

Peer assessment of language proficiency

Winnie Cheng and Martin Warren *The Hong Kong*

Polytechnic University

This article describes part of an investigation into the reliability and potential benefits of incorporating peer assessment into English language programmes. Undergraduate Engineering students attending a university in Hong Kong were asked to assess the English language proficiency of their peers - among other assessment criteria such as preparation, content, organisation, and delivery - as exhibited in the seminar, oral presentation and written report of an integrated group project. The paper compares the students' attitudes towards assessing both the English language proficiency and the other aspects of performance of their peers. It also compares peer and teacher assessments. The findings suggest that students had a less positive attitude towards assessing their peers' language proficiency, but they did not score their peers' language proficiency very differently from the other assessment criteria. Students and teachers were different in their respective marking behaviours and the ways oral and written language proficiency were interpreted. While students derived benefits from the peer assessment exercise, a question mark hangs over incorporating peer assessment for both language proficiency and the other criteria into the regular assessment process until such differences are resolved. Suggestions are made for improvement in procedures and future research.

I Introduction

Assessment is a critical activity in any instructional operation. One school of thought, which is increasingly gaining acceptance, argues that it is important for both learners and teachers to be involved in and have control over the assessment methods, procedures and outcomes, as well as their underlying rationale. Jafarpur (1991), for example, pointed out that if we are to increase the responsibility of the learner in EFL (English as a Foreign Language) study programs, this necessitates the adjustment of testing procedures.

Peer assessment, defined as ‘an arrangement in which individuals consider the amount, level, value, worth, quality of success of the products or outcomes of learning of peers of similar status’ (Topping, 1998: 250), is becoming more important as an alternative assessment method, among others such as self-assessment (see, for example, Boud, 1989; 1995; Tudor, 1996) and portfolio assessment (see for example, Mondock, 1997; Hamp-Lyons and Condon, 2000).

The benefits of incorporating peer assessment into the regular assessment procedures have been discussed in a number of studies (see for example, Burnett and Cavaye, 1980; Earl, 1986; Goldfinch and Raeside, 1990; Webb, 1995; Kwan and Leung, 1996). Peer assessment is believed to enable learners to develop abilities and skills denied to them in a learning environment in which the teacher alone assesses their work. In other words, it provides learners with the opportunity to take responsibility for analysing, monitoring and evaluating aspects of both the learning process and product of their peers. Research studies examining this mode of assessment have revealed that it can work towards developing students’ higher order reasoning and higher level cognitive thought (Birdsong and Sharplin, 1986), helping to nurture student-centered learning among undergraduate learners (Oldfield and MacAlpine, 1995), encouraging active and flexible learning (Entwhistle, 1993), and facilitating a deep approach to learning rather than a surface approach (Entwhistle, 1987; 1993; Gibbs, 1992). Peer assessment can act as a socializing force and enhances relevant skills and interpersonal relationships between learner groups (Earl, 1986).

Studies in which marks or grades were awarded by peers in the context of group work have been carried out in different disciplines, and data are mainly limited to student perceptions (Topping, 1998). Most of these studies are related to differentiating individual contributions to group projects (Conway *et al.*, 1993; Falchikov, 1993; Goldfinch, 1994). Others (Falchikov, 1993) involved group members and the teacher in negotiating peer, and self-assessment checklists of group process behaviours (Topping, 1998).

II Peer assessment in EFL contexts

A study of the literature in the EFL context has shown that peer assessment has been more commonly incorporated into English language writing instruction where peers respond to and edit each other's written work with an aim of helping with revision (see for example, Hogan, 1984; Birdsong and Sharplin, 1986; Lynch, 1988; Devenney, 1989; Jacobs, 1989; Rothschild and Klingenberg, 1990; Rainey, 1990; Bell, 1991; Mangelsdorf, 1992; Murau, 1993; Caulk, 1994; Mendonca and Johnson, 1994; Jones, 1995). All of these studies underscore the role and value of peer evaluation in TESOL writing instruction, generally in terms of developing the learners' writing ability, writing performance, and autonomy in learning. In Duke and Sanchez's (1994) study, students were given greater control in the writing assessment process in English classrooms and students' assessment criteria were compared with those of the Pennsylvania Department of Education. It was found (Duke and Sanchez, 1994: 50-53) that the students' assessment criteria were similar to those devised by the Department of Education; in addition, the assessment process helped to unify the writing process and the students also became willing evaluators. The study, however, did not compare peer and teacher assessments or examine students' attitude towards the assessment procedure.

Relative to research on peer assessment of writing, there have been much fewer studies on peer assessment of oral presentation skills. The relatively few studies reported improvement in marks and perceived learning (Watson, 1989; Falchikov, 1995a, 1995b) and significantly improved performance (Mitchell and Bakewell, 1995).

Some studies specifically compare teacher and peer assessment in the writing instruction in ESL/ EFL contexts. Topping (1998: 262) reviewed the literature relating to outcome studies of peer assessment of writing and found that it 'appears capable of yielding outcomes as least as good as teacher assessment and sometimes better'. Caulk's (1994) study found that the comments of the teacher and peer on L2 (second-language) writing serve important and complementary functions, and Devenney (1989) observed that the role and function of teacher evaluation differs from that of peer evaluation. Freeman (1995) compared assessments of oral presentation skills by groups of peers and teachers and found that the two populations were different in the marks awarded.

Some studies focus on the learners' feelings toward peer assessment. Birdsong and Sharplin (1986), for instance, reported that the overwhelming majority of students in their

study had a positive attitude towards evaluating the written work of their peers. Some research has, however, taken a more cautious view of the usefulness of peer feedback in writing, arguing for a greater degree of intervention in the process. Studies like Newkirk (1984) and Jacobs (1987) have suggested ways for the teacher to prepare the peers, for example, by demonstrating how peer evaluation works, explaining peer feedback to the learners, and setting up a structure for groups to work effectively.

To date, however, in an EFL context, there has not really been a comprehensive research on peer assessment that combines the following elements: investigating peer assessment of oral and written components in a group project, which leads to assigning marks or grades to the final pieces of work; comparing peer and teacher assessments of the same pieces of work; and finding out the students' attitudes toward participating in peer assessment exercise. What's more, studies that specifically compare different assessment criteria used in EFL peer assessment EFL contexts can rarely be found.

III Background of the study

In Hong Kong, the current education system has been found to over-emphasise academic knowledge and achievements measured solely by tests and examinations (Morris, 1996). The regular testing procedures in the classroom are basically administered by the teacher only and the notion of peer assessment is relatively new. A few studies, however, have been carried out in which students assessed their peers. Miller and Ng (1994), for example, conducted a study that involved 41 BA TESOL students assessing and assigning grades to their peers' oral language proficiency, as well as expressing their attitude towards participating in peer assessment activities. The results showed that students were able to realistically assess their group members' oral language ability under certain conditions, namely that the student assessors were high proficiency language learners, the group was homogenous, they had previous exposure to each other's oral language ability, testing environment was unthreatening, and the students were involved in the preparation of the test (Miller and Ng, 1994). The results also showed a relatively high level of agreement between peer assessments and lecturer assessments. Regarding attitudes, Miller and Ng (1994) found that their language students in general had a very negative attitude towards peer assessment, quoting reasons such as subjectivity of the

task, unfairness of the whole exercise, the strangeness of the activity, loss of face in front of classmates, feelings of being inexperienced, unqualified, and not proficient enough in English to assess their peers' oral proficiency.

Two other studies have been conducted in Hong Kong. Forde's (1996) study examined the consistency with which 50 Cantonese speakers of English in Hong Kong, enrolled in a preparation course for the International English Language Testing Systems (IELTS) examination, rated the oral ability of their peers. The study had shown that most of the learners were able to accurately assess their peers' oral proficiency, when compared with the assessments given by the IELTS-trained examiner. Patri (2002) studied the agreement between teacher, self assessment and peer assessment of students' oral skills with and without the use of peer feedback. It was found that there was higher correlation between teacher and peer assessments when the process was augmented by peer feedback, but that students were not able to judge each other's performances in the same way as the teacher had done regardless of whether or not peer feedback was incorporated into the process (Patri, 2002: 120-121).

Nevertheless, there appear to have been no studies of peer assessment that involves undergraduates being asked to assess the language proficiency of both the written and spoken English of their peers, nor any to compare peers' assessment of language proficiency as opposed to non-proficiency criteria in both oral and written work. The researchers of this study thought that in an EFL context, the inclusion of peer assessment of language proficiency would benefit learners by increasing their awareness of the variety and type of errors made and serve to illustrate the potential barrier to effective communication posed by these errors. If learners could be successfully trained to assess the language proficiency of their peers, this would have a positive effect when it came to self-correction.

The present study sought to add to the growing body of knowledge concerning the use of peer assessment in EFL contexts.

When preparing for this paper, the researchers were driven by three main concerns. First, we were interested in comparing the attitudes of these students towards peer assessment in general and towards peer assessment of language proficiency in particular. Second, we were aware that we needed evidence that our students were able to fairly and

responsibly assess their peers, both their language proficiency and non-language related aspects such as their capabilities on subject matter and organization, if we were to advocate this form of assessment across our department's English language programs. Third, we were interested in ascertaining the reliability of first-year Engineering students in supplementing teacher marks in peer assessment of the group project. The following research questions were therefore investigated:

1. Is there uniformity in students' attitude to the different assessment criteria?
2. How does peer assessment for language proficiency compare with teacher assessment of the same criterion?
3. How do peer and teacher assessment for language proficiency compare with their assessment for other criteria?

IV Method

1 Participants

Those who took part in the study were 51 (49 male and 2 female) first-year full-time Electrical Engineering undergraduates at the Hong Kong Polytechnic University, studying an English for Academic Purposes (EAP) subject. English is the official medium of instruction in this university. Prior to their entry into the university, these students had to have obtained an 'E' or above in Use of English in the Hong Kong Advanced Level Examination. Most of the participants scored an 'E' grade (which roughly corresponds to a score of 500 on the TOEFL examination) which is a typical score for first-year undergraduates in Hong Kong on non-language programs.

2 The study context

Given that this study assumes that it is reasonable to compare the students' marks with those of the teachers, the rationale for this assumption needs to be given. The three class teachers involved in the study were all experienced and used to working together as a teaching team. They all underwent rater training to facilitate inter-rater reliability and double blind marking along with moderation meetings to help to ensure the reliability of the teachers' assessment of the subject. In addition, the teachers were all experienced and used to working together as a teaching team and, based on all of these factors, we

assumed that the teachers' marks would serve as a reliable benchmark with which to compare the grades awarded by the students to their peers.

The students were taught by three teachers in class sizes of 16, 17 and 18. Peer assessment was a new concept for most of the students in the study and so a period of orientation and training was organized at the start of the EAP subject which lasted for 14 weeks with three semester hours in each week. In the first part of the subject, students were engaged in tasks in which they practised assessing their peers' spoken and written English, to discuss the assessment criteria, and any difficulties that they encountered. Halfway through the EAP subject, each class was divided into groups of 4-5 students to work on a group project that was composed of a seminar, an oral presentation, and a written report in sequential order of occurrence.

Across the three components, there were 17 assessment points altogether. First, the students are assessed on the 'language' used in the seminar, the oral presentation and the report. A second set of scales required individual students to assess the contribution of other members within the project group at various stages of preparing for and delivering the project components. The third set focused on peer assessment of nonlanguage criteria, namely 'preparation and content' and 'delivery' for both the seminar and oral presentation; and 'preparation and content', 'organization', 'layout and presentation' and 'writing style' for the report. These assessment criteria had been determined and validated by the materials writers based in the English Department.

3 Data collection procedures

At each stage of the study, the assessment criteria for each project component were explained to and discussed with the students in class with reference to the assessment forms (see Appendices 1a, 1b & 1c) for the assessment forms that list the assessment criteria for the seminar, the oral presentation and the written report). Then these components were assessed on a scale from 1 to 5 (1 being the lowest and 5 the highest) by both the peers and the teacher. When a group of students was giving a seminar and later on an oral presentation, both the teacher and the rest of the class assessed the performance of the individual members of the group by completing a form. Group reports were assessed in a similar way, except that individual members of the group were

awarded the same set of marks by both the teacher and students belonging to other groups. Within each group, individual students also gave an assessment of the contribution of other members (see Appendix 2). It should be pointed out that the marks awarded by the students constituted 50% of the marks awarded for the group project and 20% of the final grade for the entire EAP subject.

Students' attitudes were monitored before and after peer assessment exercise by means of a four-item questionnaire (see Appendices 3a & 3b), which was partly based on one used by Burnett and Cavaye (1980). Then three student feedback forms (see Appendices 4) were designed and administered to investigate the extent to which the students felt comfortable in assessing the performance of their peers and the extent to which they thought they assessed their peers fairly and responsibly on each of the 17 assessment scales. The purpose was to find out first, how the students' attitudes towards individual assessment criteria compared and second, whether the language proficiency criteria for the seminar, the oral presentation and the report respectively were viewed more or less positively than the other criteria (research questions 1 and 2).

Once the pre- and post questionnaire data had been analysed, students who displayed a marked change in responses and attitudes were interviewed. A marked change in response was taken as being cases where students responded differently in three out of the four questions asked. As a result, one third of the students were interviewed (see Cheng & Warren, 1997) at the end of the peer assessment exercise. The interviews were semi-structured and each lasted approximately 15 minutes. The main questions concerned the reason for the shift in attitude towards peer assessment, the students' feeling toward assessing different criteria, and whether they had had any training in peer assessment in secondary school.

V Results and discussion of attitudes towards assessing language proficiency

Tables 1 and 2 below show students' responses to the two items which they rated for each criterion on a five-point scale (1 = strongly agree, 5 = strongly disagree). The two items are 'I felt comfortable in assessing the performance of my peers on each of the assessment criteria' and 'I think I assessed my peers fairly and responsibly on each of the assessment criteria'.

The tables show that the students' attitudes towards the individual assessment criteria differed; however, the ranges of means were quite narrow. It was nonetheless interesting to observe the relative ranking of the assessment criteria and, as language teachers, we were particularly interested in finding out how the language proficiency criteria were ranked relative to the other criteria. The *F* values in the two tables represent the results of a series of analyses of variance to test whether the mean response to the question about each criterion was significantly different from the overall mean response to the questions about the criteria collectively. It was decided to use multiple inferential statistics in this study (see also Table 4) as it is argued that the observations collected and analysed are independent of each other.

Table 1 Means and standard deviations of students' responses to the question:

I felt comfortable in assessing the performance of my peers on each of the assessment criteria.

(1 = strongly agree, 5 = strongly disagree)

Assessment Criteria	M	SD	N	F	Sig.
1 Contribution of Group Members - Preparation	2.52	.76	50	2.54	.09
2 Contribution of Group Members - Ideas & Suggestions	2.54	.86	50	0.62	.54
3 Oral Presentation – Preparation & Content	2.56	.73	50	0.81	.45
4 Oral Presentation – Delivery	2.56	.73	50	0.51	.60
5 Seminar – Delivery	2.59	.83	51	3.20	.05*
6 Report – Organization	2.62	.81	50	2.29	.11
7 Report - Layout & Presentation	2.62	.81	50	1.44	.25
8 Report – Preparation & Content	2.63	.75	51	4.44	.22
9 Seminar – Language	2.66	.87	50	0.25	.78
10 Contribution of Group Members – Oral Presentation	2.68	.74	50	1.29	.29
11 Report - Writing Style	2.72	.76	50	1.34	.27
12 Contribution of Group Members – Writing	2.72	.78	50	3.64	.03*

13 Oral Presentation – Language	2.74	.83	50	0.33	.72
14 Contribution of Group Members - Literature Analysis	2.74	.85	50	0.41	.67
15 Seminar - Preparation & Content	2.80	.81	50	0.50	.61
16 Contribution of Group Members - Literature Search	2.80	.76	50	0.59	.56
17 Report – Language	3.02	.65	50	2.18	.12

* Criteria that are significantly different.

Table 2 Means and standard deviations of students' responses to the question:
I think I assessed my peers fairly and responsibly on each of the assessment criteria.

(1 = strongly agree, 5 = strongly disagree)

Assessment Criteria	M	SD	N	F	Sig.
1 Contribution of Group Members - Preparation	2.44	.74	48	0.92	.40
2 Oral Presentation – Preparation & Content	2.48	.65	48	0.16	.85
3 Contribution of Group Members - Ideas & Suggestions	2.50	.71	48	1.78	.18
4 Report - Layout & Presentation	2.50	.71	50	0.59	.56
5 Oral Presentation – Delivery	2.54	.68	48	3.18	.05*
6 Seminar – Delivery	2.57	.71	49	4.86	.01*
7 Report – Organisation	2.58	.73	50	1.35	.27
8 Contribution of Group Members – Literature Search	2.58	.74	48	0.20	.82
9 Contribution of Group Members – Oral Presentation	2.60	.76	48	1.34	.27
10 Contribution of Group Members – Writing	2.60	.68	48	2.95	.06
11 Report – Preparation & Content	2.61	.81	49	1.44	.25

12 Seminar – Preparation, Overall presentation & Content	2.66	.77	50	2.30	.11
13 Seminar – Language	2.68	.71	50	1.01	.37
14 Oral Presentation – Language	2.69	.62	48	1.83	.17
15 Report – Language	2.71	.71	49	2.07	.14
16 Report - Writing Style	2.72	.81	50	0.38	.68
17 Contribution of Group Members - Literature Analysis	2.77	.66	48	0.92	.41

* Criteria that are significantly different.

In Table 1, the language proficiency criteria were ranked 9 (seminar), 13 (oral presentation) and 17 (written report), where 17 was the assessment criterion which the students felt the least comfortable assessing. These rankings would seem to show that the students, generally, were less comfortable with awarding marks for their peers' language proficiency than for most of the other criteria. In Table 2, the three language proficiency criteria were ranked down at the bottom: 13 (seminar), 14 (oral presentation) and 15 (written report). In other words, students thought that they assessed their peers' language proficiency less fairly and responsibly, compared to their assessment of almost all of the other criteria. Among seminars, oral presentations and written reports, assessing the language of their peers' written reports was in both cases rated the lowest. Possible reasons for this were suggested in the follow-up interviews. However, it needs to be pointed out that the results of a Bonferroni test indicated only four criteria in all (marked with asterisks in Tables 1 and 2) were significantly different from any other, and none of these were language criteria which suggests that the relative ranking of the criteria should be read with extreme caution.

In the interview, all of the interviewees stated that there was not necessarily a contradiction in feeling uncomfortable about assessing one's peers but still doing it fairly and responsibly; as one interviewee stated: 'it's always going to be difficult at times to give true marks to classmates'. Four students stated that they believed the marks they gave to peers with whom they had a closer relationship were not awarded entirely fairly

and responsibly. Comments such as ‘avoiding hurting feelings’, ‘not wishing to embarrass my friends’ were made by this small group to justify giving, in the words of one student, a ‘reasonable 3 rather than an actual 2’. Fifteen out of 17 students, however, claimed to have managed the tension between feeling uncomfortable and assessing their peers fairly and responsibly despite feelings of ‘embarrassment and unease’.

When asked specifically why they were less comfortable when assessing their peers’ language proficiency and why they had doubts as to being capable of doing this fairly and responsibly, most of the students interviewed responded that they had no idea how to go about assessing the spoken and written language proficiency of their peers. This in turn was a result of ‘not knowing’ what constituted high and low language proficiency. The students felt they lacked experience in terms of what constituted language proficiency and would benefit from ‘more practice at grading classmates’. In addition, half of those interviewed felt that they were unable to assess their peers because of their own poor levels of English language and so felt themselves to be unqualified for the task. As one student said ‘my poor English standard and abilities in English communication made me confusion to give a fair mark’. Consequently, most of the interviewees were more uncomfortable and more uncertain when it came to assigning marks for the language criteria fairly and responsibly. The only exception to this general observation was that they were able to assess more comfortably, fairly and responsibly the ‘oral fluency’ (i.e. pronunciation, pausing, timing, pacing, etc.) of their peers manifested in the seminars and oral presentations. Presumably the reason for this is that assessing what the students characterized as ‘oral fluency’ does not require the same linguistic competence deemed by the students to be necessary for assessing, for example, grammatical accuracy. It seems that when the students assessed the spoken language proficiency of their peers, they gave greater weight to ‘oral fluency’ than vocabulary, structures and so on, assessment criteria which the teacher was equally concerned with.

What happened, then, when the students were evaluating the language of written reports which did not exhibit ‘oral fluency’? Our follow-up interviews suggested that the students tended to conflate judgements regarding content, layout, organization and so on with language. In other words, the mark awarded for language seems to have amounted to a kind of aggregate of the marks they had awarded for the other less ‘problematic’

assessment criteria. Devenney (1989: 85-86) also notes that when evaluating writing, peers in an ESL context, unlike their English teachers, did not use grammar as a basis but they made proportionately more comments on content than the teachers.

To sum up, we have evidence from the students in this study that when the peers awarded marks for spoken language proficiency, they may not have included all the elements of language proficiency assessed by teachers; and in the case of written language proficiency, strictly speaking, the students may not have been assessing language proficiency at all.

VI Results and discussion of peer and teacher assessment of language proficiency

As shown in questionnaires and interviews, the students were more uncomfortable and less confident of their ability to assess fairly and responsibly when it came to assessing the English language proficiency of their peers. We wanted to see if these concerns had been translated into differences in the actual marks that the students awarded for the language proficiency compared with the marks they gave for other criteria. We therefore established first, the degree of agreement in judgement between peer and teacher ratings for language; and second, peer ratings for language as opposed to peer ratings for non-language assessment criteria in the group project.

1 Comparison of teacher and student means

We began analyzing peer and teacher assessments by examining the descriptive statistics for each of the twelve assessment criteria that both students and teachers assessed (the remaining five assessment criteria were only applicable to the students as they assess contributions made by group members). One measure of agreement between the two sets of results is that the students' mean mark lies within one standard deviation of the teacher's mean mark (Kwan and Leung, 1996: 207-208). Table 3 shows the means and standard deviations for the marks given by teachers and students for each assessment criterion of the three project components.

Table 3 Means and standard deviations of marks awarded for each of the assessment criteria

Seminar	Class A				Class B				Class C			
	Teacher's marks		Students' Marks		Teacher's Marks		Students' Marks		Teacher's marks		Students' marks	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Preparation, Presentation & Content	2.56	0.81	3.38	0.26	4.12	0.70	3.33	0.28	4.06	0.64	3.08	0.29
Delivery	2.81	0.91	2.41	0.56	4.00	0.50	3.15	0.21	3.39	0.61	3.02	0.33
Participating in discussion	2.81	0.91	3.13	0.17	2.82	1.29	2.59	0.92	3.50	0.71	2.91	0.17
Language	2.75	1.00	3.39	0.31	3.65	0.61	3.16	0.18	3.44	0.78	2.93	0.25

Oral Presentation	Class A				Class B				Class C			
	Teacher's marks		Students' Marks		Teacher's Marks		Students' Marks		Teacher's marks		Students' marks	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Preparation & Content	3.50	0.97	3.34	0.34	3.41	0.62	3.39	0.21	3.78	0.43	3.03	0.22
Delivery	3.19	0.91	3.24	0.29	3.35	0.79	3.27	0.30	3.00	0.69	2.94	0.27
Language	3.38	0.96	3.26	0.26	3.24	0.66	3.20	0.20	3.17	0.62	3.01	0.32

Report Writing	Class A				Class B				Class C			
	Teacher's marks		Students' Marks		Teacher's Marks		Students' marks		Teacher's marks		Students' marks	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Preparation & Content	3.75	0.45	3.46	0.21	3.82	0.39	3.72	0.41	4.00	0.00	3.42	0.16
Organisation	3.25	0.45	3.23	0.07	3.59	0.51	3.60	0.29	3.00	0.00	3.20	0.32
Language	3.00	0.00	3.31	0.31	2.29	0.47	3.23	0.08	3.00	0.00	2.90	0.11
Writing style	3.00	0.00	3.23	0.16	3.00	0.00	3.40	0.26	3.00	0.00	2.92	0.17
Layout & presentation	3.50	0.89	3.29	0.25	3.65	0.79	3.26	0.38	4.28	0.46	3.16	0.18

This initial measure of agreement shows that for seminars and oral presentations, the students' mean marks for the language criteria lay within one standard deviation of the class teacher's. In other words, there is agreement between the two sets of marks. This is also the case for all of the other assessment criteria for these two project components except for the first criterion (preparation and content) for Classes B and C in seminars and Class C in oral presentations and, on one occasion, the delivery assessment criterion for Class B in seminars.

This method of measuring agreement is, however, somewhat problematic when it comes to the report writing component. Since the reports were written by groups of four to five students, there were only four written reports in each class and, in some instances, the class teacher awarded the same mark for certain assessment criteria to all of the written reports in the class. The consequence of this is that the standard deviation becomes zero and so agreement only exists if the students' and teacher's marks are exactly the same. We are therefore unable to evaluate the consistency of these sets of marks for written reports using this form of measurement, but other statistical tests were applied later to measure agreement between teachers' and students' marks for report writing.

On the whole, the standard deviations of the students for both language proficiency and non-language assessment criteria were consistently lower than those of the teachers, reflecting the tendency of the students to mark within a narrower range. In fact, students in this study have been found to be generally marking their peers within a narrower range than the class teachers to the benefit of the weaker students and to the detriment of the more able students. This observation has been made in other studies on peer assessment (see, for example, Freeman, 1995; Kwan and Leung, 1996) and is usually ascribed to the reluctance on the part of students to mark their peers up or down. Here we are primarily concerned with the marking behavior of students when assessing language proficiency compared to the other assessment criteria. We can therefore conclude that in terms of agreement with the teachers' marks and size of standard deviations, the students' scoring patterns for language were not noticeably different from those for the other criteria.

Table 4 presents the results of paired *t*-tests applied to the mean scores of the peer and teacher marks for individual assessment criteria of the three project components for all classes. Our null hypothesis was that there were no differences between the mean scores of the teacher marks and student marks of Classes A, B and C for all the assessment criteria of the group project components.

Table 4 Paired *t*-tests: teachers' and students' marks for each of the assessment criteria

Seminar	Class A		Class B		Class C	
	<i>T</i>	<i>p</i>	<i>T</i>	<i>p</i>	<i>t</i>	<i>P</i>
<i>Preparation,</i>	4.45	<0.01	5.09	<0.01	6.00	<0.01

Content						
Delivery	1.72	NS	7.19	<0.01	2.80	NS
Participating in discussion	-1.52	NS	1.27	NS	3.61	<0.01
Language	-3.08	<0.01	3.40	<0.01	2.89	<0.01

Oral Presentation	Class A		Class B		Class C	
	<i>T</i>	<i>p</i>	<i>T</i>	<i>p</i>	<i>t</i>	<i>P</i>
Preparation, content	0.78	NS	0.18	NS	7.76	<0.01
Delivery	-0.23	NS	0.49	NS	0.38	NS
Language	0.53	NS	0.27	NS	1.00	NS

Report Writing	Class A		Class B		Class C	
	<i>T</i>	<i>p</i>	<i>T</i>	<i>p</i>	<i>t</i>	<i>P</i>
Preparation, content	3.84	<0.01	1.32	NS	15.69	<0.01
Organization	0.21	NS	-0.10	NS	-2.69	NS
Language	-4.02	<0.01	-9.57	<0.01	3.90	<0.01
Writing style	-5.84	<0.01	-6.19	<0.01	2.09	NS
Layout and presentation	0.92	NS	2.93	<0.01	12.32	<0.01

$p < 0.01$; NS = Not Significant

The results of paired *t*-tests were not consistent across all components of the group project and we will therefore discuss each one in turn. In seminars, for all classes, the *t*-tests for the language proficiency criterion showed a significant difference ($p < 0.01$) between teacher's and students' marks. The same was the case for the first criterion (preparation and content). Significant differences were also found for two of the six remaining assessment criteria (all cases where there is a significant difference are in bold type). For oral presentations, the null hypothesis failed to be rejected with only one exception. For all classes, no statistically significant differences were found for the language proficiency criterion. Only in Class C did the peer mean scores given for the first criterion (preparation and content) differ significantly ($t = 7.76$; $p < 0.01$) from the teacher's mean mark.

A possible reason for the difference between seminars and oral presentations is that the seminar preceded the oral presentation by approximately five weeks and the students discussed and received feedback on their seminars which had hopefully raised their awareness. The assessment of the seminars had also given students useful practice in assessing their peers and may have helped to bring their marks and the class teacher's marks closer together in the assessment of the oral presentations.

In the case of the written reports, the *t*-tests showed that the students' mean marks for the language proficiency criterion differing significantly ($p < 0.01$) across Classes A, B and C. This phenomenon was not only observed in this particular criterion. In fact, for most of the other criteria and for all three classes, significant differences were also found (i.e. 9 out of the 15 criteria showed significant differences between teachers' and students' marks). The findings suggest that in assessing their peers' written language proficiency in reports, the students were more consistently different from their class teacher but such differences were also evident in the non-language assessment criteria. It needs to be added that claims based on a relatively small number of *t*-tests need to be substantiated with a larger number of subjects before they can be generalized.

Why is there this apparent difference between oral presentations and written reports, both of which were assessed at the end of the course? We venture to suggest that they involve different kinds of skills and the students had more practice in the former within the course. In addition, students in this study found it less problematic to assess certain aspects of the spoken language of their peers than written language. They felt more able to assess 'oral fluency', but when faced with a written text which lacks this element, they effectively generated an aggregate of the marks awarded for content, presentation, layout, and so on, and assigned it to language proficiency.

2 *The range of marks awarded by teachers and students*

Given the students' lower standard deviations evidenced across the data, we thought it would be useful to look in more detail at the phenomenon by comparing the actual ranges of marks awarded by the teacher and individual students respectively. We have done this in two ways: first by comparing the actual maximum and minimum marks given by the two sets of assessors (Table 5), and second by comparing individual students' language marks against teachers' language marks and individual students' overall project marks against teachers' overall project marks.

Table 5 Maximum and minimum marks awarded for each of the assessment criteria

Seminar	Class A		Class B		Class C	
	Teacher's	Students'	Teacher's	Students'	Teacher's	Students'

	marks		Marks		marks		marks		Marks		marks	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
Preparation, Presentation & Content	1.00	4.00	2.00	4.00	3.00	5.00	2.00	5.00	3.00	5.00	2.00	4.00
Delivery	2.00	5.00	2.00	5.00	3.00	5.00	2.00	5.00	2.00	4.00	2.00	4.00
Participating in discussion	2.00	5.00	2.00	5.00	1.00	5.00	1.00	5.00	2.00	5.00	2.00	4.00
Language	2.00	5.00	2.00	5.00	3.00	5.00	2.00	5.00	2.00	5.00	1.00	4.00

	Class A				Class B				Class C			
	Teacher's marks		Students' Marks		Teacher's marks		Students' marks		Teacher's Marks		Students' marks	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
Preparation & Content	2.00	5.00	2.00	5.00	2.00	4.00	2.00	4.00	3.00	4.00	2.00	4.00
Delivery	2.00	5.00	2.00	5.00	2.00	5.00	2.00	4.00	2.00	4.00	2.00	4.00
Language	2.00	5.00	2.00	5.00	2.00	4.00	2.00	4.00	2.00	4.00	2.00	5.00

	Class A				Class B				Class C			
	Teacher's marks		Students' Marks		Teacher's marks		Students' marks		Teacher's Marks		Students' marks	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
Preparation & Content	3.00	4.00	2.00	5.00	3.00	4.00	3.00	5.00	4.00	4.00	2.00	4.00
Organisation	3.00	4.00	3.00	4.00	3.00	4.00	2.00	5.00	3.00	3.00	2.00	4.00
Language	3.00	3.00	2.00	4.00	2.00	3.00	2.00	5.00	3.00	3.00	2.00	3.00
Writing style	3.00	3.00	2.00	4.00	3.00	3.00	2.00	5.00	3.00	3.00	2.00	4.00
Layout and presentation	2.00	4.00	2.00	5.00	2.00	4.00	2.00	5.00	4.00	5.00	2.00	4.00

Table 5 shows that when assessing their peers' oral language proficiency in seminars and oral presentations, individual students in each class awarded marks that fell within a range similar to that of their class teacher's, and in the case of Class B, at times even a wider range than the teacher's. In the assessment of written reports, however, the students' maximum and minimum marks awarded for language proficiency were respectively higher and lower than the teacher's maximum and minimum marks. This phenomenon is not confined to the language marks only. When compared to the class teacher's, the students' marks for the other report writing criteria were also found to be in a wider range. This is possibly a result of the four written reports effectively being

marked by different assessment criteria used by the students and teacher, with the students conflating content, layout and so on with language, a behaviour which was suggested in the follow-up interviews and discussed earlier.

In all classes, when language marks were combined, students' marks fell within a narrower range than those of their teachers. The ranges were not just narrower. It is also the case that the teachers' upper and lower limits were consistently higher and lower than the students' upper and lower limits (Class A: students: 61-82, teacher: 45-85; Class B: students: 42-67, teacher: 34-74; Class C: students: 37-68, teacher: 34-80). These findings confirm the evidence provided by the standard deviations discussed earlier that the students marked within a narrower range compared with their class teacher.

The patterns of marks when the teachers' and students' assessed the entire project with all the criteria combined were also examined. The ranges of marks were as follows: Class A: students: 59-73, teacher: 46-81; Class B: students: 60-71, teacher: 57-85; Class C: students: 60-65, teacher: 59-78. It can be seen that the ranges of marks given by the students are narrower, and both the upper and lower limits are lower and higher respectively, compared to the ranges of marks given by the class teachers. This shows that, in this regard, the marking behavior of students did not differ regardless of whether they were marking language proficiency or the other assessment criteria.

VII Conclusions

The first-year undergraduate Engineering students in this study were found to feel less comfortable and more uncertain of their ability to assess the language proficiency of their peers compared to the other criteria they were asked to assess. Most of the students felt unqualified to assess their peers' language proficiency for two main reasons: they felt unsure as to what constituted proficiency in English as a foreign language and they thought that their own levels of linguistic competence were insufficient for the task. Assessing what students described as 'oral fluency' in the seminar discussions and oral presentations was the one exception to this widely held view.

Agreement in judgement between student and teacher assessments was found in both language and non-language related criteria. Our analysis revealed that students tended to mark within a narrower range than their class teacher. The standard deviations

of students were approximately half those of their teachers which suggests that the two marker populations were different. Despite this, we might expect the students to award a wider range of marks if they were given more opportunities to practice and experience peer assessment procedures.

When assessments for oral language proficiency were analyzed, significant differences between peer and teacher assessments were more prevalent in seminars than in oral presentations for both the language proficiency and other assessment criteria. This phenomenon is probably due to the fact that during the intervening weeks the students discussed, received feedback on, and practiced the peer assessment techniques. Concerning assessment of the language proficiency of written reports, peer assessments were significantly different from the teachers', though differences were also found in most of the non-language assessment criteria.

An interesting observation emerged from the follow-up interviews with the students, i.e., the students did not perceive their peers' oral and written language proficiency in the same way. Oral language proficiency was associated more with 'oral fluency' and written language proficiency was often turned into an aggregate of the other criteria. This therefore suggests that the students did not in effect assess the same elements as their class teachers.

The pedagogical and developmental benefits of involving students in the assessment process were confirmed by both the teachers and the students who took part in this peer assessment exercise. Nonetheless, this study has demonstrated that our first-year Engineering students did not reliably supplement their teachers' marks in assessing both the language and non-language assessment criteria. Jafarpur (1991) also concludes from his study of English major undergraduates in Iran that EFL learners are not able to make sound judgements about their own or their peers' oral English proficiency.

We need to be as confident as our students in their ability to assess language proficiency along with other aspects of performance before we make peer assessment a part of the regular assessment process. We believe that if language learners could be trained to confidently and reliably assess the language proficiency of their peers, they would also be able to confidently evaluate their own language skills, a valuable precondition for improving them. This positive wash-back effect was evident during the

initial orientation process and in the discussions with students concerning the assessment criteria and procedures. By the end of the peer assessment exercise, students generally had acquired a reasonable grounding in peer assessment procedures and were favorably disposed to participating in peer assessment in the future. Both teachers and students found the peer assessment exercise beneficial in terms of developing students' higher level cognitive thinking and facilitating a deep approach to language learning.

The exercise reported in this paper has raised the awareness of both teacher and students of a range of assessment issues. Most important of all, decisions regarding whether to include, or exclude, peer assessment in academic programmes should not be solely based on the level of agreement between peer and teacher marks, but should rather be made after consideration of the positive impact peer assessment can have in other respects.

VIII Implications for classroom and research

In terms of classroom implications and future research, the following could be incorporated into the peer assessment design and procedures. First, students reported a low level of comfort and a low degree of confidence in their ability to fairly and responsibly assess their peers' language proficiency. It is therefore worth investigating whether a classroom with supportive learning climate would lead to positive attitudes towards peer assessment and willingness to give an objective assessment. It would also be worth investigating whether more, and more carefully, structured awareness-raising and training activities across a spread of subjects within a study program might change these views. Some studies have observed a positive connection between prior training, accurate peer assessment, and a favorable attitude toward the notion of peer assessment (Williams, 1992; Forde, 1996). Second, students could be involved in the design and development of the assessment criteria for the various components of the group project, a factor which others have found to be beneficial (Williams, 1992; Patri, 2002). A participatory and negotiable process as such would enable the students to perceive the peer assessment exercise more positively and boost their confidence in their ability to carry out the task. Any assessment criteria used need to be clarified and exemplified. Third, this study used a group of non-language specialist students as subjects. Miller and

Ng (1994) stated that proficient and highly motivated L2 learners are able to more realistically assess their peers' language ability. Similar studies could be done with groups of participants of very different discipline backgrounds and L2 proficiency levels for comparison purposes. Fourth, the fact that nearly all of our subjects (49 out of 51) were male may have affected the results in some ways. Future studies should take possible gender differences into consideration when designing their methodology and interpreting their findings. Fifth, we are aware that due to the homogeneity of our participants in terms of discipline of study and gender, our results can not be over-generalized. To a certain extent, our findings may have been a result of the possible effects of the nature and characteristics of our subjects, who are technically-oriented, analytical and mathematical rather than language-oriented. Future studies need to be conducted to confirm our findings. Last, before the study was conducted, efforts had been made to ensure that the three teachers had a common understanding of the method of study and prepared their respective classes in the same manner, however, future studies might consider the possible effects of teacher differences by, for example, interviewing the teachers both prior to and after the study.

We share the view of Etheridge (1995) who argues that peer assessment can work effectively if the teacher is more concerned with the long-term, cumulative educational benefits rather than simply the immediate success or failure of students' attempts to imitate or supplement the assessment behavior of their teacher.

Acknowledgements

We are grateful to Ronald Chan for his help in preparing the statistical data and to the anonymous reviewers for their invaluable comments and suggestions. The work described in this paper was substantially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region (Project No. G-S252).

IX References

Bell, J. 1991: Using peer responses in ESL writing classes. *TESL Canada Journal* 8, 65-71.

- Birdsong, T. and Sharplin, W.** 1986: Peer evaluation enhances students' critical judgement. *Highway One* 9, 23-28.
- Boud, D.** 1989: The role of self-assessment in student grading. *Assessment and Evaluation in Higher Education* 14, 20-30.
- Boud, D.** 1995: *Enhancing Learning Through Self-assessment*. London; Philadelphia: Kogan Page.
- Burnett, W. and Cavaye, G.** 1980: Peer assessment by fifth year students of surgery. *Assessment in Higher Education* 5, 273-278.
- Caulk, N.** 1994: Comparing teacher and student responses to written work. *TESOL Quarterly* 28, 181-188.
- Cheng, W. and Warren, M.** 1997: Having second thoughts: Student perceptions before and after a peer assessment exercise. *Studies in Higher Education*, 22(2), 233-239.
- Conway, R., Kember, D., Sivan, A. and Wu, M.** 1993: Peer assessment of an individual's contribution to a group project. *Assessment and Evaluation in Higher Education* 18, 45-56.
- Devenny, R.** 1989: How ESL teachers and peers evaluate and respond to student writing. *RELC Journal* 20, 77-90.
- Duke, C. and Sanchez, R.** 1994: Giving students control over writing assessment. *English Journal* 83, 47-53.
- Earl, S.E.** 1986: Staff and peer assessment: Measuring an individual's contribution group performance. *Assessment and Evaluation in Higher Education* 11, 60-69.
- Entwhistle, N.J.** 1987: A model of the teaching process. In Richardson, J.T.E. and Eyesynck, M.W. and Piper, D.W., editors, *Student learning: research in education and cognitive psychology*. Guildford, Society for Research into Higher Education. Miltyn Keynes, Open University Press.
- Entwhistle, N.J.** 1993: Recent research on student learning and the learning environment. Paper presented at 'New Developments in Learning' Conference, Napier University, Edinburgh.
- Etheridge, C.** 1995: What's wrong is what's right: Setting realistic expectations for peer evaluation. *English in Texas* 27, 4-7.
- Falchikov, N.** 1993: Group process analysis: Self and peer assessment of working

- together in a group. *Educational & Training Technology International* 30, 275-284.
- Falchikov, N.** 1995a: Peer feedback marking: Development peer assessment. *Innovations in Education and Training International* 32, 175-187.
- Falchikov, N.** 1995b: Improving feedback to and from students. In Knight, P., editor, *Assessment for Learning in Higher Education: 1*. London: Kogan Page: 157-166.
- Forde, K.** 1996: The effects of gender and proficiency on oral self- and peer-assessments. *English Language Studies Working Papers, City University of Hong Kong* 1, 34-47.
- Freeman, M.** 1995: Peer assessment by groups of group work. *Assessment and Evaluation in Higher Education* 20, 289-300.
- Gibbs, G.** 1992: *Improving the Quality of Student Learning*. Bristol: Technical & Educational Services Ltd.
- Goldfinch, J.** 1994: Further developments in peer assessment of group projects. *Assessment and Evaluation in Higher Education* 19, 29-35.
- Goldfinch, J.** and **Raeside, R.** 1990: Development of a peer assessment technique for obtaining individual marks on a group project. *Assessment and Evaluation in Higher Education* 15, 210-231.
- Hogan, P.** 1984: Peer editing helps students improve written products. *Highway One* 7, 51-54.
- Jacobs, G.** 1987: First experiences with peer feedback on compositions: Student and teacher reaction. *System* 15, 325-333.
- Jacobs, G.** 1989: Miscorrection in peer feedback in writing class. *RELC Journal* 20, 68-75.
- Jafarpur, A.** 1991: Can naive EFL learners estimate their own proficiency? *Evaluation and Research in Education* 5, 145-157.
- Jones, N.** 1995: Business writing, Chinese students and communicative language teaching. *TESOL Journal* 4, 12-15.
- Kwan, K.P.** and **Leung, R.** 1996: Tutor versus peer group assessment of student performance in a simulation exercise. *Assessment & Evaluation in Higher Education* 21, 205-214.

- Lynch, T.** 1988: Peer evaluation in practice. In Brookes, A. and Grundy, P., editors, *Individualization and Autonomy in Language Learning ELT Documents: 131*. Modern English Publications: 119-125.
- Mangelsdorf, K.** 1992: Peer reviews in the ESL composition classroom: What do students think? *ELT Journal* 46/3, 158-172.
- Mendonca, C. and Johnson, K.** 1994: Peer review negotiations: Revision activities in ESL writing instruction. *TESOL Quarterly* 28, 745-769.
- Miller, L. and Ng, R.** 1994: Peer assessment of oral language proficiency. *Perspectives: Working Papers of the Department of English, City Polytechnic of Hong Kong* 6, 41-56.
- Mitchell, V.W. and Bakewell, C.** 1995. Learning without doing: enhancing oral presentation skills through peer-review. *Management Learning* 26, 353-366.
- Mondock, S.** 1997: Portfolios: The story behind the story. *English Journal*, January 1997, 59-64.
- Morris, P.** 1996: *The Hong Kong School Curriculum*. Hong Kong: Hong Kong University Press.
- Murau, A.** 1993: Shared writing: Students' perceptions and attitudes of peer review. *Working Papers in Educational Linguistics* 9, 71-79.
- Newkirk, T.** 1984: Direction and misdirection in peer response. *College Composition and Communication* 35, 301-311.
- Oldfield, K. and MacAlpine, M.** 1995: Peer and self-assessment at tertiary level - an experiential report. *Assessment & Evaluation in Higher Education* 20, 125-132.
- Patri, M.** 2002: The influence of peer feedback on self-and peer-assessment of oral skills. *Language Testing* 19, 109-131.
- Rainey, K.** 1990: Teaching technical writing to non-native speakers. *Technical Writing Teacher* 17, 131-135.
- Rothschild, D. and Klingenberg, F.** 1990: Self and peer evaluation of writing in the interactive ESL classroom: An exploratory study. *TESL Canada Journal* 8, 52-65.
- Topping, K.** 1998: Peer assessment between students in colleges and universities. *Review of Educational Research* 68/3, 249-276.

- Tudor, I.** 1996: *Learner-Centredness as Language Education*. Cambridge: Cambridge University Press.
- Watson, H.M.** 1989: Report on the first year of research into developing an evaluative technique for assessing seminar work. *Collected Original Resources in Education (CORE)* 13/2, Fiche 12 Cl.
- Webb, N.M.** 1995: Group collaboration in assessment: Multiple objectives, processes, and outcomes. *Educational Evaluation and Policy Analysis* 17, 239-261.
- Williams, E.** 1992: Student attitudes towards approaches to learning and assessment. *Assessment & Evaluation in Higher Education* 17, 45-58.

Appendix 1a Seminar assessment criteria (class)

Assessing your peers is not an easy task. You need to try to be fair and objective. Your assessment scores will *only* be seen by your teacher.

Use the following scale when assessing your fellow students.

1	2	3	4	5
poor	below average	average	above average	excellent

A. Preparation, overall presentation and content of seminar paper

- evidence of rehearsal
- consideration of audience
- relevance and interest of material
- quality and appropriacy of visual aids
- well-structured
- clear conclusion(s)

B. Delivery of seminar paper

- rapport with and sensitivity to audience
- body language

timing and pacing
 sensitivity to audience feedback
 use of visual aids
 clarity of delivery
 confidence
 clarity and coordination of group delivery
 satisfactory answers when required
 encouragement of discussion

C. Participating in seminar discussion

relevant questions, comments, ideas
 appropriate academic language

D. Language

accuracy and appropriate use of

- vocabulary
- structures
- register

 conciseness
 clarity of expression

Appendix 1b Oral presentation assessment criteria (class)

Assessing your peers is not an easy task. You need to try to be fair and objective. Your assessment scores will *only* be seen by your teacher.

Use the following scale when assessing your fellow students.

1	2	3	4	5
poor	below average	average	above average	excellent

A Quality of preparation and content

consideration of audience
research of topic
relevance and interest value of topic
structuring of presentation
quality of content
quality and appropriacy of visual aids
evidence of rehearsal

B Quality of delivery

rapport with and sensitivity to audience
use of visual aids
use of eye contact, voice, speed of delivery, gestures, movement
timing and pacing
confidence
handling of questions
clarity and coordination of group delivery

C Language

accuracy and appropriate use of :
- vocabulary
- structures
- register
conciseness
clarity of expression

Appendix 1c Report assessment criteria (class)

Assessing your peers is not an easy task. You need to try to be fair and objective. Your assessment scores will *only* be seen by your teacher.

Use the following scale when assessing your fellow students.

1	2	3	4	5
poor	below average	average	above average	excellent

A Preparation and content

consideration of readers

good choice of topic

clear objective/purpose statement

appropriate methods/procedures for collecting information/data

selection of information relevant to topic and purpose

organisation of parts of report relevant to objective and reader interest and needs

B Organisation

good connection of ideas

appropriate use of a variety of cohesive devices

C Language

accurate and appropriate use of vocabulary, structures, and register

concise and clear expression of ideas

D Writing style

objective, concrete, and organised facts

appropriate use of reference conventions

E Layout and presentation

attractive

appropriate use of paragraphing, headings, numbering, spacing, illustrations, etc.

Appendix 2 Form for assessing group members' contributions to group work

Use the following scale when assessing your fellow students' level of contribution to the group project.

0	1	2	3	4	5
Did not contribute in this way	Poor	Below average	Average	Above average	Excellent
Assessment criteria	Ideas and suggestions for group project	Literature search	Literature analysis	Preparation and planning of seminar presentation	Preparation and planning of oral presentation

Group member _____

Appendix 3a BEng (Hons.) Electrical Engineering Year 1: Peer assessment: group project (pre-questionnaire)

Name: _____ Class: _____

How do you feel about peer assessment? Answer the following questions by circling your answer to each question.

1. Do you think students should take part in assessing their peers?
A. Yes. B. No. C. Not sure.
2. Do you believe a First Year student should be able to assign grades to peers in a responsible manner?
A. Yes. B. No. C. Not sure.
3. Do you think you will feel comfortable in making peer assessments?
A. Yes. B. No. C. Not sure.
4. Do you think you will make a fair and responsible assessment of your peers?
A. Yes. B. No. C. Not sure.

Thank you.

Appendix 3b BEng (Hons.) Electrical Engineering Year 1: peer assessment: group project (post-questionnaire)

Name: _____ Class: _____

You have assessed your peers' performance in seminars, reports, and oral presentations. You have also assessed the level of contribution of each of your group members while doing the project.

What are your feelings about peer assessment when you think back on it? Answer the following questions by circling your answer to each question.

1. Do you think students should take part in assessing their peers?
A. Yes. B. No. C. Not sure.
2. Do you believe a First Year student should be able to assign grades to peers in a responsible manner?
A. Yes. B. No. C. Not sure.
3. Did you feel comfortable when you made peer assessments?
A. Yes. B. No. C. Not sure.
4. Do you think you have made a fair and responsible assessment of your peers?
A. Yes. B. No. C. Not sure.

Thank you.

Appendix 4 Student feedback on peer assessment

Name: _____ Class: _____

Seminar

[Note: The questionnaires for oral presentation and report were the same, except for the different assessment criteria]

In semester one, you assessed your peers' performance in seminars, reports, and oral presentations. You have also assessed the level of contribution of each of your group members while doing the project. Then you answered a questionnaire which asked you about your feelings about peer assessment.

This questionnaire aims to ask you for more specific information about how you felt when you assessed your peers. Please indicate, by ticking the appropriate box, whether you Strongly Agree (SA), Agree (A), are Neutral (N), Disagree (D), or Strongly Disagree (SD) to each of the statements below.

	I felt comfortable in assessing the performance of my peers on each of the assessment criteria.					I think I assessed my peers fairly and responsibly on each of the assessment criteria.				
	SA	A	N	D	SD	SA	A	N	D	SD
A. Preparation, overall presentation and content of seminar paper	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
evidence of rehearsal										
consideration of audience										
relevance and interest of material										
quality and appropriacy of visual aids										
well structured										
clear conclusion(s)										
B. Delivery of seminar paper	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
rapport with and sensitivity to audience										
body language										
timing and pacing										
sensitivity to audience feedback										
use of visual aids										
clarity of delivery										
confidence										
clarity and coordination of group delivery										
satisfactory answers when required										
encouragement of discussion										
C. Participation in seminar discussion	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
relevant questions, comments, ideas										
appropriate academic language										