

## Learning to self-assess oral performance in English: A longitudinal case study

Yuh-Mei Chen *National Chung Cheng University, Taiwan*

This paper reports on a study that investigated students' learning to self-assess oral performance in English by comparing student assessment with teacher assessment. Twenty-eight Chinese students at a university in southern Taiwan participated in the study. The assessment program involved training, observation, evaluation, discussion, feedback, and response, including two weeks of training and 10 weeks of two-cycle assessment. The assessment components were developed by the teacher and students collaboratively during the first week. The criteria included four elements within a five-level scoring standard. Comparison between self- and teacher assessment was analyzed in terms of scores and comments given over two assessment cycles. Results showed that self- and teacher ratings differed significantly in the first cycle of assessment, but were closely aligned in the second. Comments generated by students themselves in the second cycle were more similar to the teacher's, becoming more positive and constructive. A majority of the students favored participation in assessment and considered self-assessment conducive to learning. This study demonstrated that through feedback and practice, participating students make significant progress in learning to assess their own oral performance and that inviting students to be assessment partners helps Chinese students achieve desired learning outcomes.

**Keywords:** Assessment feedback; assessment procedure; collaborative assessment; effective student assessment; English as a foreign language; oral performance assessment; self-assessment

In line with theories of constructivism and learner autonomy, self-assessment is currently assuming a larger role in language teaching. The procedure involves students in judging their own learning, particularly their achievements and learning outcomes. Many have argued that teachers should help students construct knowledge through active involvement in assessing their own learning performance, and that students are empowered by gaining ownership of their learning and life-long learning skills. Research on language pedagogy especially recommends that teachers should provide opportunities for students to assess their language level so as to help them focus on

Address for correspondence: Yuh-Mei Chen, National Chung Cheng University, 168 University Road, Min-Hsiung, Chia-Yi 621, Taiwan; email: folcym@ccu.edu.tw

their own learning (Blanche, 1988; Blue, 1994; Dickinson, 1987; Harris, 1997; Henner-Stanchina & Holec, 1985; Oskarsson, 1989). Hunt, Gow, and Barnes (1989) even claim that without learner self-evaluation and self-assessment 'there can be no real autonomy' (p. 207). Oskarsson (1989) mentions six advantages of using self-assessment in the language classroom: promotion of learning, raised level of awareness, improved goal-orientation, expansion of range of assessment, shared assessment burden, and beneficial post-course effects. Blue (1994) identifies benefits such as encouraging greater effort, boosting self-confidence, and facilitating awareness of distinctions between competence and performance as well as self-awareness of learning strengths and weaknesses.

Despite the importance of self-assessment, learners are rarely put in charge of rating their own performance (Luoma & Tarnanen, 2003, p. 440). This is especially true in the context of Chinese learners of English as a foreign language. Some scholars have found that Chinese students tend to rely on the teacher and shy away from direct involvement in assessment. For example, Littlewood (1999) observed in his studies that East Asian students expect the teacher, as the holder of authority and knowledge, to be responsible for the learning assessment. Chan (1995, cited in Littlewood, 1999) reported that fewer than half of Chinese students wish to be involved in the evaluation process. Likewise, based on a questionnaire survey of Chinese English learners in Hong Kong, Chan, Spratt, and Humphreys (2002) found that the majority (76.3%) of respondents felt that evaluating their learning was a major part of the teacher's job. It should be noted that these findings were mostly based on learners' self-reports on the questionnaires. No information was provided regarding whether students had participated or received training in assessment before their opinions were collected for analysis.

In contrast to the above reports on Chinese learners' perspectives on participating in assessment, Chen (2006), in a study which asked 40 university students learning English as a foreign language to assess their own and peers' speaking performances, found that her students had positive attitudes towards participating in assessment and felt that student assessment, though less objective than teacher assessment, could supplement teacher assessment. However, like all other studies of student self-assessment, Chen's focused on comparing and correlating teacher and student generated marks, overlooking students' learning development through the assessment process.

Major reviews (Black & Wiliam, 1998a; Boud & Falchikov, 1989; Dochy, Segers, & Sluijsmans, 1999; Falchikov & Boud, 1989) and subsequent studies (e.g. AlFallay, 2004; Chen, 2006; Orsmond, Merry, & Reiling, 1997, 2000; Taras, 2001) have concluded that the ability to self-assess is extremely helpful and useful, but comparisons of student self-assessment and teacher assessment yield mixed results. Orsmond et al. (1997) warn that to serve as an effective assessment tool, self-assessment should not simply depend on students' ability to mark themselves according to criteria, nor on the similarity

of student and teacher marks; rather, effectiveness should be judged according to the extent of students' development during the assessment process. That is, in addition to correlation analysis of self- and teacher-awarded marks, how students apply the criteria to assessment, which indicates their understanding of the criteria, should not be overlooked in comparing self-assessment with teacher assessment.

Given limited knowledge of how Chinese students learn to assess their own performance in comparison with teacher assessment, this study investigated student development based on the marks and comments they gave themselves over two assessment cycles in a university course in English oral training.

## **I Theoretical framework**

The use of self-assessment is supported by theories of constructivism and learner autonomy. Epistemologically, constructivism asserts that knowledge is actively constructed by, not passively transferred to, individuals and that knowing is an adaptive activity in which one continually modifies one's knowledge of the world based on interaction with the environment (von Glasersfeld, 1989). Aligned with this constructivist concept of knowledge and knowing, the practice of self-assessment creates a setting for students to actively engage in discussions of how their learning performance will be evaluated, and what desired performance consists of, leading to reflection on what they have achieved with the help of peer and teacher feedback. Their knowledge of the assessed ability or skill and themselves is constructed and reconstructed through a dialectic process that examines their learning performance according to imposed or student–teacher jointly developed criteria.

Another reason for using self-assessment is based on conceptualizations of autonomous and life-long learning. In the early 1970s, the Council of Europe recognized that it was important for an individual to act autonomously in the rapidly changing society, and thus established self-directed learning as an integral part of permanent European education policy. Since then, rigorous efforts have been made to explore learner-centered procedures in order to enhance individual autonomy (Oscarson, 1997). Self-assessment, deemed a reification of autonomy, has been introduced into the classroom to motivate students and raise the levels of learning outcomes (Boud, 1995; Holec, 1988; Dickinson, 1987). Underpinning the concept of autonomy is the notion of lifelong learning. Many practitioners, especially in higher education, have delved into the use of periodic self-assessment to prepare students for effectiveness and improvement in their lives. Self-assessment is thus linked with the goal of lifelong learning and integrated into various subjects and domains. It has become not only a means to an end (autonomous lifelong learning), but an end itself (a crucial component of autonomy). Through the process of self-assessment, students learn to discern patterns of strength and weakness that can help them become better learners. Equipped with self-assessment skills,

students gradually develop a critical attitude toward learning throughout their lives and in the long run achieve the fullest autonomy.

Along with these theories of learning, the practice of self-assessment also reflects new thinking about classroom assessment: assessment for learning and empowerment evaluation. The assessment-for-learning approach shifts the focus from summative to formative assessment, from making judgments to prove that students have learned to providing feedback to help them learn. It strives to help close the gap between learners' current positions and desired goals. In order to involve everyone in monitoring the self-learning process, self-assessment is often recommended as an assessment procedure. For example, Black and Wiliam (1998a), after an extensive research review, concluded that self-assessment coupled with frequent performance feedback is an effective formative strategy that yields substantial learning gains. They argue that student self-assessment is an essential component of formative assessment and suggest that students should be trained in self-assessment if formative assessment is to be productive (Black & Wiliam, 1998b).

Another new notion implied in the practice of self-assessment is the democratic nature of classroom assessment. Self-assessment not only requires students to develop knowledge of standards of good work, make judgments about how well they have met the standards, and decide what to do next (Boud, 1995) but also empowers them by engaging them in assessment partnership with the teacher (Stefani, 1998). Traditionally, assessment has been assumed to be the realm of teachers, not students. But when power is shared in the classroom, assessment becomes a dialogue of sharing and negotiating understanding of the assessment criteria and standards between the teacher and students. This process fosters improvement and self-determination (Fetterman, Kaftarian, & Wandersman, 1996).

## **II Factors affecting self-assessment**

Research into the usefulness of self-assessment among ESL/EFL students to date has mainly investigated relationships between sets of scores and has yielded mixed results. Some studies have reported agreement between students' self-awarded ratings and ratings awarded by their teacher (AlFallay, 2004; Chen, 2006) or between scores that students expected to get on a test and those they actually obtained (Bachman & Palmer, 1989; LeBlanc & Painchaud, 1985). However, discrepancies have also been found between students' self-ratings and ratings from other sources (Blue, 1988, 1994; Patri, 2002; Wangsotorn, 1981; Yang, 2002).

Previous studies of self-assessment overemphasize students' ability to mark or predict their language performance in accordance with the given criteria or the authority (teacher), and overlook the value of self-assessment as a learning tool. Orsmond et al. (1997, 2000) alert us to the importance of students' development during all stages of the assessment process, and argue that in classroom practice,

we should not simply compare self-assessment with teacher assessment by looking at the correlation of self- and teacher-awarded marks. We should also consider how students apply the criteria to assessment and whether they are learning as a result of the assessment task. That is, the process of student self-assessment should be no less important than the product (rating), especially when it is being used as a tool to enhance learning.

Speaking of how to increase the effectiveness of self-assessment and engage students in assessment to activate deep learning, several factors are considered essential: clear criteria, training, intervention and feedback, and sufficient practice (AlFallay, 2004, Chen, 2006; Orsmond et al., 2000; Patri, 2002; Stefani, 1998; Taras, 2001). For instance, Airasian (1997) argues that clear criteria should be articulated to students in advance. When students are aware of the criteria for success, assessment is valid. Stiggins (2001) also urges teachers to identify and discuss performance criteria with students. When they clarify their expectations and explicitly state the desired performance outcomes, students' standards of achievement are likely to be raised. A typical example is Falchichov's (1986) model of self-, peer, and tutor collaborative assessment, in which the teacher first generates a set of characteristics of excellent essays, students then construct their own criteria as individuals and in groups, and after comparisons and discussions of teacher/student generated criteria, a list of agreed criteria is adopted for assessment. In support of this procedure, Orsmond et al. (2000), in a study in which students constructed their own marking criteria for peer and self-assessment, note more spontaneous and active student discussions.

Many scholars have recommended that before the actual assessment, training through a workshop or a tutorial should be practiced as a necessary step in developing students' understanding of the criteria and enhancing the quality of student assessment (e.g. AlFallay, 2004; Chen, 2006; Orsmond et al., 2000). For example, AlFallay (2004) implemented a three-hour workshop in the first week to ensure the accuracy of student assessment. Chen (2006) also provided four hours of training over two weeks. During the training, her students were asked to practice assessing two videotaped performances using teacher-student jointly developed criteria, then to discuss and share their evaluations and observations within groups and with the class. The teacher demonstrated her evaluations and comments on students' assessment results at the end. To facilitate students' understanding of the assessment procedure, Wiggins (1993) used examples of previously marked products for class discussion and analysis. Orsmond et al. (2002) found that the use of examples and formative feedback could help students demonstrate greater understanding of both marking criteria and subject standards while achieving higher-quality learning outcomes.

To direct students' learning effectively through assessment, some scholars have claimed that guidance and effective feedback are indispensable. In the writing classroom, Min (2005) and Stanley (1992) found that effective

teacher intervention was significant in helping students become successful peer reviewers and better writers. In the translation classroom, Taras (2003) reported that self-assessment integrated with tutor feedback helped students recognize their true strengths and weaknesses and focus on what work was really needed. Effective feedback refers to specific and constructive information that encourages students to look at their performance according to the criteria and standards. According to Chen (1998), it consists of *praise* (describing merit in accordance with the agreed criteria), *questions* (pointing out anything not up to the standards), and *suggestions* (providing specific remedial actions).

### III Method

#### 1 *The oral training course*

The study was conducted within a two-credit English oral training course at a national university in southern Taiwan. The class, meeting 100 minutes each week for a semester, aimed to help students express their ideas in clear and appropriate English, think critically and reflectively, assume responsibility for learning, collaborate with their peers, and evaluate their performance. Assessment was integrated into learning and teaching in this course. In addition to speaking tasks such as individual speech, pair talk, and group projects, students were required to assess their own and peers' oral performances as well as evaluate their participation in pair talk and group projects. The assessment task investigated in the study concerned only individual oral performances.

#### 2 *Participants*

Twenty-eight students participated: 18 females and 10 males, 22 English majors and six non-English majors. Before the study, all had studied English for 6–12 years. When asked to compare their English skills with those of their classmates on a 5-point scale, on average they rated their English listening skills at 3.14, speaking skills at 2.75, and learning attitude at 3.52. Of all participants, seven had experience in assessing and making comments on their own oral performances; only two had graded peer performances in other classes before.

#### 3 *Instruments*

An evaluation form (see Appendix A) and questionnaire (see Table 6) were used. The assessment components were developed by the teacher and students collaboratively. The assessment criteria included four elements: content (30%), language (30%), delivery (30%), and manner (10%). Scoring standards

included five levels: *excellent*, 90% and above, *good*, 80%–89%, *fair*, 70%–79%, *ok*, 60%–69%, and *poor*, below 60%. The questionnaire contained three parts, the first based on Chen's (2006) to elicit students' opinions about the practice of self-assessment; the second adapted from Falchikov's (1986) questionnaire asking students about their perspectives on the benefits of self-assessment. The questionnaire items were stated in both English and Chinese to avoid ambiguity and misunderstanding. Internal consistency was examined through calculating the Cronbach alpha coefficient. The resulting alpha value of 0.96 indicated high inter-item correlation.

#### 4 Procedures

Boud (1995) alerts teachers to some potential problems of the practice of self-assessment: students believing they are doing the teacher's job, self-marking without fully engaging in the whole process of assessment, and using self-assessment activities in isolation. To tackle these problems, the assessment program in the present study embraced the following steps: training, observation, evaluation, discussion, feedback, and response. Table 1 summarizes the time, goals, and activities of each step.

In the first two weeks of the class, students received training in peer and self-assessment. At the first class meeting, students were divided into groups of their own preference and informed of the assessment task required for the course. The class discussed and generated a set of criteria and scoring standards for assessing oral performance. In the second meeting, students practiced assessing two videotaped performances against the marking criteria. They discussed their observations with their group members and shared evaluation results in class. The teacher then commented on the videos to demonstrate her evaluation and scoring of the performances and gave students feedback on their evaluations.

After the training, each week the individuals in one group took turns giving oral presentations on topics of their choice for about 3–5 minutes. During each performance, both the teacher and peers simultaneously assessed the performance using the criteria and standards developed during the training. The performing students used the same form to complete self-assessment after their performance. After observing the performances, students exchanged comments or suggestions and observations within groups. Then peer groups and the teacher provided feedback to the performers. In the end, the whole assessment was concluded with performers' responses to the given comments.

Different from some procedures reported in the literature, self-assessment conducted in the present study asked students to assess their own performance in the classroom setting using the same criteria as those used for peer and teacher assessment. Students were not asked to rate their language abilities and/or predict their test scores in questionnaire format (e.g. Bachman &

**Table 1** Steps, time, goals, and activities of the assessment procedure

Step	Time	Goal	Activity
Training	2 weeks before assessment	<ul style="list-style-type: none"> <li>Comprehend the learning objectives of English oral performance/presentation</li> <li>Understand the assessment content and criteria</li> <li>Become conversant with the assessment procedure and process</li> </ul>	<p><i>Week 1</i></p> <ul style="list-style-type: none"> <li>Groups discuss criteria for English oral performance/presentation</li> <li>Class shares group discussion results</li> <li>Teacher summarizes and synthesizes students' opinions and guides students to generate assessment criteria</li> <li>Class determines criteria weights and scoring levels</li> </ul> <p><i>Week 2</i></p> <ul style="list-style-type: none"> <li>Class watches videos of two English oral performances</li> <li>Each group practices scoring and commenting on the video performances</li> <li>Groups share assessment results</li> <li>Teacher demonstrates her assessment of observed performances, explains her evaluation of each group's assessment, and comments on students' feedback</li> </ul>
Observation	2 cycles over 10 weeks. Each week 4-6 students perform for 3-5 mins. each	<ul style="list-style-type: none"> <li>Comprehend peers' English speeches</li> <li>Observe and learn from peers' speech contents, English usage and use, delivery skills, and demeanor</li> </ul>	<ul style="list-style-type: none"> <li>Individuals of one group engage in oral presentations and self-assessment, rest of class in peer assessment</li> </ul>



Evaluation	2–3 mins. after each performance	<ul style="list-style-type: none"> <li>Identify weaknesses of peers' speeches</li> <li>Analyze and distinguish strengths and weaknesses of peers' oral performances</li> <li>Use agreed criteria and scoring levels to evaluate each performance</li> </ul>	<ul style="list-style-type: none"> <li>Each student does 2 oral presentations and self-assessments 8 peer and assessments</li> <li>Each presentation based on topic of student's choice</li> <li>After each performance, peers and teacher use agreed on criteria and standards to assess the performance</li> </ul>
Discussion	12–15 mins.	<ul style="list-style-type: none"> <li>Communicate and negotiate assessment results with group members</li> <li>Use English to express opinions and positions</li> <li>Develop mutual respect and awareness of peer collaboration</li> </ul>	<ul style="list-style-type: none"> <li>Peer groups discuss observed performances</li> <li>Group speakers synthesize assessment comments</li> </ul>
Feedback	20 mins.	<ul style="list-style-type: none"> <li>Construct knowledge of self and others through multidimensional assessment</li> <li>Develop critical thinking ability</li> </ul>	<ul style="list-style-type: none"> <li>Peer groups and teacher comment on observed performances</li> <li>Commentaries must include praise, problem identification, and suggestion for improvement</li> </ul>
Reflection/Response	Each performer has 1 min. to respond to comments	<ul style="list-style-type: none"> <li>Listen to diverse opinions.</li> <li>Reflect on one's own performance</li> <li>Establish self-confidence and self-awareness</li> <li>Develop self-monitoring and self-regulated learning attitude</li> </ul>	<ul style="list-style-type: none"> <li>Each performer reflects on own performance or responds to given feedback</li> <li>Reflections or responses must include comments on strengths and analysis of weaknesses</li> </ul>

Palmer, 1989; Blue, 1994; LeBlanc & Painchaud, 1985). The self-assessment process was not used like learning logs for facilitating reflective and critical analysis of learning progress (e.g. Baldwin, 2000; Schon, 1987). Self-assessment was practiced as a tool to monitor students' own performance through critical feedback from peers and the teacher in the classroom. It was not 'an isolated or individualistic activity' (Boud, 1999, p. 122) but conducted with peer and teacher assessment. It also became an iterative process when students subconsciously compared peers with themselves in peer assessment.

### 5 *Data collection and analysis*

To ensure methodological triangulation, this study adopted multiple methods to gather quantitative and qualitative data. The quantitative data embraced scores obtained from two cycles of assessment and questionnaire responses. Because the sample was small and not normally distributed, the Spearman correlation test was chosen to detect the existence of a monotonic relationship between self- and teacher ratings, and the Wilcoxon match-pairs signed-ranks test was used to examine whether there was a symmetrical distribution of the median difference between the two assessment measures. The questionnaire responses were analyzed with the chi square test. The qualitative data contained written comments on the evaluation form and selective interviews. The comments were classified and coded by two EFL university teachers. The coding scheme included two categories: topic and nature. Comment topics were mainly classified in accordance with the assessment criteria: *content, language, delivery, manner*, and an emerging type of comment, *others*. Classification of the nature of the comments was based on Falchikov's (1996) taxonomy: *positive* comments are those identifying strengths, *negative* comments those identifying weaknesses, and *neutral* comments those revealing reflection and suggestions for improvement. (Coding examples are provided in Appendix B.) The inter-coder reliability indices of coding comment topics and nature categories ranged from 0.82 to 0.87. To examine whether the categories of assessors' comments differed between two cycles of assessment, the chi square test was used. The significance level of all the statistical tests in the study was set at 0.05.

### 6 *Reliability*

Because the assessment task in this study involved human raters, several procedures were adopted to ensure the reliability of the task. Assessment training was provided to ensure that students understood what and how to assess. Students were involved in developing the criteria and scoring levels with the teacher, and were asked to practice assessing oral performances with group members before undertaking the assessment task. All assessors were asked to use the same evaluation form to record analytical scores and comments. Assessment consistency was scrutinized by correlation analysis of the

scoring given in and outside class. It was hypothesized that student assessors were more experienced and accurate after the first cycle of practice, so the second was chosen for reliability analysis. All performances in the second cycle were videotaped and assessed again by the teacher and students outside the class at the end of the cycle. The Spearman correlation coefficients of two teacher and two student ratings yielded 0.83 and 0.80 respectively ( $n = 28$ ,  $p < .0001$ ).

## IV Results

### *1 Comparison of self- and teacher scoring*

Table 2 summarizes the descriptive statistics of self- and teacher scoring of oral performance in the two cycles of assessment. A look at the means showed that self-awarded marks tended to be lower than those given by the teacher. In the first cycle of assessment, students under-marked themselves so much that their mean scores deviated from the teacher's mean scores by about five points. However, the mean difference in the second cycle decreased to 1.4. Another look at the standard deviations indicated that scores obtained from self-assessment and teacher assessment spread over an identical range and were internally discriminating. The range of scores covered four intervals, although scores were mostly centered at the interval of 80–89. Initial readings of the scoring descriptive statistics indicated that self- and teacher assessment became more similar in the second cycle of assessment.

Further statistical tests were performed to examine the relationship between the two forms of assessment. As shown in Table 3, the correlation tests of self-and teacher assessment in the first and second cycles yielded indices of 0.5521 and 0.7938, respectively, and their  $p$  values were both less than 0.05. The result suggested that students' self-assessment scoring significantly varied from that of teacher assessment in the same direction. But the greater coefficient in the second cycle indicated stronger correlation between the two variables. This result was confirmed by the Wilcoxon matched-pairs signed-ranks test. In the first cycle the  $p$  value of the test was less than 0.05, suggesting rejection of the null hypothesis that both scorings were a match; however, in the second cycle the hypothesis was accepted with a  $p$  value greater than the significance level. That is to say, a significant difference was detected between the two sets of scores in the first cycle but not in the second.

To ascertain whether students had made significant progress in learning to assess, the Wilcoxon matched-pairs signed-ranks test of student-teacher scoring difference in two cycles was conducted. As shown in Table 4, the  $p$  value was less than 0.05, the set significance level, suggesting rejection of the null hypothesis that the median difference of the two was zero. The result revealed that the difference between self- and teacher scorings in the first cycle was not identical to that of the second cycle. A look at the mean scores shown in Table 2

**Table 2** Descriptive statistics of self- and teacher assessment in two cycles

	n	Mean	SD	Range	Min.	Max.	S ≥ 90	89 ≥ S ≥ 80	79 ≥ 70	69 ≥ S ≥ 60
<i>First cycle</i>										
SA	28	78.18	5.44	21	67	88	0	16	9	3
TA	28	83.07	5.54	26	69	95	4	20	3	1
<i>Second cycle</i>										
SA	28	83.07	4.79	24	67	91	3	21	3	1
TA	28	84.47	5.59	26	70	96	5	17	6	0

SA = self-assessment, TA = teacher assessment, SD = standard deviation, S ≥ 90: Score equal to or above 90

**Table 3** Spearman correlation and Wilcoxon matched-pairs signed-ranks test results of self-and teacher scorings in two cycles of assessment

	SA-TA Spearman correlation			Wilcoxon matched-pairs signed-ranks test			
	n	rho	p (2-tailed)	W+	W-	NSR	p (2-tailed)
<i>First cycle</i>	28	0.5521	0.002319	37.50	340.50	27	0.0003
<i>Second cycle</i>	28	0.7938	<0.0001	111.50	239.50	26	0.1068

NSR = number of signed ranks

**Table 4** Wilcoxon matched-pairs signed-ranks test of student-teacher scoring difference in two cycles

S-T diff.	W+	W-	NSR	p (2-tailed)
First cycle vs. Second cycle	79	272	26	0.01476

NSR = number of signed ranks

indicated that students' self-awarded scores differed less from those given by the teacher in the second cycle of assessment.

The above inferential statistics indicate that there was a significant match between self- and teacher scorings in the second cycle of assessment. Students' knowledge and skills obtained from the first cycle of assessment might have contributed to their improvement in assessment accuracy. The assessment task involved students in the process of internalizing a set of standards for good oral performance by observing and discussing peer performance, and reflecting on self-performance and responding to peer feedback. This learning process could have been the basis for validating self-assessment judgments.

**Table 5** Comparison of written comments from self- and teacher assessment in the first and second cycles

	1st SA	1st TA	$\chi^2$	p	2nd SA	2nd TA	$\chi^2$	p
<i>Topic of comments</i>			27.78	0.0000138			12.37	0.014802
Content	9 (15%)	25 (37%)			8 (19%)	19 (31%)		
Language	8 (13%)	18 (27%)			6 (14%)	20 (33%)		
Delivery	13 (22%)	20 (30%)			12 (29%)	16 (26%)		
Manner	23 (38%)	4 (6%)			10 (24%)	6 (10%)		
Others	7 (12%)	0 (0%)			6 (14%)	0 (0%)		
Effort/Preparation	3				2			
Audience response	2				2			
Lack of confidence	2				2			
Improvement								
Total	60 (100%)	67 (100%)			42 (100%)	61 (100%)		
<i>Nature of comments</i>			28.78	0.00000056299			2.92	0.23224
Positive	11 (18%)	35 (52%)			13 (31%)	30 (49%)		
Negative	40 (67%)	12 (18%)			13 (31%)	11 (18%)		
Neutral	9 (15%)	20 (30%)			16 (38%)	20 (33%)		
Total	60 (100%)	67 (100%)			42 (100%)	61 (100%)		

## 2 *Comparison of student and teacher comments*

Table 5 summarizes the frequencies of self-and teacher-given comments in terms of the topic and nature as well as the chi square test results. The topics of student comments in both cycles of assessment show that students paid attention to manner and delivery most. A majority identified nervousness and lack of eye contact as serious flaws in performances. In addition to the approved criteria – content, language, delivery, and manner – other variables were also mentioned in the comments. Some emphasized effort, audience amusement, confidence, and improvement as factors in determining the success or failure of their oral performance. However, the teacher commented more on aspects of content, language, and delivery. She noted students' language and delivery problems, explaining her preferences or suggestions for improvement.

The chi square test on the frequency distribution of the identified topics of comments found that variance existed between self- and teacher assessment in both cycles. Both  $p$  values obtained in the first and second cycles were less than the significance level, 0.05, indicating rejection of the null hypothesis that categories of self- and teacher-given comments have similar frequency distributions. The result suggested that students and the teacher emphasized different aspects of the established criteria in both cycles of assessment. Apparently students did not seem to observe their own performances against the criteria in the proportion expected (as the teacher did) nor did they show significant development in the second cycle.

Statistical analysis of comment topics did not support the finding that self- and teacher scorings became closely aligned in the second cycle of assessment. It was necessary to look into what helped students to bring their scoring more into line with the teacher's in the second cycle. A close look at the frequencies of these three types of comments showed that the ratio of negative comments in self-assessment was reduced from 67% to 31%, and those of positive and neutral comments increased. Apparently, in the first cycle students were extremely strict when evaluating their own performances and tended to look at their shortcomings while ignoring their merits and ways of solving their problems. However, in the second cycle, the frequencies of the three comment types changed. Students gave neutral feedback most and provided roughly equal amounts of positive and negative comments.

The chi square test found that self- and teacher assessment in relation to the nature of comments differed in the first cycle ( $p < .05$ ), but not in the second ( $p > .05$ ). The result suggested that after a cycle of assessment practice, the students not only began to notice their strengths and weaknesses but also learned to analyze their deficiencies and see how to eliminate them. Compared to teacher comments, students' comments tended to be more constructive and positive in the second cycle than the first. This change may have resulted from the teacher's guidance and encouragement. The teacher gave more positive and neutral feedback to students in both cycles of assessment. Below are some examples of student comments.

## 1st Cycle

### *Positive*

- \*I believe I had clear enough pronunciation. (A1S08)
- \*I spent so much time preparing for this topic. I'm glad that it's interesting. (A1S26)

### *Negative*

- \*Bad, bad, bad. I was so nervous than I expected. I think I did a bad job. (A1S22)
- \*I was very nervous when speaking English in front of others. (A1S21)
- \*My content is so flat, not interesting enough. (A1S20)
- \*I used too many "and" to connect sentences and too many pauses. That doesn't let listeners feel comfortable. (A1S05)
- \*Poor delivery, not lively at all. (A1S12)
- \*I spoke too fast. I was afraid I would forget what to say next. (A1S13)

### *Neutral*

- \*I should practice more to speak more fluently. I would practice by myself and record it and listen to it many times and modify my delivery. (A1S27)

## 2nd Cycle

### *Positive*

- \*I like the content I prepared for this show & tell. It's more interesting than last time, I think. (A2S24)
- \*I think I did better than last time, though my voice is still low. (A2S11)
- \*I slowed down my speech and talked more clearly this time. (A2S20)

### *Negative*

- \*My content was lack of vivid description of my experience. (A2S14)
- \*I was too nervous. I couldn't remember what I was going to tell. I have to be calmer next time!! (A2S07)

### *Neutral*

- \*I should practice more at home. I spoke too fast, so my talk became a little too short. (A2S06)
- \*I used a lot of new words for audiences. It's a little hard to understand the meanings of those words immediately. I should slow down my speech and checked whether they can follow me. (A2S21)
- \*The exact preparation of some words was not so well-said. I need to improve my pronunciation (especially the vowel/e/) and watch my grammar. (A2S10)

It is clear from the above that students learn through interaction with their learning environment. In this oral training class, students were required to assess their own and peers' oral performance for 10 weeks. After each performance, the class shared observations and evaluative feedback with the performers through teacher supervision. Giving and receiving feedback was an essential stage of the assessment procedure. Students' written comments demonstrated that they were more likely to reflect on their problems and identify areas for improvement over the course of time. A close reading of their written remarks also revealed a change from a negative tone of 'bad, bad, bad performance' to a reflective or suggestive one of 'I should ...' (meaning 'I should have') corresponding to the tone of the structure used by the teacher when giving suggestions, 'if ... it would be better'.

**Table 6** Students' questionnaire responses

I. Practice of self-assessment		A	NS	DA	$\chi^2$	p
Item/Statement						
1. Assessment should be the teacher's job only	5 (18%)	4 (14%)	19 (68%)	15.07	0.0005	
2. Students should take part in assessment	26 (93%)	2 (7%)	0 (0%)	44.86	<.0001	
3. Student assessment should not be used to supplement teacher assessment	5 (18%)	5 (18%)	18 (64%)	12.07	0.0024	
4. I feel comfortable about grading and commenting on my oral performance in this course	16 (57%)	10 (36%)	2 (7%)	10.57	0.0051	
5. I assess my oral performance based on the agreed on criteria and scoring levels	20 (71%)	6 (21%)	2 (7%)	19.14	<.0001	
6. I assess my own performance honestly	23 (82%)	5 (18%)	0 (0%)	31.36	<.0001	
7. Peer feedback is useful. I receive suggestions about how to improve my oral performance	22 (79%)	4 (14%)	2 (7%)	26	<.0001	
8. Teacher feedback is useful. I receive suggestions about how to improve my oral performance	27 (96%)	0 (0%)	1 (4%)	50.21	<.0001	

A = Agree, NS = Not Sure, DA = Disagree



II. Nature of self-assessment

Item/Statement	Choice	n (%)	NS	n (%)	Choice	n (%)	$\chi^2$	p
9. Self-assessment is	Fair	14 (50%)	NS	12 (43%)	Unfair	2 (7%)	8.86	0.0119
	Aids learning	18 (64%)	NS	8 (29%)	Doesn't aid learning	2 (7%)	14	0.0009
	Inaccurate	3 (11%)	NS	14 (50%)	Accurate	11 (39%)	6.93	0.0313
	Unbiased	11 (39%)	NS	11 (39%)	Biased	6 (21%)	1.79	0.4086
Good	Easy	6 (21%)	NS	5 (18%)	Hard	17 (61%)	9.5	0.0087
	Good	17 (61%)	NS	9 (32%)	Bad	2 (7%)	12.07	0.0024

III. Benefits of self-assessment

Item-assessment	Choice	n (%)	NS	n (%)	Choice	n (%)	$\chi^2$	p
10. Self-assessment is	Independent	16 (57%)	NS	5 (18%)	Dependent	7 (25%)	7.36	0.0252
	Not think	3 (11%)	NS	0 (0%)	Think	25 (89%)	39.93	<0.0001
	Learn more	19 (68%)	NS	6 (21%)	Not learn more	3 (11%)	15.5	0.0004
Lack confidence	Lack confidence	3 (11%)	NS	14 (50%)	Confident	11 (39%)	6.93	0.0313
	Critical	15 (54%)	NS	10 (36%)	Uncritical	3 (11%)	7.79	0.0203

A = Agree, NS = Not Sure, DA = Disagree

### 3 Students' perceptions of self-assessment

Table 6 summarizes students' questionnaire responses and chi square test results. The questionnaire was generally intended to elicit students' perceptions of self-assessment, including the practice, its nature and benefits. The chi square test was conducted to examine whether the frequency distribution of students' categorical opinions of each item differed. The results showed that a significant majority had positive perceptions of self-assessment.

Concerning the practice of self-assessment, 68% of students disagreed with the idea that assessment is exclusively the teacher's responsibility, 93% thought that students should take part in assessment, 64% believed that student assessment can be used to supplement teacher assessment, 57% stated that they felt comfortable with self-assessment, 71% claimed that their assessment was based on the agreed-on criteria and scoring levels, 82% stated that they assessed their own performances honestly, 79% thought peer feedback is useful, and 96% considered teacher feedback useful. Overall, the students were supportive and satisfied with the practice of self-assessment. They had been honest in their assessments and valued teacher and peer feedback.

As to the nature of self-assessment, a significant majority (50%–64%) of the students considered it fair, effective, helpful for learning, but difficult. When it comes to the accuracy of self-assessment, students seemed to reserve their opinions more. Although more tended to think it accurate than inaccurate, significantly, 50% showed uncertainty. Similar opinions were also revealed concerning whether self-assessment is biased or unbiased. No significant difference was detected in students' opinions of assessment bias; however, the frequency of 'not sure' and 'unbiased' was similar and each appeared more than 'biased'. Scoring involved students' subjective judgments about the quality of a performance. Perhaps awareness of potential bias had led some students to think it was difficult to be 'accurate and objective'. Nevertheless, more students largely considered self-assessment fair but challenging.

With regard to the benefits of self-assessment, a significant majority of students claimed that self-assessment made them independent, critical, think, and learn more, though half were not sure about whether the practice increased their confidence. In general, self-assessment was regarded as beneficial for learning.

## V Discussion

This study investigated students' development in learning to assess their oral performance in a university EFL class. After two weeks of training plus a cycle of assessment practice, students made significant progress in assessment accuracy. The statistical analysis of self- and teacher scoring demonstrated that the two sets of scores differed significantly in the first cycle of assessment but became nearly identical in the second. The results extended

the findings of previous studies (e.g. AlFallay, 2004; Chen, 2006) which concluded that practice contributes to the accuracy of self-assessment. While the previous studies simply reported correlation coefficients of the average grades obtained from several assessments over a period of time, this study developed direct evidence of students' progress in assessment practice.

Literature on student self-assessment often discusses its validity and reliability, or effectiveness, as an assessment tool based on agreement between self- and teacher scorings. How students' scoring comes to resemble the teacher's has been ignored. In this study, the data on students' comments suggest that teacher feedback has an impact on students' learning to look at their own performances neutrally. Comparison of written comments given by students and teacher shows that students seemed to pay attention to individual assessment criteria in varied degrees. Their emphases were different from the teacher's in both cycles of assessment. Yet, analysis of the nature of their comments found that in the second cycle students made comments more comparable to the teachers', moving toward the more positive and constructive. Such results indicate that students might have had a different understanding of some individual criteria compared with that of the teacher, but they gradually internalized the teacher's strategies used for commentaries. In the first cycle, students were very critical of themselves, focusing on their problems and mistakes 67% of the time. The teacher's feedback given during the assessment procedure apparently helped students to develop a more 'neutral' perspective on their own performances. In the second cycle, students took more note of their strengths and reflected more on what and how to improve. Significantly changing from overemphasizing weaknesses to highlighting merits and suggestions for improvement, the students learned to appreciate their own performances and naturally gave themselves higher scores than during the first cycle. Learning to assess in the present study indeed involved a process of achieving an understanding of one's own performance under the guidance of or in collaboration with the teacher. Students' changes presented evidence of their learning to assess effectively through interaction with teacher feedback.

The questionnaire data also confirmed students' learning through the assessment task. Students stated that they had learned more and become more critical and independent. The fact that more students favored participation in self-assessment did not differ from previous findings (e.g. Orsmond et al. 1997, 2000; Chen, 2006), but refuted the claim that students were reluctant to be involved in assessment or regarded assessment as the teacher's responsibility (Chan, 1995; Chan et al., 2002). Although their Chinese cultural background has been posited in the literature as a hindrance to autonomous learning (Littlewood, 1999), students who participated in the present study seemed to react to the practice of self-assessment autonomously and favorably. They may not have taken a proactive role in their learning, but their growth in self-assessment suggested that the assessment practice can be a successful practice of reactive autonomy. According to Littlewood, there are two types of

autonomy: proactive and reactive. Proactive learners 'take charge of their own learning, determine their objectives, select methods and techniques and evaluate what has been acquired' (Littlewood, 1999, p. 75), while reactive learners do not create their own directions. But once a direction has been initiated, they organize learning resources autonomously in order to reach the goals (1999, p. 75). As such, this study demonstrates that assessment partnership promises to help Chinese students achieve reactive autonomy.

Despite students' positive questionnaire responses, their uncertainty about self-assessment accuracy and its effect on self-confidence indicated transitional learning behavior. Their hesitation could be related to cultural values and learning experience. Chinese culture emphasizes strict self-discipline and generosity to others. Although students claimed they were honest in assessment, subjectivity of student assessment was noted. In the present study, peer assessment was conducted at the same time as self-assessment. Students were excessively critical of themselves when comparing their own to their peers' performances. Modesty is a virtue in Chinese culture. Showing modesty might have been regarded as being honest by the participating students, so they tended to give themselves moderate scores or scores lower than they deserved. Self-confidence refers to how much worth one sees in oneself and is closely related to self-identity. The participating students had grown up learning to obey authority and to be introspective, so they might have been confused about how their confidence could be enhanced by becoming public targets. During the assessment process, they were asked to let go of their concerns about face and examine themselves objectively compared to others. Those who tended to obey authority and accept themselves according to how the authority had defined them could have sensed that they were losing their self-identity when their identification had not yet been reconstructed. Although their awareness of their own strengths and weaknesses had been raised, students might not have been confident about becoming more self-aware as a step in developing and enhancing confidence. In this study, praise was required as part of the feedback given to each performer, but questions and suggestions were provided as well. As students observed problems or areas for improvement in self-assessment, their uncertainty about its impact on their self-confidence might have been inevitable.

## **VI Conclusion**

In terms of classroom practice and future research, this study includes the following implications and suggestions.

First, students made significant progress in assessing their own oral performance and benefited from this learning process. The assessment procedure used involved two processes: learning to assess and assessing to learn. When conducting self-assessment, students not only examined closely their performance of the desired skills and knowledge but also used the target language in

an authentic context. The evidence presented in this and a previous study (Chen, 2006) demonstrated that for university freshman and sophomore students, self-assessment is both a viable alternative to teacher assessment of oral performance and a useful learning task. Interested teachers could well use the implemented assessment procedures to encourage students to take greater responsibility for their own learning and engender deeper learning.

Second, students' supportive attitudes toward self-assessment in the study especially suggested the need for training and feedback. Training is a must for Chinese students, who are accustomed to a centralized education system and tend to accept so-called 'standard' or authoritative evaluations. In addition, students need to learn to assess through receiving and giving feedback. Self-assessment is best practiced in tandem with peer and teacher assessment. Interacting with the learning environment, students can judge their performance skills more accurately in relation to those of others. This self-assessment process allows students to seek knowledge from the surroundings to make sense of uncertainty, conflict, and doubt, and to draw a clearer self-profile. Training or tutorials are the first and foremost step in preparing students for the assessment procedure, and teacher and peer feedback is indispensable to monitoring assessment accuracy.

Third, students' personal traits or psychological characteristics may govern their assessment behavior. Students in this study were likely to under-mark themselves and emphasize affective factors such as confidence, nervousness, and audience feedback in oral assessment. Many researchers have claimed that the degree of leniency increases as the subjects' level of expertise decreases. For example, Boud and Falchikov (1989), in their extensive review of educational self-assessment studies, note that more able and experienced students tend to under-mark themselves in comparison with teacher marking. Blue (1994) and Orsmond et al. (1997) also report that lower level students tend to over-mark while better students tend to err on the side of underestimation. In addition to experience and proficiency, Blue (1994) speculates that nationality can be an important factor in self-assessment, proposing that some nationalities have a tendency to overestimate their level and others tend to underestimate. Considering the roles of some psychological and personality traits, Alfalay (2004) observes that those possessing the positive side of a trait are more accurate than those with its negative side. To understand the relation between Chinese students' personality traits and their self-assessment accuracy, correlation analysis and in-depth interviews are needed.

Fourth, the data were collected over a period of 12 weeks; what students learned was detected only in how they made comments or reflections on their own performances, not in what criteria they highlighted in assessment. In addition, their uncertainty regarding the impact of self-assessment on self-confidence could have been a transitional reaction to the nature of self-assessment. A study conducted over a longer period may improve our understanding of students' learning from the practice of self-assessment.

*Acknowledgements*

This paper reports on a project granted by the National Science Council.

**VII References**

- Airasian, P. W. (1997). *Classroom assessment* (3rd ed.). New York: McGraw-Hill.
- AlFallay, I. (2004). The role of some selected psychological and personality traits of the rater in the accuracy of self- and peer-assessment. *System*, 32, 407–425.
- Bachman, L., & Palmer, A. (1989). The construct validation of self ratings of communicative language ability. *Language Testing*, 6, 14–29.
- Baldwin, M. (2000). Does self-assessment in a group help students to learn? *Social Work Education*, 19(5), 451–462.
- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education*, 5(1), 7–74.
- Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–148. (Available online: <http://www.pdkintl.org/kappan/kbla9810.htm>)
- Blanche, P. (1988). Self-assessment of foreign language skills: Implications for teachers and researchers. *RELC Journal*, 19(1), 75–93.
- Blue, G. (1988). Self-assessment of listening comprehension. *International Review of Applied Linguistics*, 16, 149–156.
- Blue, G. (1994). Self-assessment of foreign language skills: Does it work? *CLE Working Paper*, 3, 18–35.
- Boud, D. (1995) *Enhancing learning through self assessment*. London: Kogan Page.
- Boud, D. (1999). Avoiding the traps: Seeking good practice in the use of self-assessment and reflection in professional courses. *Social Work Education*, 18(2), 121–133.
- Boud, D., & Falchikov, N. (1989). Quantitative studies of self-assessment in higher education: A critical analysis of findings. *Higher Education*, 18, 529–549.
- Chan, V., Spratt, M., & Humphreys, G. (2002). Autonomous language learning: Hong Kong Tertiary students' attitudes and behaviours. *Evaluation and Research in Education*, 16(1), 1–18.
- Chan, V. N. K. (1995). Communicative language teaching: Focus on the learner. Unpublished master's thesis, Hong Kong Polytechnic University.
- Chen, Y.-M. (1998). Peer review and learning styles. *Proceedings of the Seventh International Symposium and Book Fair on English Teaching* (pp. 189–198). Taipei: Crane.
- Chen, Y.-M. (2006). Peer and self-assessment for English oral performance: A study of reliability and learning benefits. *English Teaching and Learning*, 30(4), 1–22.
- Dickinson, L. (1987). *Self-instruction in language learning*. Cambridge: Cambridge University Press.
- Dickinson, L. & Carver, D. J. (1980). Learning how to learn: Steps towards self-direction in foreign language learning in schools, *English Language Teaching Journal*, 35, 1–7.

- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24(3), 331–350.
- Falchikov, J. N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, 59, 395–430.
- Falchikov, N. (1986). Product comparisons and process benefits of collaborative peer group and self assessments. *Assessment and Evaluation in Higher Education*, 11(2), 146–165.
- Falchikov, N. (1996). Improving learning through critical peer feedback and reflection. *HERDSA*. Retrieved on 20 October 2005 from <http://www.hersda.org.au/confs/1996/falchikov.htm>.
- Fetterman, D., Kaftarian, S., & Wandersman, A. (1996). *Empowerment evaluation*. Sage Publications.
- Harris, M. (1997). Self-assessment of language learning in formal settings. *English Language Teaching Journal*, 51(1), 12–20.
- Henner-Stanchina, C. & Holec, H. (1985). Evaluation in an autonomous learning scheme. In P. Riley (Ed.), *Discourse and learning* (pp. 255–263). London: Longman.
- Holec, H. (1981). *Autonomy and foreign language learning*. Oxford: Pergamon.
- Holec, H. (Ed.). (1988). *Autonomy and self-directed learning: Present fields of application*. Strasbourg: Council of Europe.
- Hunt, J., Gow, L. & Barnes, P. (1989). Learner self-evaluation and assessment – a tool for autonomy in the language learning classroom. In V. Bickley (Ed.), *Language teaching and learning styles within and across cultures* (pp. 207–217). Hong Kong: Institute of Language in Education.
- LeBlanc, R., & Painchaud, G. (1985). Self-assessment as a second language placement instrument. *TESOL Quarterly*, 19(4), 673–687.
- Littlewood, W. (1999). Defining and developing autonomy in East Asian contexts. *Applied Linguistics*, 20(1), 71–94.
- Luoma, S., & Tarnanen, M. (2003). Creating a self-rating instrument for second language writing: From idea to implementation. *Language Testing*, 20(4), 440–465.
- Min, H.-T. (2005). Training students to become successful peer reviewers. *System*, 33(2), 293–308.
- Orsmond, P., Merry, S., & Reiling, K. (1997). A study in self-assessment: Tutor and students' perception of performance criteria. *Assessment and Evaluation in Higher Education*, 22(4), 357–369.
- Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment and Evaluation in Higher Education*, 25(1), 23–38.
- Orsmond, P., Merry, S., & Reiling, K. (2002). The use of exemplars and formative feedback when using student derived marking criteria in peer and self-assessment. *Assessment and Evaluation in Higher Education*, 27(4), 309–323.
- Oscarson, M. [formerly spelt Oskarsson]. (1997). Self-assessment of foreign and second language proficiency. In *The encyclopedia of language and education, Vol. 7: Language testing and assessment* (pp. 175–187). Dordrecht: Kluwer Academic.

- Oskarsson, M. (1989). Self-assessment of language proficiency: Rationale and applications. *Language Testing*, 6(1), 1–13.
- Patri, M. (2002). The influence of peer feedback on self and peer-assessment of oral skills. *Language Testing*, 19(2), 109–131.
- Schon, D. (1987). *Educating the reflective practitioner*. San Francisco: Jossey-Bass.
- Stanley, J. (1992). Coaching student writers to be effective peer evaluators. *Journal of Second Language Writing*, 1, 217–233.
- Stefani, L. J. (1998). Assessment in partnership with learners. *Assessment & Evaluation in Higher Education*, 23(4), 339–350.
- Stiggins, R. J. (2001). *Student-involved classroom assessment* (3rd ed.). Upper Saddle River, NJ: Merrill.
- Taras, M. (2001). The use of tutor feedback and student self-assessment in summative assessment tasks: Toward transparency for students and for tutors. *Assessment and Evaluation in Higher Education*, 26, 605–614.
- Taras, M. (2003). To feedback or not to feedback in student self-assessment. *Assessment & Evaluation in Higher Education*, 28(5), 549–565.
- von Glasersfeld, E. (1989). Constructivism in education. In T. Husen & N. Postlewaite (Eds.), *International encyclopedia of education: Research and studies. Supplementary Volume* (pp. 162–163). Oxford: Pergamon Press.
- Wangsotorn, A. (1981). Self-assessment in English skills by undergraduate and graduate students in Thai universities. In J. A. Read (Ed.), *Directions in language testing: Selected papers from the RELC seminar on evaluation and measurement of language competence and performance* (pp. 240–260). Singapore: Singapore University Press.
- Wiggins, G. P. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. San Francisco: Jossey Bass.
- Williams, E. (1992). Student attitudes towards approaches to learning and assessment. *Assessment and Evaluation in Higher Education*, 17, 45–58.
- Yang, T. H. (2002). Self-assessment as a placement tool for oral training courses. *Proceedings of the 19th International Conference on English Teaching and Learning in the Republic of China* (pp. 479–88). Taiwan: Crane.



## Appendix A: Evaluation form

Evaluator: \_\_\_\_\_ Group: \_\_\_\_\_ Date: \_\_\_\_\_

Scoring Criterion/Level	Excellent (90% & above): Performed at a very high level, i.e. at a level that makes it exceptional for the class.	Good (80%–89%): Performed at a high level, i.e. at a level that clearly exceeds competency.	Fair (70%–79%): Performed without being exceptional in any way but thought of as competent.	Ok (60%–69%): Performed at a minimally acceptable level with flaws that are not serious enough to merit a failing mark.	Poor (below 60%): Performed at a very low level, without showing any positive quality. Having serious flaws.
Content (30%) Language (30%) Delivery (30%)	27–30	24–26	21–23	18–20	17–
Manner (10%)	9–10	8	7	6	5–
Content	Wonderful, surprising, interesting, fascinating, flat, boring, not organized, not having a point		Delivery	Fluent, good-guidance, with expressive voice, with various gestures, with visual aids, with eye contact with the audience, average, not lively at all, dull	
Language	Precise, error-free, clear pronunciation, appropriate, good use of transitions, poor grammar, unacceptable use of words		Manner	Calm, polite, graceful, dramatic, appropriate, no indication of nervousness or irrelevant body language	
<b>Performer</b>	<b>Criteria</b>	<b>Score</b>	<b>Comments</b>		
	Content 30%				
	Language 30%				
	Delivery 30%				
	Manner 10%				
	Total 100%				
<b>Performer</b>	<b>Criteria</b>	<b>Score</b>	<b>Comments</b>		
	Content 30%				
	Language 30%				
	Delivery 30%				
	Manner 10%				
	Total 100%				
<b>Performer</b>	<b>Criteria</b>	<b>Score</b>	<b>Comments</b>		
	Content 30%				
	Language 30%				
	Delivery 30%				
	Manner 10%				
	Total 100%				
<b>Performer</b>	<b>Criteria</b>	<b>Score</b>	<b>Comments</b>		
	Content 30%				
	Language 30%				
	Delivery 30%				
	Manner 10%				
	Total 100%				

## Appendix B: Coding examples of student and teacher comments

Student	Teacher
<i>Topic of Comments</i>	
Content	
I think my content was nice for my feeling is true (A1S15)	Interesting topic, talking about her idol. (A1T25)
My content is so flat, not interesting enough (A1S20)	Talked about two things: her comb and her volleyball practice in senior high. But better if told the audience how the comb is related to her senior high school life at the beginning. (A1T14)
Although I skipped some content, what I said is still complete. I'm glad that it's interesting to you. (A2S28)	Interesting topic (a new product), better if explained its functions clearly. (A2T11)
I think my content is not bad, it's very interesting. (A2S26)	Touching story, but needs to explain what is sent to her in the package at first. (A2T17)
Language	
I used too many 'and' to connect sentences and too many pauses. That doesn't let listeners feel comfortable. (A1S05)	Good grammar and wording. Clear pronunciation. (A1T02)
I need to enhance my pronunciation. I made some mistakes in /ɛ/ and /ei/. (A1S03)	Loud, clear pronunciation. Used a variety of adjectives to talk about his favorite music. (A1T05)
I slowed down my speech and talked more clearly this time. (A2S20)	Intonation tends to rise in the end. Adding an extra vowel in the end (ghost becomes ghoster). Could use some singing to better her pronunciation. (A2T21)
The exact preparation of some words was not so well-said. I need to improve my pronunciation and watch my grammar. I still made some errors. (A2S10)	Tends to have false starts too often. Could use recording to check and cut her ums and ahs. (A2T16)
Delivery	
Poor delivery, not lively at all (A1S12)	Reading notes only, little connection with the audience. Better if she could've memorized some key phrases or sentences and outlined the points on smaller cards for occasional reference. (A1T10)
I spoke too fast. I was afraid I would forget what to say next. (A1S13)	Good guidance for the audience. Asked questions at the beginning to involve the audience. (A1T28)
I did not speak fluently and expressively as others. (A2S16)	Amusing ways to attract the audience attention (A2T28)

## Appendix B (Continued)

---

I used a lot of new words for audiences. It's a little hard to understand the meanings of those words immediately. I should slow down my speech and check whether they can follow me. (A2S21)

Natural gestures with expressive voice. (A2T04)

### Manner

Bad, bad, bad. I was so nervous than I expected. I think I did a bad job. (A1S22)  
I was very nervous when speaking English in front of others. (A1S21)  
I was too nervous. I couldn't remember what I was going to tell. I have to be calmer next time!! (A2S07)  
I felt nervous so I know I stammered. (A2S08).

Calm and polite. Greeting to the audience at the beginning. (A1T02)  
Dramatic. He's a natural performer. (A1T04)  
Smiling while talking. Gives the audience a pleasant feeling. (A2T26)

### Others

There's still room for me to improve, but since I've worked hard to prepare my show & tell, I would give myself a big hand anyway. (A1S07)  
I'm very delightful that I heard some people laughing when I take out my pillow. I deeply appreciated my audiences gave some good response to me! (A1S17)  
I was shy. It was a pity that some of my content I prepared was not performed because of my lack of confidence. (A1S01)  
I did not prepare very well. The history of FIFA World Cup is hard to memorize. (A2S27)  
I think I did better than last time, though my voice is still low. (A2S11)  
Through the laughters of the audience I know I had good interaction with them. (A2S08)

### Nature of Comments

#### Positive

I believe I had clear enough pronunciation. (A1S08)  
I spent so much time preparing for this topic. I'm glad that it's interesting. (A1S26)

Content substantial and organized. Informative and interesting to students of your age. (A1T02)  
Enthusiastic in delivery. Eager to build a connection with the audience. Used a funny dialogue to interest the audience. (A1T01)

---

<p>I think my content is quite special. (A2S09)</p>	<p>With demo and pictures to explain what happened to her when the analyst injected a need to her spine. Effective way to guide the audience. (A2T11) Using humorous expressions to make the audience interested and amused. (A2T15)</p>
<p>I like the content I prepared for this show &amp; tell. It's more interesting than last time, I think. (A2S24)</p>	
<p>Negative</p>	
<p>I was too nervous. I forgot what to say and had a little stammer. (A1S23)</p>	<p>Tended to add schwa to the ending voiceless stops such as /t/ in note, /p/ in hope, /k/ in take. (A1T21)</p>
<p>I used too many 'and' to connect sentences and too many pauses. That doesn't let listeners feel comfortable. (A1S05)</p>	<p>Misuse 'medal' for 'citation,' 'athletic meeting' for 'athletic meet'. (A1T15)</p>
<p>I was too nervous. I couldn't remember what I was going to tell. I have to be calmer next time!! (A2S07)</p>	<p>No eye contact, no communication with the audience. Reading notes only. (A2T18)</p>
<p>My content was lack of vivid description of my experience. (A2S14)</p>	<p>Problem with the th sound in brother, mother, father. (A2T13)</p>
<p>Neutral</p>	
<p>I need to enhance my pronunciation. I made some mistakes in /ɛ/ and /ei/. (A1S03)</p>	<p>Better if she could've memorized some key phrases or sentences and outlined the points on smaller cards for occasional reference. (A1T10)</p>
<p>I should practice more to speak more fluently. I would practice by myself and record it and listen to it many times and modify my delivery. (A1S27)</p>	<p>Talked about two things: her comb and her volleyball practice in senior high. But better if told the audience how the comb is related to her senior high school life at the beginning. (A1T14)</p>
<p>I should practice more at home. I spoke too fast, so my talk became a little too short. I think stage fright really affected me a lot. (A2S06)</p>	<p>Like self-examination, self-revealing. But could've stressed more on how to adjust himself to the new learning environment. (A2T19)</p>
<p>I used a lot of new words for audiences. It's a little hard to understand the meanings of those words immediately. I should slow down my speech and check whether they can follow me. (A2S21)</p>	<p>Talked about her family, mocking herself. To enrich her content and link her ideas, she could've talked more about how she felt about her parents, then explained she changed her views of her parents' love for her. (A2T13)</p>

---

A1S27: example from Student 27 in the first cycle of assessment

A2S21: example from Student 21 in the second cycle of assessment

A1T15: example from teacher assessment of Student 15 in the first cycle of assessment

A2T13: example from teacher assessment of Student 13 in the second cycle of assessment