicated that this method of assessment was sensitive to both below- and above-average contributors. In addition, knowledge that individual contributions to the group project were to be assessed likely resulted in a low incidence of free-riding and, consequently, more involvement in group-based learning. We would endorse the inclusion of such a method of peer-assessment in estimating individual contributions to group productivity when assessing group-based projects at the tertiary level.

Acknowledgements

The contributions of Alice Boyes, Myron Friesen, Geraldine Henry and Nickola Overall are gratefully acknowledged.

Notes on contributors

Lucy Johnston is a Senior Lecturer in Psychology at the University of Canterbury, New Zealand. Her main research and teaching interests are in stereotyping and stigmatization.

Lynden Miles is a Ph.D. student in the Department of Psychology at the University of Canterbury. His research focuses on implicit processes in the perception of facial expressions of emotion.

References

Beard, R. & Hartley, J. (1984) Teaching and learning in higher education (4th edn) (London, Paul Chapman).

Burke, R. (1969) Self-assessments and peer ratings, *Journal of Educational Research*, 61, 444–448.

Conway, R., Kember, D., Sivan, A. & Wu, M. (1993) Peer assessment of an individual's contribution to a group project, *Assessment & Evaluation in Higher Education*, 19, 45–56.

Earl, S. E. (1986) Staff and peer assessment—measuring an individual's contribution to group performance, *Assessment & Evaluation in Higher Education*, 11, 60–69.

Falchikov, N. (1988) Self and peer assessment of a group project designed to promote the skills of capability, *Programmed Learning and Educational Technology*, 25, 327–339.

Falchikov, N. (1991) Group process analysis: self and peer assessment of working together in a group, in: S. Brown & P. Dove (Eds) *Self and peer assessment* (Birmingham, SCED Publications).

Fletcher, G. J. O., Simpson, J. A., Thomas, G. & Giles, T. (1999) Ideals in intimate relationships, *Journal of Personality and Social Psychology*, 76, 72–89.

Freeman, M. (1995) Peer assessment by groups of group work, *Assessment & Evaluation in Higher Education*, 20, 289–299.

Garvin, J. W., Butcher, A. C., Stefani, L. A. J., Tariq, V. N., Lewis, M. H. R., Blumson, N. L., Govier, R. N. & Hill, J. A. (1995) Group projects for first-year university students: an evaluation, *Assessment & Evaluation in Higher Education*, 20, 273–288.

Gibbs, G. (1992) *Improving the quality of student learning* (Bristol, Technical and Educational Services Ltd).

Goldfinch, J. (1994) Further developments in peer assessment of group projects, *Assessment & Evaluation in Higher Education*, 19, 29–35.

Goldfinch, J. & Raeside, R. (1990) Development of peer assessment technique for obtaining individual marks on a group project, *Assessment & Evaluation in Higher Education*, 15, 210–225.

Hughes, I. E. & Large, B. J. (1993) Staff and peer-group assessment of oral communication skills, *Studies in Higher Education*, 18, 379–385.

Habeshaw, S., Gibbs, G. & Habeshaw, T. (1993) *53 interesting ways to assess your students* (3rd edn) (Melksham, Cromwell Press).

Harvey, L. & Green, D. (1994) *Employer satisfaction* (Birmingham, University of Central England).

Hindle, B. P. (1993) The 'Project': putting student-controlled, small-group work and transferable skills at the core of a geography course, *Journal of Geography in Higher Education*, 17, 11–20.

Hufford, T. L. (1991) Increasing academic performance in an introductory biology course, *Bioscience*, 41, 107–108.

Hughes, I. E. & Large, B. J. (1993) Staff and peer-group assessment of oral communication skills, *Studies in Higher Education*, 18, 379–385.

Jacques, D. (1984) *Learning in groups* (London, Croom Helm).

Johnson, D. W. & Johnson, R. T. (1975) *Joining together. Group theory and group skills* (Englewood Cliffs, NJ, Prentice Hall).

Klimoski, R. & Brickman, M. (1987) Why do assessment centers work? The puzzle of assessment center validity, *Personnel Psychology*, 30, 353–361.

Krause, J. E. & Popovich, N. G. (1996) A group interaction peer/self assessment process in a pharmacy practice course, *American Journal of Pharmaceutical Education*, 60, 136–145.

Lejk, M., Wyvill, M. & Farrow, S. (1996) A survey of methods of deriving individual grades from group assessments, *Assessment & Evaluation in Higher Education*, 21, 267–280.

Lejk, M. & Wyvill, M. (2001) The effect of the inclusion of self-assessment with peer assessment of contributions to a group project: a quantitative study of secret and agreed assessments, *Assessment & Evaluation in Higher Education*, 26, 551–561.

Mello, J. A. (1993) Improving individual member accountability in small group settings, *Journal of Management Education*, 17, 253–259.

Michaelsen, L. K. (1992) Team learning: a comprehensive approach for harnessing the power of small groups in higher education, *To Improve the Academy*, 11, 107–122.

Orpen, C. (1982) Student vs. lecturer assessment of learning: a research note, *Higher Education*, 11, 567–572.

Rafiq, Y. & Fullerton, H. (1996) Peer assessment of group projects in civil engineering, *Assessment & Evaluation in Higher Education*, 21, 69–81.

Rowntree, D. (1987) Assessing students: how shall we know them (revised edition) (London, Kogan Page).

Schechtman, Z. (1992) Interrater reliability of a single group assessment procedure administered in several educational settings. *Journal of Personnel Evaluation in Education*, 6, 31–39.

Simpson, J. A., Fletcher, G. J. O. & Campbell, L. (2001) The structure and function of ideal standards in close relationships, in: G. J. O. Fletcher & M. Clark (Eds) *Blackwell handbook of social psychology: interpersonal processes* (Oxford, Blackwell), 86–106.

Somervell, H. (1993) Issues in assessment, enterprise and higher education: the case for self-, peer and collaborative assessment, *Assessment & Evaluation in Higher Education*, 18, 221–233.
Thornton, S. C. & Byham, W. C. (1982) *Assessment center and managerial performance* (New York, Academic Press).