

# Peer and self evaluation in spoken English: The views of first and second years in a nursing college

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 views of first and second year students at a nursing college on their experiences with peer and self evaluation. The students were asked to award two grades to each student, one for a role play on nursing English with a partner chosen of their choosing, and one for an unprepared dialogue on everyday English with a partner drawn by lots. Their views on the experience of peer and self-evaluation were canvassed through an optional questionnaire. A brief review of previous testing styles and the relevant literature are discussed, before the students' responses to the questionnaire are discussed. The students, on the whole, found peer evaluation to be enjoyable and helpful in developing their English speaking abilities. First years were less confident than second years both in their ability to evaluate and in performing before their peers, but a significant majority appreciated the experience. Self-evaluation was also viewed positively, and both will be amended and continued with in the future.

---

**[Keywords]** Peer evaluation, Self Evaluation, Student feedback

## Introduction

Over the last nine years the author has been on a journey to reach the best way of evaluating spoken English at a Japanese college. Direct testing, as recommended by Hughes (1989) has been central to all the testing performed, and as Heaton (1988:89) says 'oral tests can have an excellent backwash effect on the teaching that takes place' as the teacher strives to improve on how the students learn. Furthermore, if the aim is to test communication, the pressure of time that comes when speaking (Brazil, 1995) can not be replicated in a written test.

Initially students were evaluated in face to face individual interviews with the teacher. There were some advantages to this style, first and foremost being that the best students excelled as being on their own enabled them to display the full range of their abilities; something which is not the case in paired or group interviews as the better students then feel they must speak in simple English for their peers. It is also the easiest speaking test in which to replicate the same conditions as a teacher can control what they ask and say, making it as consistent as such a test can be. The weaknesses were that an interview can become a question and answer session as Madsen (1983) warns, and that the usual dynamics of a conversation (Cornbleet and Carter, 2001) cannot be achieved because of the power difference

between a teacher and a student (Bailey, 1998; Kormos, 1999). Furthermore, interviews also seemed to be too stressful for weaker students.

Consequently, this was then abandoned in favour of using groups of four as this should be sufficient for more natural dialogue (Underhill, 1987; Hughes, 2002) and students were able to choose the topic for discussion from a wide selection, which should also have benefitted students (Jennings et al., 1999). It was certainly less intimidating for weaker students, students did need to interact more and were required to ask questions, but the main failing was that it enabled weaker students to do the bare minimum, and stronger students were more reluctant to speak for fear of exposing the shortcomings of classmates.

The next change was to have students in pairs and to pair them according to ability, so that they would be of similar levels. Again this worked well for those who were accomplished at speaking English, but it felt as though putting the weaker students together stigmatised them and became self-fulfilling. The following year students were allowed to choose partners and though this eased the problem of compatibility, it meant that the stronger individual would do the majority of the speaking, and it became too easy for the weaker students. It also became apparent that some students would memorize chunks of dialogue and shoe-horn them into any topic thus defeating the purpose of

the test, so the following year paired interviews were performed to prevent memorization. In the same year, a new fourth year elective class began and peer and self evaluation were experimented with (Evans, 2008) and as it seemed to be a success it has been gradually introduced with younger students in compulsory English courses.

The attempt at self-evaluation by using grades was soon abandoned as the 4<sup>th</sup> year students conspicuously under-marked and an early trial with 2<sup>nd</sup> years showed some doing the same and others over-marking, so it was replaced with self-evaluation in written comments rather than as part of the grade. Peer evaluation was continued with, but unlike the 4<sup>th</sup> year students who evaluated each other's class work and a presentation, these students evaluated a role-play between a nurse and patient and an unprepared timed dialogue. The students' grades accounted for 50% of the final grade, the remainder being determined by the teacher, who as Underhill (1987) points out, will know the student's English level better than anyone, save for the student.

This paper examines the opinions of the students towards peer and self-evaluation as any problem areas will be improved next year so that the experience of the test-takers is enhanced (Alderson et al., 1995; Boud, 1995).

## Methods

### *The students and the classes*

The students who responded to the questionnaire are first and second years at a nursing college. There are 100 students in each year, and the classes are divided into 4 classes of 25 when college begins. They are not streamed, but divided according to the starting letter of the surname. The classes will be referred to as 1:1 and 1:2 for the first years, and 2:1 and 2:2 for the second years. There are two teachers, the author, an English native and a Japanese native. Each teacher has two classes of 25 in the first year, and this is repeated in the second year, except that the classes are swapped.

The first year students study general English from a regular English textbook (Helgesen et al. 1995) with some nursing English added in the form of handouts, whereas the second years study nursing English from a nursing English textbook (Chinen, 2001), with some everyday English added in the form of handouts. The focus of all classes is on communicative spoken English.

Classes are 90 minutes in length and in both semesters there are 7 classes. The classes within each year are as near as identical as is possible, as the book, the lesson plan and the teacher are the same. The students are predominantly female.

Both courses involve a large amount of speaking practice. Pair-work is the most common method of practice, though students do also work in groups of three and in larger groups on occasion.

As all the students have studied English at high school for 6 years, the main goal of the class is to get the students to turn their passive knowledge into active usage so that the students speak. As Richards (1993) notes, speaking is generally the poor relation in English study, though this is being gradually addressed through the use of a listening test in the centre exam. One of the most common activities is for students to practice a particular topic or function by walking around the room talking to as many different students as possible in the time allowed. This is unmonitored so that students can experiment with the English language. By changing the partner students can become more fluent in saying what they have said once already but without the same person having to hear it twice. The other person in the pair also hears two different voices and answers, as well as interacting with someone new, which keeps the practice interesting.

The activities are relatively straightforward, such as asking your partner about the weekend, plans for the holidays, and other common topics. Students might be given some preparation time if the task is particularly challenging, but usually they are not as the aim is to improve fluency. Students associate this activity with peer assessment, when in fact it is only a teaching technique.

Students will do more conventional pair work with the person sitting next to them, and though the teacher does request students not to always work with the same person, most of this work is done with a friend. They are encouraged to be ambitious in what they say and not to worry about making mistakes. Students can be 'frightened' of speaking English as in their previous experiences; English is 'right' or 'wrong', rather than simply a means of communication. The teacher will occasionally point out a class mistake but individual correction is postponed until English becomes an elective subject.

As all the students are training to be nurses, testing nursing English and the technical language involved is important and real for the students (Douglas, 2000).

### *Test preparation*

At the end of the summer term, and a week before the real test, students did practice tests in both years. In addition to this, first years had dummy tests mid-way through each term. This is important as without practice, such an exam would be too demanding (Angelo & Cross, 1993; Cohen, 1994). The teacher

also took in the practice grading sheets from students to see that they had correctly understood what to do. The most common problems were that students did not use a sufficiently wide range, rendering the grades redundant and did not make subtle enough distinctions between students; using 0 and 5 as in 80, 85 and not those in between. Students also filled in self-evaluation forms during the year.

### The exam

The exam format was the same for both 1<sup>st</sup> and 2<sup>nd</sup> years. It was a spoken exam taken in pairs and consisting of a prepared nursing dialogue of up to 3 minutes in length, and an unprepared 2 minute conversation. For the nursing dialogue students role-played a nurse-patient interaction. They chose their partner for this dialogue. The teacher had hoped that students would agree on an outline for their dialogue and then trust themselves to improvise but most did learn the dialogues by heart. The students drew lots to determine the order.

Each pair performed their dialogue to the class, and each pair was evaluated by both the teacher and the watching students. As well as entering a grade, students wrote comments. Once the last pair finished, the students handed in their assessment sheet to the teacher and had a fifteen minute break.

For the second dialogue the first question to begin the dialogue was written on the blackboard 5 minutes before the test began so that students could have an idea of what they would talk about but not be able to learn it by heart. The pairs were drawn at random from a bag.

Before the test began, the students were advised to judge the first few pairs more sympathetically as they were disadvantaged by going first. The students were also reminded not to see the students as a pair but as individuals. They were encouraged to try and differentiate in their grading as much as possible. The aim of the questionnaire was to find out how the students felt about self and peer evaluation and to identify any problems so that they could be addressed in future years. There are 12 questions in total which are a mixture of Likert scale selections (percentages, where given, have been rounded up) and guided responses in which students could write what they wished in response to a question.

### Ethical considerations

Approval from the ethics committee was obtained in 2008. At the beginning of the year the students were told that there would be peer evaluation (50% of the final grade comes from peer evaluation and 50% from the teacher). The students were asked to fill in the questionnaire after the grading was completed but

they could opt out of having their comments used in research. Seven first year students did so, but no second years, although some second years did not return the questionnaire.

## Results

### 1 Have you enjoyed doing peer evaluation?

	NO				YES
	1	2	3	4	5
1:1	2(8%)	3(12%)	4(17%)	9(38%)	6(25%)
1:2	1(4%)	4(17%)	9(39%)	7(31%)	2(9%)
Total	3(6%)	7(15%)	13(28%)	16(34%)	8(17%)
2:1	0(0%)	6(32%)	5(26%)	8(42%)	0(0%)
2:2	2(8%)	3(12%)	5(21%)	10(42%)	4(17%)
Total	2(5%)	9(21%)	10(23%)	18(42%)	4(9%)
<b>All</b>	<b>5(5%)</b>	<b>16(18%)</b>	<b>23(26%)</b>	<b>34(38%)</b>	<b>12(13%)</b>

Taken as a whole the first years are broadly in favour of peer evaluation, though not overwhelmingly so. An interesting difference emerges between the two classes, with more students in 1:1 class awarding peer evaluation a rating of 4 or 5. This might well be due to the 'mood maker' impact of particularly positive students which seems to influence others (Norcini, 2003).

As with the first years, the second years are broadly in favour of peer evaluation, and similarly, there is a noticeably more positive response from one class, class 2:2. Again, this can be attributed to 'mood makers' who positively influence their peers.

### 2 What did you enjoy or not enjoy about it?

The first years most commonly mentioned 'listening to classmates', 'getting to know classmates', and 'talking to my friends'. These have been grouped together as they all involve the notion of bonding. There were 22 mentions of this by first year students in comparison with only 8 by second years. The following are quotes from students.

*I could listen to many friends' opinions and conversation, so I enjoyed it.*

*I could make friends with people I wouldn't normally speak with.*

Listening to classmates is something that the students always excel at, and for first years peer evaluation gave them an early opportunity to know about their peers.

*It was interesting to hear my friend's English –it was very beneficial to me.*

This comment illustrates not only the positive effect of listening to peers but also how it allows students to learn from each other (Luoma, 2004) and to rank themselves within a class. For more competitive students this can be beneficial.

*I enjoy dialogues they're very interesting- I could get to know my friends well.*

There were some other isolated comments which were encouraging about the benefits peer evaluation.

*At first I was very shy but finally I could speak before other students.*

*Making conversations in public is difficult but enjoyable.*

These comments show that by being exposed to public speaking students gradually overcome their fears and reticence. This is beneficial not just for their English but for their general development.

*I enjoyed speaking to many friends but in public, I was very nervous.*

*Making dialogues was interesting, but talking in class was nerve-wracking.*

The above comments show that it needs to be approached sensitively, and for the atmosphere in the class to be as non-threatening as possible.

The most common difficulties for first years all related to the assessing of other students. Some felt that placing students within a range was difficult, others felt grading friends was, and some felt unprepared for writing comments.

*Making a score is very difficult. I don't know a lot of words, so I hardly comment, but I listen intently to the dialogues.*

*Difficult to judge because I put scores just by my impression of the speaker-I'm not sure if it's reliable or not. Basically I enjoyed being like the teacher.*

The students award an impressionistic grade, and the comments they write reveal the criteria they adopt. This means they use those measures they feel comfortable using. As Luoma (2004) says, students are not able to evaluate linguistic criteria reliably so it is best if they chose areas in which they feel competent.

One other important benefit of peer evaluation is hinted at in the last comment. By 'being teacher' students begin to develop their critical faculties, and this should lead to them becoming more efficient learners of English (Race et al., 2005). Not only do they analyse what makes one student successful or unsuccessful, but they also, if they practice giving feedback, develop the skills necessary for their future lives as leaders, in telling others how they need to improve.

A further benefit is that they can actually use what other students say. If one says 'By the way' to change the topic, they can copy it.

*I enjoy speaking to my friends in English in this class on many topics, but I didn't enjoy scoring my friends in a test.*

Only two students mentioned this, but for some it is hard to be

objective when grading close friends.

*Peer evaluation is difficult for me, because everyone is good, but I am not good. I found making the range difficult.*

This is a perennial problem with peer evaluation, and one that is difficult to overcome. Weak students, if they are evaluating other students, can not know if their inability to understand is due to the high level of the speakers or not.

The other most commented on problem was that of writing in English.

*It was a bother. I don't have the vocabulary, so don't write the things I think.*

*I don't have English knowledge, so I couldn't enjoy speaking English and writing comments.*

This is a minor problem of not having sufficient English to write comments, but it is problematic for weaker students. Some students are remarkably adept at pinpointing a strength or weakness, and some also felt that writing in English improved their ability.

One interesting dissenting opinion is written below.

*I didn't enjoy being compared with my classmates.*

In peer evaluation, as opposed to individual interviews, it is much more obvious that a student is being compared to his or her peers whereas in an individual interview this is not apparent to the test-taker, even though the interviewer will be doing so.

For the second years the most common response was that they had enjoyed the speaking aspect of the test. Some were pleased that it was a spoken test rather than a written one, others that they could do 'free talking' (the non-prepared test) and others because they learnt about their friends. There were 16 comments in total along these lines.

*Thinking in a conversation is interesting.*

Above is one of the more encouraging comments, which shows a student who is beginning to make a break through in conversation, and has reached a stage where talking without preparation is no longer intimidating but enjoyable. The comment below is less effusive, but gives cause for hope.

*I don't like speaking English, but it was a good experience.*

That a student who is not an English enthusiast should see the benefits of peer evaluation is very encouraging.

Another comment highlighted a different benefit of doing peer evaluation.

*I talked about many things and I could become confident.*

As with many things, 'practice makes perfect' or at least it takes away the fear factor and develops the students' confidence. The comment below shows that such a way of testing also can appeal to high achievers and competitive students.

*I could listen to many original dialogues.*

Both for the nursing dialogue and the unprepared dialogue each interaction is unique, and so holds the interest of the audience.

As for the negative points, eight students said they lacked the skill to do peer evaluation. As with the first years, this is a definite problem with peer evaluation.

*I have no ability to evaluate other people-I didn't enjoy it.*

Another comment similar to one from the first years is that it is all too much effort. Certainly, a more uniformed, positive reaction is likely from an elective class, but some dissent is inevitable.

*I thought to talk to all my classmates is a bother.*

### 3 Did you feel able to do peer evaluation reasonably well?

	NO				YES
	1	2	3	4	5
1 <sup>st</sup> years	2(4%)	12(24%)	20(41%)	14(29%)	1(2%)
2 <sup>nd</sup> years	1(2%)	6(14%)	16(38%)	14(33%)	5(12%)
<b>All</b>	<b>3(3%)</b>	<b>18(20%)</b>	<b>36(40%)</b>	<b>28(31%)</b>	<b>6(6%)</b>

Not surprisingly, the first years are not convinced by their ability to do peer evaluation well. In fact the results show an almost mirror image with 15 students choosing a rating above 3 and 14 choosing a rating below it. Evaluating spoken English is difficult, and will be more daunting the younger you are.

The second years are more confident in their ability to do peer evaluation, with 19 feeling they could do it very well and only 7 feeling they were poor. It might be the case that peer evaluation is more suited to older, more mature and confident students, or it might be that first years need more practice before they feel confident enough to do it.

If the two years are taken as a whole, the students are more in favour of peer evaluation than they are against it. Thirty-seven percent gave a rating of 4 or 5 as opposed to the 23 percent who gave it a 1 or 2. As mentioned earlier, this is more a cautious approval than a whole-hearted one and indicates that nearly a quarter of students are not in favour of peer evaluation.

### 4 What did you find most difficult in doing peer evaluation?

For this question first and second years could write what they wished. About half the students in both years mentioned either awarding a grade or difficulty in writing comments.

Awarding a grade is very difficult even for those with years of experience. The problems are alluded to in the following comments.

*When I decided the score I couldn't decide it easily! I usually give a high score to the first pair. After that I had to change the first decision.*

This comment shows the difficulty in doing peer evaluation, but also shows the evolving thought processes that peer evaluation triggers. Students are advised to listen to the first 3 pairs before awarding a grade so that they can have some context into which to decide the marks. The first marks that you award 'frame' the range that you will use, so it is important to become skilled at this. By doing frequent practices throughout the year, students should find this less difficult (Dudley-Evans & St. John, 1998).

*The difficult thing is scoring precisely.*

*To make the difference between 1 or 2 points. Sometimes I was confused-who is better than whom.*

As the comments here show, it is very difficult to make a distinction of 1 or 2 points in a hundred, but the process of doing so makes the students think carefully as to what they think makes a good speaker. In terms of the final grade such a difference for one marker is insignificant when there are 24 others, but the process is important.

*I think evaluating grammar is difficult.*

This is obviously very difficult for students with poor grammar, but the choice of what to evaluate was left to the individual students, so it is not necessary for all students to attempt to do so. With the next cohort, students are experimenting with evaluating a different aspect of speech in practice evaluations which should make them more confident when deciding what to evaluate in the final exam.

*When everyone prepares students deserve almost the same evaluation.*

This comment has highlighted a common feature of the student's grade sheets, which is the reluctance to differentiate between those in a pair in the nursing dialogue. In many ways the nursing dialogue is a way of measuring effort, as after the course there is no reason why even the weakest students should not be able to deliver a strong dialogue. As students cooperate in its preparation, the reasoning goes; the students should receive the same grade. Certainly, in cases where the division of speaking is the same, and when students' pronunciation is similar it is difficult to fault the logic.

*I thought every pair was good and the same, but I had to evaluate.*

*When everyone prepares students deserve almost the same evaluation.*

The above comments also add credence to the notion that the prepared dialogue is the easier of the two to perform but the harder to evaluate as effort in preparation can erode the advantage of better speakers of English. As the teacher, the concept of one test measuring hard work and the other

measuring the English proficiency is acceptable, as the main aim of the test is to evaluate what students have achieved in the 14 lessons rather than their proficiency in many cases would be excellent regardless of whether they had attended class or not.

The format for the unprepared dialogue is practiced prior to the exam, but it is inevitable that this will measure actual communicative ability due to the fact that it is unprepared.

*It isn't always good English to use difficult words.*

This is a very interesting observation and shows that the student is becoming more aware of how language is used. If the audience does not know a 'difficult' word, there is no point in using it.

One of the better students wrote the following comment

*Many students aren't good at English, so evaluation is very difficult. I think we can do them in a list of contents, e.g. intonation 1, 2, 3 etc.*

What is pleasing here for the teacher is that the student is identifying a problem and proposing a solution to it.

### 5 Did you find it enjoyable to see your classmates do their dialogues?

	NO				YES
	1	2	3	4	5
1 <sup>st</sup> years	0(0%)	1(2%)	4(8%)	21(43%)	23(47%)
2 <sup>nd</sup> years	0(0%)	1(2%)	2(5%)	18(42%)	22(51%)
<b>All</b>	<b>0(0%)</b>	<b>2(2%)</b>	<b>6(7%)</b>	<b>39(42%)</b>	<b>45(49%)</b>

This is perhaps the most favourable result of all. Ninety percent of the students gave a mark of 4 or 5 to this question. Students listen to each other more carefully than to anything else and this not only improves their listening ability but it acts as a bond within the class as they all 'endure' or 'enjoy' the experience together.

### 6 What was enjoyable/not enjoyable?

The answers to this question, in terms of what was enjoyable, were different between 1st and 2nd years. For first years the most enjoyable aspect was that they made friends with, and found out about classmates that they did not know. This illustrates both the benefit of changing partners regularly, especially in the first term when relationships are in flux, and again emphasises the importance of shared performance.

*It was very enjoyable. To talk about many things is very interesting and I can know my friends.*

*Listening to classmates I hardly speak to is interesting, but sometimes I can't hear what they say.*

*I enjoyed their unique ideas and different backgrounds.*

The comments above are a small sample of the sentiment.

There was only one negative comment, 'I was stressed so it was not enjoyable'. As mentioned before, performance anxiety is unavoidable when speaking a foreign language, but it should lessen with greater practice.

*Enjoyable thing was listening practice, not enjoyable was conversation, because it's very difficult, but it's good for our studies.*

The second years, as might be expected, made less of finding out about classmates, but made more of the challenges it presented.

*I could listen to many original dialogues.*

*Everybody has a different set and characters. It's very enjoyable.*

These comments show both an appreciation of their classmates' efforts, but also highlight the variety that emerges when students are free to script their own dialogues in the nursing practice, which in turn keeps the listening fresh and interesting. One student commented on the unprepared dialogue that she liked, 'The effort to try and keep speaking.' A comment such as this reveals that the student will become a very accomplished speaker.

### 7 Did you find it helpful to your performance to see other students?

	NO				YES
	1	2	3	4	5
1 <sup>st</sup> years	0	2(4%)	9(18%)	21(43%)	17(35%)
2 <sup>nd</sup> years	4(10%)	2(5%)	9(23%)	9(23%)	16(40%)
<b>All</b>	<b>4(4%)</b>	<b>4(4%)</b>	<b>18(20%)</b>	<b>30(34%)</b>	<b>33(37%)</b>

This is another encouraging result with the vast majority of both 1st and 2nd years giving a 4 or 5 to the question. Another advantage of peer evaluation is that students learn from each other (Luoma, 2004). Even in the test, it is not too late to learn something from a fellow student. When peer evaluation works effectively students both teach and learn.

### 8 Please give any examples.

Again there is a slight difference between the two years in what they focus on. The first years are much more aware of how the dialogue is presented and the demeanour of the speaker whereas second years concentrate more on the content and the kind of English used. The following are comments from first years.

*I was aware of other students' reactions, questions, subject and so on.*

*I know how to speak fluently, how to gesture, to smile and so on.*

First year students were also impressed by the help that was given by partners.

*I often forget words but my partner helps me.*

*Today's exam, when one student couldn't speak, her partner helped her. There were many cases like this.*

Students are encouraged to help their partners in the nursing dialogue when they become unsure of what to say as the exam is not meant to be an ordeal or a memory test. More importantly, in the unprepared dialogue, being able to help out a floundering partner is a very important conversational skill even in one's own language, and students that can do this in a foreign language are displaying excellent language ability.

Some students did also comment upon the language that they learnt from others.

*Other students' performance gave me ideas for vocabulary.*

*I found many ways of English expression*

Yet others made interesting discoveries of their own.

*I learned how to keep speaking.*

*When one person speaks too much, it's not a conversation it's a speech. A conversation is one short phrase and changing the speaker often.*

*I became courageous speaking to many people.*

The second years' comments were more connected to the nuts and bolts of conversation.

*I could learn accurate grammar.*

*The timing of the changing of the topic.*

*When students use good phrases I copied them and improved my English level!*

One other important facet of conversation was noted by another student.

*We knew what kind of attitude is good for the audience to enjoy.*

### 9 Did you find doing self-evaluation helped you?

Students were asked to fill in an evaluation form of their performance at various stages throughout the year. They were asked to write down things they had done well and things that they could improve upon for their next performance.

	NO				YES
	1	2	3	4	5
1 <sup>st</sup> years	1 (2%)	5 (11%)	14 (31%)	18 (40%)	7 (16%)
2 <sup>nd</sup> years	2 (5%)	5 (12%)	16 (39%)	8 (20%)	10 (24%)
<b>All</b>	<b>3 (3%)</b>	<b>10 (12%)</b>	<b>30 (35%)</b>	<b>26 (30%)</b>	<b>17 (20%)</b>

These figures are not an overwhelming vote of confidence in self-evaluation, but this is probably because the way of doing it is not as effective as it might be. Students were asked to self evaluate what they had done well and what could be improved,

but were not asked to suggest ways of improving.

An interesting finding in the scores above is that it is the two weaker classes in each year that more highly valued self-assessment. Although this might initially seem surprising, it is in fact quite logical. Students who find English difficult are more likely to benefit from thinking about their performance than those who are already accomplished.

### 10 In what way?

There were no great differences in response between first and second years and nothing particularly interesting or surprising. Samples from both years are below.

*I could use what I wrote for my self evaluation for my next performance.*

*Not only in my speaking, but also in thinking about my dialogue.*

*We can understand are own tendencies and this helps us to improve our skills.*

*By doing that I can know my lacking point and try to make it good.*

The one comment that will change next year's approach is this; *Self-evaluation is good for remembering my mistakes and the points that I need to improve but I don't know how to do that! That's the big problem for me.*

As the student correctly points out it is one thing to know what the problem is, but quite another to correct it. A section on how any deficiencies can be overcome will be included in next year's self-evaluation sheets.

### 11 Do you think peer evaluation should be continued with next year?

	YES	NO
1 <sup>st</sup> years	34 (76%)	11 (24%)
2 <sup>nd</sup> years*	32 (86%)	5 (14%)
<b>All</b>	<b>66 (80%)</b>	<b>16 (20%)</b>

\*one student wrote "don't know"

Overall 80% of all students feel that peer evaluation should continue so it does seem worthwhile to persevere and refine it. An interesting difference is that the second years are noticeably more in favour of peer evaluation than the first years. Maturity would seem to be an important factor in appreciating peer evaluation so this difference is not so surprising.

### 12 Why?

A vast array of answers was given by both first and second years and a combined selection of the most interesting is given below.

*It is a good chance for students to speak in the presence of*

others.

*Because it makes us try and use words or sentences appropriate for the occasion.*

*Can think of the next sentence by oneself.*

*It was very stressful for me but it helped my English speaking ability.*

*If there is no evaluation some students will not look at their performance carefully.*

*Because it makes us listen to English and to pay attention to the way it is used.*

*In the future we need to have the evaluation skill in nursing practice.*

There are many positive comments above and show that peer and self evaluation can continue to benefit students in speaking English. Those who were not in favour of wrote simply that it was 'hard' or that they did not have enough time to write comments.

### Conclusion

The students were predominantly positive about the experience of doing peer and self evaluation, and they do think that it should be continued. The first years were particularly appreciative of the bonding that it promoted whereas second years were more appreciative of how they were able to learn from each other. For some students in both years peer evaluation provided them with motivation by being able to rank one's own performance in comparison with others. Practice reduces the anxiety factor in both performance and in grading, and, not surprisingly, first years would seem to need more opportunities than second years.

Grading remains the most difficult feature for students, and as a consequence of the comments made by these students, the following years are experimenting with using different categories in practice evaluations so that students can find those areas they feel able to grade. As previously mentioned there is no point in stipulating categories such as use of grammar in the final exam if the students do not feel able to grade them. It is very difficult to reduce conversation to 4 or 5 distinct categories, so by leaving the choice up to individual students it should mean a wider range of features is being evaluated.

The nursing dialogue part of the test is a little problematic in so far as students tend to award the same grade for the pair. By practice it should be possible to tease out some differences but it might be necessary for the teacher to highlight potential areas of difference. Whereas a category such as 'a realistic dialogue' would warrant equal scores (assuming students had jointly

prepared it), 'a confident manner' might show one student to be better than her partner.

The unprepared dialogue is beneficial in the wash back effect it has on students in that it means all talking activities are useful for the exam. There is an element of luck in this section as you might draw a student who is non-communicative or very slow thinking, but this is a challenge that peer evaluation presents and, as students have complete autonomy in the grading, it is up to each individual as to how to respond. As the teacher, the challenges like this are a benefit of peer evaluation as they make the students think for themselves. Students could, for example, use their knowledge of the student's usual performance from class, and such use of continual assessment would mitigate too against a student underperforming due to anxiety.

Although some students were concerned about their lack of ability in writing comments the aim of the comments is to make the students think rather than just writing down a grade, and as such it is not so important if they can not express themselves perfectly. Rather than provide students with stock expressions, the current system enables students to say exactly what they want. The problem is more a lack of time than ability, so allowing more time would seem to be a better solution.

The use of self-evaluation is a potentially fertile area for benefitting students' English. Having abandoned the awarding a grade and substituting it with comments (submitted to the teacher), it means students are forced to think about how they can improve. By making students think about this it should encourage learner autonomy and for some it might even encourage them to apply this thinking to areas other than language.

### References

- Alderson, C. J., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Angelo, T., & Cross, P. T. (1993). *Classroom assessment techniques*. San Francisco: Jossey Bass Publishers.
- Bailey, K. (1998). *Learning about language assessment*. Boston: Heinle & Heinle.
- Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.
- Brazil, D. (1995). *A grammar of speech*. Oxford: Oxford University Press.
- Chinen, C. L. (2001). *'Christine no level up kango eikaiwa.'* Tokyo: Igaku Shoin.
- Cohen, A. (1994). *Assessing language ability in the classroom*. Boston: Heinle and Heinle.



- Cornbleet, S., & Carter, R. (2001). *The Language of Speech and Writing*. London: Routledge.
- 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.
- Evans, D. (2008). Reflections on peer evaluation in an English language course. *The Journal of Nursing Studies, National College of Nursing, Japan*, 7(1), 41-49.
- Heaton, J. (1988). *Writing English Language Tests*. New York: Longman.
- Helgesen, M., Brown, S., & Mandeville, T. (1999). *English Firsthand*. Hong Kong: Pearson.
- Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Hughes, R. (2002). *Teaching and Researching Speaking*. Hong Kong: Pearson.
- Jennings, M., Fox, J., Graves, B., & Shohamy, E. (1999). The test-takers' choice: an investigation of the effect of topic on language-test performance. *Language Testing*, 16(4), 426-456.
- Kormos, J. (1999). Stimulating conversations in oral-proficiency assessment: a conversation analysis of role plays and non-scripted interviews in language exams. *Language Testing*, 16(2), 163-188.
- Luoma, S. (2004). *Assessing Speaking Skills*. Cambridge: Cambridge University Press.
- Madsen, H. S. (1983). *Techniques in Testing*. Hong Kong: Oxford University Press.
- Norcini, J. J. (2003). Peer assessment of competence. *Medical Education*, 37, 539-543.
- Race, P., Brown, S., & Smith, B. (2005). *{500 Tips} on Assessment* (2<sup>nd</sup> edition). Oxon: Routledge Falmer.
- Richards, J. (1993). Real-World Listening in the Japanese Classroom. In *A Handbook for Teaching English at Japanese Colleges and Universities*. Oxford University Press.
- Underhill, N. (1987) *Testing Spoken Language*. Cambridge: Cambridge University Press.

受付日 2009 年 9 月 1 日 採用決定日 2009 年 11 月 26 日