

Reflections on Peer Evaluation in an English Language Course

David Evans

National College of Nursing, Japan : 1-2-1 Umezono, Kiyose-shi, Tokyo, 〒 204-8575, Japan
evans@adm.ncn.ac.jp

[Abstract] This paper reports the experiences of both a 4th year elective English discussion class and its teacher on peer and self evaluation. The students were asked to award two grades to each student, one for their ability an effort during course-work and the other for a presentation of the student's graduate thesis. There views on the experience of peer evaluation were canvassed through an optional questionnaire. A brief review of the relevant literature is discussed, and the results of the peer evaluation are presented in terms of the student rankings. These show that there are great similarities between the teacher evaluation and the student evaluation, and that the teachers can become aware of their own idiosyncrasies. The students, on the whole, found peer evaluation to be beneficial and fair, and made suggestions as to how the process could be improved for future courses. The teacher's recommendations for the course are also included and these include classroom management and clearer instruction on choosing criteria.

[Keywords] peer evaluation, self evaluation, student feedback

Introduction

Peer and self evaluation have always existed informally as students constantly compare their own performance with those of their classmates (Underhill, 1987; Race, Brown & Smith, 2005). Peer assessment naturally helps self-assessment, as by becoming aware of others' performance students will reflect on their own. Peer and self assessment are skills that should benefit the student throughout their studies and professional life (Brown, Rust & Gibbs, 1994).

There are many different ways in which both are used by educators, sometimes as isolated activities within a course but also for final assessment. For self evaluation it is common for students to rate what they feel are their strengths and weaknesses, whereas in peer evaluation, students may provide feedback to fellow students. Some academics caution against using student assessments for final assessment, as self-evaluation is clearly open to abuse as students could award themselves undeservedly high marks; similarly there are doubts about the reliability of peer evaluation and consequently the degree to which it should be used in assessment.

Luoma (2004) believes that peer evaluation can be used as a supplement to teacher evaluation but should not replace it, whereas Race, et al. (2005) argues that it should be included in the final assessment to make the process meaningful. Walker

(2001) in a review of previous research found two studies had deemed peer evaluation as too unreliable to have a significant role in final assessment, but also cautions that another study also found it no less reliable than tutor marks. Cohen (1994) cites a study by Rolfe that found that in language learning the correlation between tutor and peer evaluation was consistently high, and that they were as reliable as the tutors, if not more so, as the students spent more time interacting with each other than the teacher did.

In high stakes testing it is easy to see why there is a reluctance to base life changing decisions on 'non-experts' (even though there are doubts about the reliability of multiple choice exams (Hughes, 1989)), but in low stakes testing, such as an optional English course in a college, with highly motivated students, this risk is less severe.

Additionally, in testing spoken ability in English, there is inevitably going to be an element of subjectivity, no matter how rigorous the training of testers, be they professionals or students as speech can not be easily broken down into a series of discrete items. However, as this author has argued previously (Evans, 2003a; Evans, 2003b; Evans & Herbert, 2003), a test of spoken ability does need to be done directly, as the fundamentals of speaking can not be successfully measured in a grammar or reading exam. Students who score highly in tests such as TOEIC do not necessarily do well in oral tests and vice versa.

The advantages and disadvantages of peer evaluation

Benefits

1. Developing students' critical faculties

By becoming more aware of the strengths and weaknesses of classmates, students will become more reflective of their own (English, Brookes, Avery, Blazeby & Ben-Shlomo, 2006) which should help them to become better speakers of English. Instead of simply thinking that another student is better, students should be asking themselves why they are better and then endeavour to emulate them.

2. Autonomous learning

Autonomous learning is a potential benefit as described by many experts including Brown, et al. (1994). As students develop their critical faculties they should become less reliant on teachers for guidance and more able to know direct their own learning. There is a limit to how much a teacher can help improve a student's English, and furthermore the teacher can not be with the students outside of the classroom.

3. Developing leadership skills

By developing the ability to appraise one's own and others' performance, students should be able to transfer this skill into other realms, such as work. One of the aims of this college is to produce nursing leaders, and by becoming aware of individual's strengths and weaknesses, this should help students when they are in senior positions, as Topping (1998) reports in his review of peer evaluation. By doing peer evaluation students ask classmates informally how they can improve and also give and receive advice. This experience should enhance their leadership abilities.

4. Motivation

In teacher evaluation it is only necessary for the student to be actively engaged when the teacher is present. Motivation increases (Abson, 1994 cited in Topping, 1998) when students are continually being assessed. Peer evaluation also means that students need to listen to classmates which keeps them involved at all times.

5. Multiple assessors

The danger with having one assessor only, even if it is the teacher, is that should the teacher make a mistake, there is no chance for it to be rectified. With a class of assessors, assuming that they are capable of performing such a role, the assessment should be fairer.

6. Continual assessment

If a course finishes only with a final exam it is possible for students who have performed well over the course to have 'a bad day' and consequently receive a poor grade, or conversely, for a student who has made little effort during the course to do well

solely in the final test. By doing peer evaluation for class work, such problems are avoided.

7. Fairness

Although the teacher should be best able to evaluate students overall proficiency in English, as they possess better English skills, coursework should be about achievement during that course and that course only. If proficiency is measured in a college course, a returnee student (one who has been schooled abroad, and then returned home) should be able to get the highest grades in a final exam, regardless of whether or not the student has attended the class. By evaluating achievement, all students will know what has been done in the class, and are therefore more able to grade achievement than proficiency.

Teachers also suffer from what is called the 'halo effect', a term used by a psychologist called Thorndike in the 1920's, to explain the phenomenon of generalising from an initial impression. If a first impression of an individual is good, other actions are seen positively. In English language teaching, this means that once a teacher forms a positive opinion of a student they continue to hold it. The opposite is also true. Thorndike calls this the 'devil effect', and although this term is not used in the English teaching community, students do have difficulties in overturning an initial negative impression. Using peer evaluation prevents such a problem.

A second way, in which peer evaluation is fairer, is that just as the 'white coat effect' is acknowledged in medicine to explain atypical results when a doctor is present, a teacher can have an inhibiting influence on some students who become tongue-tied and who under-perform when a teacher is monitoring.

8. Classroom management

Because of the need for students to assess all their classmates, this means that the seating positions change on a weekly basis. For well motivated students, this is of little consequence, but for less eager students, it is a very useful tool for breaking up small groups that do not otherwise participate in classes. A second advantage of changing seating positions is that students get to hear different accents and different pronunciations. A final advantage is that changing seating positions gives students the opportunity to develop relationships with classmates that they do not know so well.

9. Attendance and Punctuality

These are always vexing issues for a teacher, firstly, because the ideal is for a student to attend every class punctually, but also because for some students, perfect or near perfect attendance is seen to compensate for a lack of ability. With peer evaluation, this aspect can be included in class work assessment by a jury of peers.

10. Ownership

The evaluation process is shrouded in mystery for many students, and to be included in the process not only gives them an understanding of what it entails, but also it values their contribution.

Concerns

1. Maturity

Assessing is a difficult skill, and it is possible that students might show favouritism to friends.

2. Ability to assess

Assessing someone who has less ability than oneself is much easier than assessing who has more ability. To differentiate between better speakers is extremely difficult for someone with a lower ability level.

3. Student reaction

Students might feel it was an abrogation of teacher responsibility to involve students in the assessment process. Any test needs 'face validity', which means the students must believe it is a 'proper' test. If students do not have confidence in the style of the test, or its graders, it will not be successful. Rollinson (2005) points out that students might feel only a teacher or a better speaker is capable of grading accurately.

A second potential problem is that students might feel the responsibility too great.

4. Stereotyping

In the same way that teachers can suffer from the 'halo effect' so can students. It is important that a student is evaluated only on their performance in the course being evaluated and not on how they performed in previous courses. If students are unsure about their ability to grade fairly they might rely on past performance.

Methods

The students

There were 15 students; 14 female and one male. All the students were 4th years in a nursing college.

Ethical considerations

Approval was sought and granted by the college's ethics committee. Students were asked in writing if they would allow their data relating to both the scores and to the questionnaire to be used for research purposes. There was no pressure for students to agree to this. All students were content for their grades and questionnaire comments to be used anonymously.

The teachers

The two teachers for this course are the author, a native English speaker, and a Japanese English teacher. The author is the main teacher for the course and both teachers monitor the students.

The course

The course is an elective subject for 4th year students. The course is primarily designed to be a discussion course, and the main course book is 'Impact Values' by Day, Yamanaoka & Shaules (2003). The book is divided into 5 sections and has topics on values for people, relationships, the work place, the family and in society. In pre-course feedback, each student selects two topics from each section that are of most interest, then, all students choices are totalled, and the most popular are chosen.

Each week students either listen to the individual topics or read them. Fixed expressions and phrases, or alternative expressions are selected from each unit and presented to the students in handouts. Any problematic language is discussed and useful patterns highlighted, and students are encouraged, but not obliged, to use them. Students then divide into three groups of five. They are also given a list of potential questions to discuss, but are encouraged to lead the conversation in which ever way they like. The students typically discuss the subject for about 45 minutes. One student in each group makes notes and reports back to the class for 5 minutes and, if time permits, students from other groups ask questions. By the end of the course all students had reported back their findings to the class at least once.

The second major element to the course is presentation skills in English. The students present their graduation thesis to the class. As all the students are training to be nurses the topics are on nursing, and as such, are of relevance to their classmates. They are given handouts about what constitutes a good presentation, and also on how to construct a presentation with appropriate language and signposting. Students are not allowed to read their presentation, but they can have a list of bullet points to refer to when speaking. Memorization is discouraged.

In the second term, two or three students presented their work each week which not only gave the students presenting the opportunity to improve for the final exam, but also gave their peers the opportunity to improve as graders (Norcini, 2003). After their presentation, other students asked questions about their graduation topic. English was spoken at all times. They received individual written feedback by email the day after the class, including both strengths and weakness in their

performance. In addition, points that were regarded as being beneficial to all students in improving their performance were given by both the Japanese teacher of English, (focusing on lexical and grammatical areas), and from the native English speaker (on organisational and presentational issues) both orally and in handout form.

The presentations were expected to be between 10 and 15 minutes. They were asked not to rely on visual aids, as this could have been a distraction from their speaking skills.

The final strand to the course was the development of the students' nursing and technical knowledge. Students would prepare a worksheet before the class and the answers would be discussed at the beginning. This lasted for about 20 minutes each week until the presentation practice began.

Preparation of students for peer evaluation

In the first class the students were told that they would be doing peer evaluation, for both class work and for a presentation, and they were provided with information about important aspects of peer evaluation at regular intervals as the course progressed.

The information given to the students included a literature review which highlighted the benefit and problems with peer evaluation. Other discussions and handouts focused upon choosing assessment criteria and the potential difficulties in grading.

Evaluation day

The evaluations took place in the last class. The order for the presentations was decided by drawing lots. Each student had a form in which to keep notes on the performance of each speaker which were collected in afterwards.

The students were told that an A grade was equivalent to 80% and above, a B from 70-79%, and a C (should it be necessary) was below 70%. They were asked to record their evaluation in percentage terms and asked not to use the same percentage twice. Students were asked to self evaluate as well.

Class work and presentation were to account for 50% each of the final percentage score, so to calculate the students' final scores the averaged mark for each was added and divided by two. Both teachers also graded the students so that any irregularities could be double-checked.

The questionnaire

Once the students had completed their presentations and their peer evaluations, they were asked to complete a simple questionnaire, designed to record their feelings towards the

evaluation process. The questions that the students were asked were;

1. *Do you feel that peer evaluation was worth doing? Why? Why not?*
2. *Which was easier to evaluate, class work or presentations?*
3. *What did you find the most difficult thing in assessing the presentations?*
4. *What did you find the most difficult thing in assessing class work?*
5. *Would you have liked to have had stricter criteria for the evaluations?*
6. *Do you think peer evaluation should be continued next year?*

Results and Discussion

Class work

The results are presented according to rank, rather than the raw scores, as the marking scales employed varied between the students. Nearly all students graded within a 30% range, from 70-100%, but in one case grades of 30% were given, so this was excluded, as the influence on the overall scores would have been too great. This shows how important it is for students to use similar ranges when assessing.

The columns indicate the rankings awarded by the students (represented by capital letters) and the rows represent the ranking they received from fellow classmates. T1 in the 3rd column from the right represents the ranks as awarded by the Japanese teacher of English and T2 represents those of the native, British teacher. The 'All' column is the combined ranking awarded by both students and teachers.

The rankings given by the two teachers are remarkably similar except in the case of student N (underlined). This anomaly might be explained by the halo effect from a very impressive practice presentation the week before.

Although the rankings for the other 13 students by the teachers are remarkably similar, this case shows that evaluation is necessarily subjective, no matter who is evaluating, student or teacher. This highlights an advantage of peer evaluation as by using the rankings given by peers, that 'mistake' is moderated by the other graders. Conversely, a disadvantage of peer assessment is that because there are multiple assessors the higher scores, (in percentages) are lower than they would be with only one assessor, as are the lowest scores. As a consequence of both this, and the fact that the students generally graded lower than their teachers, it was felt necessary to delete each individual's three lowest scores in order to boost the final percentage mark.

Table 1 Ranking for class work

	A	B	C	D	E	F	G	H	I	*J	K	*L	M	N	T1	T2	All
A	6	1	1	1	1	1	2	3	1	5	1	9	1	2	1	1	1
B	3	4	2	2	4	7	3	6	6	1	5	4	2	4	2	3	2
C	5	3	5	4	2	2	7	2	1	8	2	6	4	3	4	7	2
D	8	8	11	11	3	3	1	5	5	4	7	2	7	5	9	6	4
E	4	7	6	10	9	6	6	7	4	2	4	11	3	1	2	2	5
F	7	2	4	3	7	13	5	1	3	11	14	3	11	7	7	4	6
G	2	5	8	7	6	4	8	8	9	7	6	1	8	10	10	5	7
H	1	6	13	9	10	9	4	11	10	3	3	8	5	8	8	8	8
I	9	12	7	13	5	5	12	14	14	10	11	7	10	6	11	10	9
J	11	9	10	5	11	10	9	9	8	14	10	12	9	14	14	9	11
K	10	13	9	15	13	15	10	12	13	9	15	13	13	15	12	13	12
L	15	11	15	12	12	14	13	13	12	13	13	10	12	12	13	13	13
M	14	14	14	14	14	12	11	15	11	12	9	14	14	13	15	15	14
<u>N</u>	12	10	12	8	8	11	14	10	7	6	8	5	15	11	4	12	10

Table 2 Ranking for presentation

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	T1	T2	All
A	10	7	5	6	8	6	2	5	6	8	9	4	4	3	3	3	6
B	4	5	2	2	3	3	3	1	1	1	1	5	3	2	1	2	1
C	8	2	7	3	6	2	5	2	3	5	5	3	2	5	8	5	4
D	2	1	3	9	2	7	1	6	5	4	3	1	5	4	5	1	3
E	1	3	6	1	10	1	7	4	2	2	2	2	1	1	2	4	2
F	7	11	10	10	11	10	8	10	7	11	12	7	8	9	9	7	10
G	9	6	4	7	5	4	9	8	9	9	7	6	6	6	7	6	8
<u>H</u>	5	4	1	4	4	5	6	7	4	3	6	8	9	7	4	11	5
I	11	10	12	14	7	11	14	11	14	14	11	13	13	11	10	9	13
J	12	12	14	13	12	8	13	9	12	10	8	12	10	12	14	14	11
K	6	9	13	12	9	12	11	13	10	7	14	9	12	14	13	8	9
L	14	13	9	8	14	13	10	12	11	12	10	14	11	10	11	12	12
M	13	14	11	11	13	14	12	14	13	13	13	11	14	13	12	13	14
N	3	8	8	5	1	9	4	3	8	6	4	10	7	8	6	10	7

The two columns with asterisks (J and L) show the rankings most divergent from the norm for the top ranked student. Both of these students were among the weaker students in the class, and as mentioned earlier, this suggests there is a problem in evaluating upwards. By contrast the students are fairly unanimous in identifying the weaker students.

In each column the figure in bold is the self-evaluation rank. What can immediately be seen from this is that as a means of assessment, self-assessment was not worthwhile. The percentage scores for self-evaluation were excluded from the final evaluations for this reason.

Presentations

When comparing the two teachers scores, the vast majority of students are ranked consistently, however, there is one obvious exception, student H (underlined), in which this author's ranking is significantly wayward compared both to the Japanese teacher and the students. Student H is an excellent speaker of English, but the author felt the student underperformed, and, it would seem, judged the student more critically than other students. Being able to make a comparison with other graders allows a teacher to understand their own grading idiosyncrasies. Similarly, in other cases in which the author is slightly awry from the other graders in 'over-grading' it becomes apparent too

great an emphasis has been placed on effort rather than achievement.

Table 2 shows more consistency in grading than was the case for class work, because for the presentation all graders witnessed the same event. Even in grading 'upwards' there seems to be more agreement as to which were the best presentations. For the weaker performances, the uniformity in ranking is particularly noticeable.

The self-evaluations for the presentations also show that students have under graded. However, in a recent peer evaluation with 2nd year students, in which they perform pair work for peer evaluations, it is not uncommon for students to rank themselves and their partner as the best in the class. This suggests that pair and group evaluation might promote a more competitive attitude or that 2nd year students may not yet be sufficiently mature to undertake peer evaluation. This finding lends weight to Lurie, Nofziger, Meldrum, Mooney & Epstein (2006, p. 845) who found in peer evaluation among medical students that 'different classes may tend to have different 'personalities''.

Another pleasing result of the peer evaluation is that the students have, as a class, made a clear distinction between class work and the presentation. There has been no assumption that a student who excels in class work must be good at presentation, or vice versa. Furthermore, some of the more proficient speakers are not at the top of the rankings, but instead, those who worked hardest and who took a more active part in class have fared better, and there is no sign of any favouritism or stereotyping in the presentation rankings. This is very reassuring.

Student reaction to peer evaluation

The students' answers to a brief questionnaire are discussed in this section.

1. Do you feel peer evaluation was worth doing? Why? Why not?

Yes	No	Not Sure
11	1	1

For those who felt it was worthwhile the following are representative of the comments.

I could compare English skills between my friends and me. I could find many good points of my classmates that I felt I should imitate.

By evaluating others it became clear what my weak points were.

In keeping with the literature this shows that students reflect on the performance of classmates and then incorporate those strengths into their own performance.

Classmates have a similar feeling and way of thinking as each

other.

We know our classmates very well (As a result of the time spent working together)

Being evaluated by other students meant I knew the points that other students will evaluate carefully.

These three comments suggest that students feel positive about being evaluated by peers, as they are more aware of what classmates regard as good practice than they are with teachers. This could mean that teachers need to communicate their aims more clearly to the students, but it could also be that fellow students are more realistic in what their classmates can achieve.

Yes, because when we did class work, teachers could not see everything.

This shows that students are aware of the limitations of solely being evaluated by a teacher. A teacher can not see everything, but a fellow classmate will and in this respect peer evaluations are fairer than only teacher evaluations.

For those who chose 'Not sure', one student did not comment and the other wrote

For peer evaluation, I could evaluate classmates, but self-evaluation was difficult.

This suggests that students need more practice in self-evaluation, or it might be cultural, in that it is hard for Japanese students to appear overly confident.

2. Which was easier to evaluate, class work or presentations?

Class work	Presentations	About the same
1	8	4

One probable reason for the preference for presentations is that everyone sees the same performance. Also class work evaluation is over time and performance is not consistent, therefore grading is harder. It is also possible that the students were not given sufficient preparation to do class work evaluations as confidently as the presentations.

3. What did you find the most difficult thing in assessing the presentations?

The teacher asked us not to give the same score twice and sometimes I could not find a difference between the presentations.

This is a very valid and important point. In trials with peer evaluation, students had often give the same grade countless times, rendering the task meaningless. To avoid this problem, students were asked not to award the same mark twice. It was hoped this would make students concentrate on fine differences between two classmates, but in truth, it is sometimes too difficult to make a distinction between two similar students. This rule was too artificial and it will be changed in the future.

Sometimes the topics were too different. It wasn't a problem of English, but some topic areas I found difficult to understand.

When I couldn't understand the research.

Both of these comments are also very insightful. By choosing the students' graduate thesis as the topic of the presentation there were both advantages and disadvantages. The key advantage was that at that stage of term, the thesis is the most important work in the students' academic lives. As Douglas (2000) points it is essential that test takers must be engaged by the task in which they are tested. Furthermore, the students are required to present the thesis in a presentation (albeit in Japanese), so the task of presenting was also relevant.

The disadvantage is that the content matter was different for each student. A student, whose graduate research was on a mainstream topic, familiar to all students, was at an advantage compared to those which were more specialised. This problem should not have mattered greatly, as students were constantly reminded about the importance of the audience, and the need to make the presentation accessible to all. Similarly, it was made explicit to the students that they were not meant to be assessing the content of the presentation, but how it was presented and the English used. However, it is still the case that it is easier to listen to a familiar topic rather than an unfamiliar one.

Sometimes I wasn't sure if the student used good English or not. Assessing the English level.

I couldn't make judgements about grammar.

These three comments highlight the main weakness of peer evaluation in an English language course. It is very difficult for students to assess correct usage of language. If a classmate uses a phrase with which they are unfamiliar, it is not possible for them to know if it is a correct expression or an incorrect one. In a previous trial with presentation evaluations, students did award high marks to a student who spoke fluently but whose syntax was flawed, whereas both teachers graded the student significantly lower. Peer evaluation might be more reliable if the scores of only the better students were used. However, in low stakes testing, this is not so critical, and it would seem unfair to exclude weaker students from the process.

4. *What did you find most difficult in assessing the class work?*

In each class we did different things so it was difficult for me to compare and evaluate the difference.

This is a similar point to that about difference in content in the presentations, but students should have been assessing speaking skills rather than the ideas expressed, so more guidance on how to evaluate seems to be necessary.

With one student we were only together for one class, so that made evaluation difficult.

There were some students I hadn't had much contact with so I found this grading difficult.

There was one student I wasn't in the same group as.

I didn't talk to all of my classmates.

This problem should not have arisen as students were regularly reminded to work with different students each week, but it shows that students' seating positions need to be regulated by the teacher. There was, in fact, sufficient opportunity for each student to have worked with all other students at least twice. However, this does highlight the difficulty of doing peer evaluation with larger class sizes.

I found it difficult to remember previous classes.

I forgot how well certain students did in detail.

This could be countered by students using records as the course progresses, but although peer evaluation proved to be a positive exercise, no course should be dominated by evaluation. Additionally, for less mature students keeping records could become a cause of anguish especially if low grades are given and talked about. Forms could be returned to the teacher at the end of the class, but even this put too strong an emphasis on grading. Remembering how well students have done from week to week is difficult even for experienced teachers so it is not so surprising that students find it difficult. Reminding students orally each week might help to keep this information in their minds.

There are some students who talk a lot and some who talk a little. I felt uncomfortable about it.

This is an interesting point and one all teachers wrestle with. Some people might have good English but just be naturally quieter personalities than others. Spoken tests do favour more extrovert characters (Cohen, 1994). Students need to be aware of how important it is to offer more in such an exam or course. Many students say little because of a concern of speaking incorrectly and being penalised, so it would be very difficult for a teacher to discern the correct cause of a student's lack of participation.

Class work doesn't lend itself to ranking; I think grades would have been better.

This is an understandable point, but it would be very tempting for students to give, for example, eighty percent of the students 'A' grades and the remainder 'B' grades, without too much thought. By requiring students to rank students (which is what in effect giving percentages without repeating the same percentage does) they are challenged to use their critical faculties more, which should be more beneficial to each student.

5. *Would you have liked to have strict criteria for evaluation?*

Yes	No
8	5

Originally the plan was for all students to use the same criteria

for grading. However, it became apparent from watching the practice presentations that there were so many facets to a good presentation that it would be impossible to distil these into 4 or 5 categories that all students would feel competent in judging. Grammatical competence or use of high level structures and phrases would be difficult for some students to grade. Similarly, judging fluency is difficult. Hesitation, should an idea spontaneously occur, is perfectly acceptable, but students tend to regard any pause as an error. In a previous experiment with presentation evaluation one student was penalised by classmates for this very manoeuvre, whereas both teachers felt it to be commendable.

As a result of these difficulties students were asked to grade the presentations holistically and to choose their own criteria. The written comments showed the same criteria were applied for each student. The advantage of allowing students to choose the criteria is that they will select those that they feel able to judge.

Class work is the area in which more help is needed for students when evaluating. Students need to have the elements that make up conversation explained to them (as this is what is primarily what takes place) and then they can choose those which they feel able to judge.

6. *Do you think peer evaluation should be continued next year?*

**Yes, because you can see the effort your classmates make.*

**Yes, because I could understand all my classmates skill.*

**Yes, it's worth doing but it should be begun in the 1st year.*

Yes, it was a good experience for me.

Yes, because we could learn not only English but also how to evaluate.

I think it should be continued because I had to use my brain for it.

Yes I do, because it's good to be evaluated by all students.

Yes, I think peer evaluation makes all students listen more carefully.

Yes, being evaluated by others is useful.

I'm not sure but I could listen carefully to each classmate's presentation.

Yes, but I think set criteria should be given.

These comments show that all but one student found the peer evaluation to be beneficial to them. The first two comments show that peer evaluation can make students more appreciative of the efforts of others and this is an unforeseen benefit.

The third comment is very interesting and perceptive. Grading becomes easier and more efficient for a teacher with more experience, and this will be even more important for novices. Dudley-Evans & St. John (1998) point out that all examiners

need meetings and training sessions to reach consistent standards and the same will be true for students. If students begin peer evaluation earlier in their college life they would be considerably better at it by the 4th year. However, issues of maturity, class size and the fact that classes are compulsory need to be considered carefully.

Self evaluation

As mentioned in the results, self-evaluation as expressed by a percentage mark was not a meaningful exercise and this mark was not included in the final percentage. One student wrote;

How about writing our own good points and weak points after our practice presentations?

This is an excellent suggestion and will be incorporated in the next course. Self-evaluation in a non-grading manner would seem to be a potentially more beneficial experience.

Conclusion

The impression of peer evaluation for both teacher and students was favourable.

The positive discoveries for the teacher were;

- 'mistakes' in grading by the teacher are moderated by the class;
- it is possible for teachers to become aware of their own evaluation style;
- even less able students can determine who the most able students are;
- students were able to distinguish clearly between class-work and presentation evaluation.

The benefits according to students are;

- they became aware of what they needed to do to improve;
- fellow student's criteria for grading are easier to comprehend than the teachers;
- teachers can not evaluate everything as they aren't always listening.

The lessons that have been learnt are;

- all students need to grade within a similar range;
- peer evaluation moderates both the highest and lowest scores which disadvantages the better students;
- using self evaluation for grades is not useful but for formative feedback it could be;
- class-work evaluation will provide more erratic grades due to its nature;
- presentation evaluation is preferred by students, so greater help is required with class work evaluation;
- weak students do not feel capable of evaluating grammar;
- it is important for the teacher to rotate the seating positions.

Peer evaluation would seem to be promising as a viable alternative or as additional form of assessment for Japanese students. Both the teacher and the students were pleased with the results. It should be noted that this paper is concerned with one class, and the experience might be different with other students.

■ References

- Brown, S. (1996). Assessment. <http://www.city.londonmet.ac.uk/deliberations/assessment/invite.html>
- Brown, S., Rust, C., & Gibbs, G. (1994). <http://www.city.londonmet.ac.uk/deliberations/ocsd-pubs/div-ass5.html>
- Cohen, A. (1994). *Assessing language in the classroom*. Boston: Heinle & Heinle.
- Day, R.R., Yamanaka, J., & Shaules, J. (2003). *Impact values*. Hong Kong: Pearson.
- Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge: Cambridge University Press.
- Dudley-Evans, T., & St. John, M.J. (1998). *Developments in English for specific purposes: A multidisciplinary approach*. Cambridge: Cambridge University Press.
- English, R., Brookes, S. T., Avery, K., Blazeby, J. M., & Ben-Shlomo, Y. (2006). The effectiveness and reliability of peer-marking in first-year medical students. *Medical Education*, 40, 965-972.
- Evans, D. (2003a). Introducing oral testing for English conversation classes: The reasons and the reaction. *The Journal of Nursing Studies, National College of Nursing, Japan*, 2(1), 49-58.
- Evans, D. (2003b). Why do oral testing? *ETJ Journal*, 4(2), 1-3.
- Evans, D., & Herbert, J. (2003). The Importance of Oral Testing for Conversation Classes: A Framework for Planning Oral Tests. *The Journal of Nursing Studies, National College of Nursing, Japan*, 2(1), 40-48.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Luoma, S. (2004). *Assessing speaking skills*. Cambridge: Cambridge University Press.
- Lurie, S. J., Nofziger, A. C., Meldrum, S., Mooney, C., & Epstein, R. M. (2006). Temporal and group-related trends in peer assessment amongst medical students. *Medical Education*, 40, 840-847.
- Norcini, J. J. (2003). Peer assessment of competence. *Medical Education*, 37, 539-543.
- Race, P., Brown, S., & Smith, B. (2005). *(500 Tips) on Assessment (2nd edition)*. Oxon: Routledge Falmer.
- Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT Journal*, 59(1), 23-30.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249-276.
- Underhill, N. (1987). *Testing spoken language*. Cambridge: Cambridge University Press.
- Walker, A. (2001). British psychology students' perceptions of group-work and peer assessment. *Psychology Learning and Teaching*, 1(1), 28-36.