Self and Peer Assessment Consistency over Time

Alireza Ahmadi [*]
Assistant professor of TEFL, Department of English & Linguistics, Faculty of Literature & Humanities, Shiraz University, Shiraz, Iran

Abstract
This article investigated and compared the consistency of self and peer assessments as alternatives for teacher assessment. Thirty sophomores majoring in TEFL were asked to assess their classmates' as well as their own speaking ability in a conversation class. They were taught how to do this using a rating scale of speaking. They did the rating twice during the term; the first rating was carried out during the 8th and 9th weeks and the second rating at the end of the term (weeks 15 and 16). The results of the study indicated that self and peer assessments were not significantly related at the end of the term and only loosely, though significantly, related in the middle of the term. Both self and peer assessments indicated consistency over time, however peer assessment enjoyed a higher consistency.

Keywords: Self assessment; Peer assessment; Consistency; Speaking rating scale

[*] *E-mail address*: ar.ahmadi@yahoo.com

Introduction

Assessment is of crucial concern in any educational operation which could be materialized in different ways. Ordinarily it's the teacher who is responsible for the assessment in classroom. But a growing number of studies support the use of peer and self assessments as alternatives for teacher assessment (e.g. AlFallay, 2001; Boud, 1989; Boud, 1995; Brown and Hudson, 1998; Falchikov, 1995; Magin and Helmore, 2001; Patri, 2002; Rada and Hu, 2002; Stefani, 1998; Tudor, 1996; Woolhouse, 1999). It has been argued that these types of assessment lead to effective learning. Students learn to be critical of others' work and receive critical appraisal of and feedback on their own work. They develop to think critically about others and their own learning**.**

Peer assessment which is defined as "an arrangement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of the learning of peers of similar status" (Topping, 1998, p. 250) has extensively been used in diverse fields (e.g. Falchikov, 1995; Freeman and McKenzie, 2002; Sluijsmans et al., 2002). The peer assessment tasks can be regarded as the learning exercises in which the assessment skills are practiced (Sluijsmans et al., 2002). Students have an opportunity to observe their peers through the learning process and often have a more detailed knowledge of the work of others than do their teachers (Dochy, Segers, and Sluijsmans, 1999). Peer assessment may lead to developing autonomy, better understanding of what happens in classes, and an objective view toward one's own and others' work. This type of assessment necessitates an open marking system and this provides an opportunity to see standards set by peers as well as their mistakes. There is also the hope that assessors gain an ability to 'stand back' from their own work and assess objectively. It is possible that as students become familiar with the way in which marking criteria are implemented they improve their understanding of assessment procedures (Langan and Wheater, 2003). Furthermore peer assessment can help students plan their own learning, identify their own strengths and weaknesses, target areas for remedial action, develop meta-cognitive and professional transferable skills, and enhance their reactive thinking

and problem solving abilities during the learning experience (Sluijsmans, Dochy, and Moerkerke, 1999; Smith, Cooper, and Lancaster, 2002; Topping, 1998). Peer assessment has also been found to be a reliable and valid method for assessment and teaching (Falchikov and Goldnch, 2000; Topping, 1998) , to increase students' interpersonal relationships in the classroom (Sluijsmans et. al., 2002), and to facilitate a deep approach to language learning (Cheng and Warren, 2005).

Like peer assessment, self assessment has also been emphasized in the realm of assessment. It is believed that self assessment helps to open up wider perspectives on the learning process (Oscarson, 1989), increase student and teacher motivation (Ross, 1998), develop critical thinking, and emphasize self reflection (Fletcher and Baldry, 2000). Through self assessment, students can also gain an insider access to the interaction between curriculum and assessment which is fundamental to any worthwhile educational enterprise (Little, 2005). Three reasons have been mentioned by Little (2005) for engaging students in self assessment. First of all, a learner-centered curriculum is not complete if it involves students in decisions concerning the content of the curriculum but prevents them from having a voice in evaluating the curriculum including their own learning. Secondly, self assessment plays a key role in developing autonomy and thirdly,

> To the extent that languages learnt in formal contexts are to be used in the world beyond the classroom, a capacity for accurate self-assessment is an essential part of the toolkit that allows learners to turn occasions of target language use into opportunities for further explicit language training (p. 322).

Review of literature

Self and peer assessments have been studied from different perspectives. Some studies have focused on the specific benefits of self and peer assessments. Tseng and Tsai (2007) for example, have indicated that on-line peer assessment can significantly enhance students' quality of

projects, as it provides students with opportunities of learning not only from other peers but also from evaluating other peers' work. They state that the learning in peer assessment comes from both students' adaptation of peers' feedback and their assessment of peers' projects. Furthermore, with the implementation of networked peer assessment, it is believed that the teaching load could be somewhat reduced for instructors.

Many studies have also shown various benefits that students receive in the writing learning process by using peer assessment (Berg et al., 2006; Boud, Cohen, and Sampson, 1999; Davies, 2006; Falchikov, 1995; Plutsky and Wilson, 2004; Stefani, 1994; Topping, 1998; Topping et al., 2000) because the major activities of writing such as editing and reviewing could be very similar to the process of peer assessment (McIsasc and Sepe, 1996).

After using a peer assessment method in teaching the process of scientific writing to 39 undergraduate students, Guilford (2001) found that participants learned more course content knowledge and technical skill-writing for publication through the peer review than through traditional term paper approaches. A similar finding was also found in a study by Venables and Summit (2003).

Some researchers have pinpointed the key role that peer feedback plays in students' learning in peer assessment. Richer (1992) for instance, compared the effects that two kinds of feedback, peer directed and teacher based, had on first year college students' writing proficiency in an experimental study with 87 participants. The study results showed that there was a significant difference in writing proficiency in favor of the peer-feedback-only group. The finding indicated that using peer feedback provides a feasible method enabling college students to enhance their writing skills and improve their learning achievement. In the same line, Topping et. al. (2000) have argued that formative assessment seems likely to be most helpful if it yields rich and detailed qualitative feedback information about strengths and weaknesses, not

merely a mark or a grade. This idea has been underlined by other researchers as well (Black and Wiliam, 1998; Chaudron, 1983; Davies, 2006; Lin, Liu, and Yuan, 2001; Paulus, 1999; Plutsky and Wilson, 2004; Stefani, 1998).

Some other studies have underlined the effect of psychological and personality traits on the accuracy of self and peer assessments. AlFallay (2004), for example, investigated the role of motivation types, self-esteem, anxiety, motivational intensity, and achievement in the accuracy of self and peer assessments. The study concluded that learners possessing the positive side of a trait are more accurate than those who have its negative side, with the exception of students with high classroom anxiety. The study also demonstrated that students with low self-esteem are the most accurate in assessing their performance, whereas learners with instrumental motivation are the least accurate. Similar studies were formerly conducted by AlFallay (2001) and Lin et al. (2001). Other studies have focused on the effect of proficiency level on self and peer assessments (e.g. Davidson and Henning, 1985; Heilenmann, 1990; Janssen-van Dieten, 1989) or gender bias and fairness in self and peer assessments (e.g. Bradley, 1993; Falchikov and Magin, 1997; Newstead and Dennis, 1990).

Different reactions to peer assessment are also found in the literature. Zhao (1998), for example, found that in face-to-face peer assessment, students frequently expressed anxiety in sharing their feedback for fear of being wrong or rejected by peers. Likewise, Macleod (1999) reported that some students doing face-to-face peer assessment were caused to be dishonest in giving feedback because of interpersonal relationships. Following the same line, Topping et. al. (2000) found that most students considered peer assessment as time consuming, intellectually challenging, and socially uncomfortable although it was effective in improving their learning.

However, a line of research which has attracted a good number of researchers is the investigation of the reliability and validity of self and

peer assessments as alternatives for the teacher assessment. Studies in this regard have yielded mixed results. A number of studies has indicated high reliability and validity for peer-assessment. Tseng and Tsai (2007), for instance, examined the consistency between peer assessment scores and expert (teacher) scores. It was found that the correlations between these scores were significantly high, implying that peer assessment could be perceived as a valid assessment method. In the same line, Bailey (1998) found high correlations (between 0.58 to 0.64) between self-rated oral production ability and scores on POI concluding that learners' self-assessments may be more accurate than one might suppose. Similar results were formerly found by other researchers (e. g. Alfallay, 2004; Bachman and Palmer, 1989; Cho et al., 2006; Haaga, 1993; Patri, 2002; Ross, 1998; Saito and Fujita, 2004; Stefani, 1994; Sullivan and Hall, 1997; Williams, 1992; Xiao and Lucking, 2008). However, the reliability and validity of self and peer assessments have also been challenged by other researchers (e. g. Blue, 1988; Cheng and Warren, 1999; Falchikov and Magin, 1997; Hughes and Large, 1993; Mowl and Pain, 1995; Mowl and Pain, 1995; Orsmond et al., 2000; Wesche et al., 1990).

Following the same track, some other researchers have tried to investigate the reliability and validity of self and peer assessments under the conditions of anonymous review (Lu and Bol, 2007; Zhao, 1998) or employing multiple raters (Cho, Schunn, and Wilson 2006; Fagot, 1991; Ferguston, 1966; Magin, 1993; Xiao and Lucking, 2008).

However most of the studies conducted so far on the validity and reliability of self and peer assessments have been mostly studies of validity not reliability. That is, even those studies which intended to investigate the reliability of the self and peer assessments actually seem to be studies of validity (Topping, 1998). Respectively, very few studies have ever tried to study the reliability of self and peer assessments in terms of the consistency of the results throughout time (e.g. Marcoulides and Simkin, 1995). Sung, Chang, Chiou, and Hou (2005) is one of the studies in this regard who found evidence for the fact that self

assessment is consistent over a short time frame. Formerly, Blatchford (1997) had found that self assessments were not stable between ages 11 and 16 in English classes.

Research questions and hypotheses

The major objective of the present study is to investigate the reliability of self and peer assessments over time in the higher educational context of Iran. It also tries to find the possible relationship between peer and self assessment. As such, the study is after the following research questions:

1. Is there any significant relationship between self and peer assessment scores?
2. Does self assessment enjoy reliability over time?
3. Does peer assessment enjoy reliability over time?

Method

Participants

Thirty sophomores majoring in TEFL participated in this study. They had passed Conversation Courses I and II during the first year of their study at the university and now they were attending the Conversation Course III in two separate classes with 15 students in each class. The course lasted for one academic semester (16 weeks), for 4 hours per week.

Instrumentation

A holistic rating scale of speaking (ILR Language Skill Level Descriptions) was utilized to assess the learners' speaking skill. This scale specifies six levels of proficiency running from 0 to 5. Zero stands for no/memorized proficiency, 1 for elementary proficiency, 2 for limited working proficiency, 3 for general professional proficiency, 4 for advanced professional proficiency, and 5 for functionally native proficiency. The scale levels are explained in detail in the appendix**.**

Data collection procedures

At the beginning of the term students were taught how to assess their peers' speaking ability using the speaking rating scale (see the appendix). During the term they had due practice in assessing their peers' and their own speaking ability based on their performance on different tasks such as lectures, panel discussions, oral reproduction of stories, etc. Problems concerning self and peer assessments were presented and discussed with the whole class to make sure that the teacher and all the students had the same idea of what constituted oral proficiency and consequently assessed the same elements in this regard. Then during the 8th and 9th weeks students were asked to rate their classmates' as well as their own speaking ability and report a score (form 0 to 5) in this regard. Each student was supposed to assign scores to all of his classmates' as well as to his own speaking ability. As such, each student received 15 scores for his speaking ability, one obtained through self assessment and 14 others from peer assessment. The mean of these fourteen scores was reported as the peer assessment score. The same procedure was employed at the end of the term (weeks 15 and 16) during which each student was asked to assign scores to his own speaking ability as well as others'. The reason that all the students rather than some of them were asked to rate their peers' performance was based on the idea that if individual students are poor judges, the reliability of averaged scores can be increased by increasing the number of raters (Cho, Schunn, and Wilson, 2006; Fagot, 1991; Ferguston, 1966; Magin, 1993; Xiao and Lucking, 2008).

Results

The data collected were subjected to a number of correlation coefficients to determine the degree of relationship between sets of scores obtained from self and peer assessments. First of all Pearson correlation coefficient was used to determine the degree of relationship between self and peer assessments in the middle of the term (weeks 8 and 9). Table 1 depicts the statistics in this regard.

Table 1
Correlation Coefficient for Self/peer Assessments in Midterm

Assessment Type	N	Mean	Std.	Pearson C	Sig
Self Assessment	30	3.00	0.64	00.38	0.05
Peer Assessment	30	2.18	0.77		

As indicated in this table, peer and self assessments are significantly related at $p < 0.05$. Of course caution should be applied in using the results because although the two assessments are significantly related, the correlation is not high enough ($r = 0.38$) to lead us to safe conclusions. Usually correlation coefficients below 0.5 are considered to be low (Ary, Jacobs, and Razavieh, 1996).

As for correlation at the end of the term (weeks 15 and 16), Table 2 below indicates that self and peer assessments are not significantly related. Hence concerning the first question of the study that asked for the relationship between self and peer assessment, the results indicated that there is a significant (though weak) relationship between the two assessment procedures in midterm and that there is no significant relationship between them in the final exam.

Table 2
Correlation Coefficient for Self/Peer Assessments at the End of the Term

Assessment Type	N	Mean	Std.	Pearson C	Sig
Self Assessment	30	3.47	0.86	00.31	0.09
Peer Assessment	30	2.15	0.71		

The same procedure was employed to see whether the scores obtained from self and peer assessments indicated consistency over time. The results are depicted in Tables 3 and 4 below.

Table 3
Correlation Coefficient for Self Assessments

Assessment Type	N	Mean	Std.	Pearson C	Sig
Midterm Self Assessment	30	3.00	0.64	00.68	0.000
Final Self Assessment	30	3.47	0.86		

Table 3 depicts a significant correlation between the two self assessments carried out in the middle and at the end of the term. Hence the answer to the second research question is positive meaning that the scores obtained from self assessment in midterm significantly correlate with those obtained in the final exam.

The third research question asked for the consistency of peer assessment scores over time. Here again Pearson correlation coefficient was utilized to answer the question. The result is depicted in Table 4.

Table 4
Correlation Coefficient for Peer Assessments

Assessment Type	N	Mean	Std.	Pearson C	Sig
Midterm Peer Assessment	30	2.18	0.77	00.90	0.000
Final Peer Assessment	30	2.15	0.71		

This table indicates that peer assessments are significantly and meaningfully related in the middle and at the end of the term ($r = 0.90$). However the results obtained from peer assessments are more satisfactory than those of self assessments. In other words, the scores obtained from peer assessments indicate a higher level of consistency ($r = 0.90$) over time than the scores obtained from self assessments ($r = 0.68$).

Discussion

This study focused on the consistency of self and peer assessments over time. The results indicated that both self and peer reviewers were

consistent evaluators. Peer assessors, however, were more consistent. In other words, the results indicated that Peer assessment enjoyed a higher consistency as an alternative for teacher assessment. That is, the correlation obtained for peer assessments within an interval of 6 weeks was quite high (r = 0.90) which was considerably higher than the one obtained for self assessments (r = 0.68) though in both cases the correlation was significant. This seems to be logical since in this study the peer score given to a student's performance was defined as the mean of 14 scores given by all the other students in class. Hence such a peer score undergoes less fluctuation throughout time and tends to be more consistent and reliable than the score given by a single student through self assessment procedure. This is in line with the idea that an individual reviewer is less consistent in his evaluation than multiple raters (Cho, Schunn, and Wilson 2006; Fagot, 1991; Ferguston, 1966; Magin, 1993; Xiao and Lucking, 2008). Furthermore, "students may be either too harsh on themselves or too self-flattering" (Brown, 2004, p. 270) and this can lead to inaccuracy or inconsistency of self assessment scores.

The study also checked for the degree of relationship between self and peer assessments. It was shown that the two types of assessment were not significantly related at the end of the term. As for the midterm, although they were significantly related, the correlation was not high enough to be considered meaningful (r = 0.38). Thus overall one might say that self and peer assessments were loosely related in this study. Formerly, a number of studies had indicated a high correlation between peer assessment and teacher assessment (e.g. Alfallay, 2004; Bachman and Palmer, 1989; Falchikov and Goldnch, 2000; Patri, 2002; Ross, 1998; Stefani, 1994; Sullivan and Hall, 1997; Topping, 1998; Williams, 1992) and concluded that peer assessment could be a valid procedure to be used as an alternative for teacher assessment. This study indicated that peer assessment also enjoyed more consistency over time. This may give us a clue as to the preferability of peer assessment over self assessment as an alternative for teacher assessment since it enjoys a higher degree of reliability (leading to more consistent results over time), and validity (having a high correlation with teacher assessment).

It is worth mentioning here that the success of the peer and self assessments depends to a large extent to how they are managed in the classroom. That is, much valuable class time may be lost if alternative assessment is not employed properly (Wheater, Langan, and Dunleavy, 2005). There could also be other problems like inaccuracy and low precision by naive markers, and variability between groups of peer-assessors (Swanson et al., 1991).

Conclusion

This study indicated that self and peer assessment scores do not necessarily correlate with each other. The results also indicated that peer assessment as an alternative for teacher assessment is more consistent over time than self assessment. In other words, peer assessment enjoys a higher reliability than self assessment and could lead us to more dependable results. However this does not necessarily mean that we should exclude self assessment from or include peer assessment into our academic programs. Each of these assessment types may have its own benefits to the students and teachers if employed at the right time (and for the right language area). As such, any decisions as to whether exclude or include self and peer assessments should be made after due consideration of their positive impacts (Cheng and Warren, 2005).

Implications of the research

A growing number of studies support the use of peer and self assessments as alternatives for teacher assessment (e.g. Falchikov, 1995; Hughes, 2001; Magin and Helmore, 2001). However this study cast doubts over such an assumption because the results indicated that peer and self assessment were not significantly related in the final exam and only loosely, though significantly, related in the midterm exam. This means that using self or peer assessment may lead to different results. This calls for more caution in employing self and peer assessments instead of teacher assessment.

The study also indicated that the results of both self and peer assessment enjoyed consistency over time, however peer assessment

indicated a higher consistency in this regard. This can lead to the idea that if we are to use an alternative for teacher assessment, then peer assessment could be a more logical option.

Further research

The following may be implied for the future line of research: First, this study investigated the consistency of the scores obtained through self and peer assessments from a group of sophomores. It is therefore worth investigating whether the same results are also true with students of different proficiency levels. Second, this study didn't consider the probable effect of gender on the consistency of the scores. It would be worth to design self and peer assessment studies investigating the consistency of the results under the effect of gender. Third, this study focused on assessing oral proficiency using a holistic rating scale. The same study could be replicated using analytic rating scales or focusing on other language skills and content areas. It has been indicated that students are better assessors in content areas when the specific course objectives are available and are less adept at assessing general proficiency (Ross, 1998). And finally a question which is still open to research and offers a variety of potentials for assessment studies is: *under what conditions and for what language or content areas should we limit the use of self and peer assessments?*

Received 10 February, 2009
Accepted 15 August, 2009

References

AlFallay, I. (2001). Motivation type and level of language proficiency: two crucial factors in self assessment, *Arab Journal for the Humanities*, 75, 313-32.

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, 407-25.

Ary, D., Jacobs, L. C., & Razavieh, A. (1996). *Introduction to research in education (5th ed.)*. Fort Worth: Harcourt Brace College Publishers.

Bachman, L. & Palmer, A. (1989). The construct validation of self ratings of communicative language ability, *Language Testing,* 6, 14-29.

Bailey, K. M. (1998). *Learning about language assessment: dilemmas, decisions, and directions*. Cambridge, MA: Heinle & Heinle.

Berg, I. V. D., Admiraal, W., & Pilot, A. (2006). Peer assessment in university teaching: evaluating seven course designs, *Assessment and Evaluation in Higher Education*, 31(1), 19-36.

Black, P. & Wiliam, D. (1998). Assessment and classroom learning, *Assessment in Education*, 5(1), 7-74.

Blatchford, P. (1997). Students' self-assessment of academic attainment: accuracy and stability from 7 to 16 years and influence of domain and social comparison group, *Educational Psychology*, 17 (3), 345-60.

Blue, G. (1988). Self-assessment of listening comprehension, *International Review of Applied Linguistics*, 16, 149-56.

Boud, D. (1989). The role of self-assessment in student grading, *Assessment and Evaluation in Higher Education*, 14, 20-30.

Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.

Boud, D., Cohen, R., & Sampson, J. (1999). Peer learning and assessment, *Assessment and Evaluation in Higher Education*, 24(4), 413-26.

Bradley, C. (1993). Sex bias in student assessment overlooked? *Assessment and Evaluation in Higher Education,* 18, 1-8.

Brown, H. D. (2004). *Language assessment: principles and classroom practices*. New York: Pearson Education, Inc.

Brown, J. D. & Hudson, T. (1998). The alternatives in language assessment, *TESOL Quarterly*, 32, 653-75.

Chaudron, C. (1983). Evaluating writing: Effects of feedback on revision. *Paper presented at the 17 Annual TESOL Convention, Toronto*, *Ontario, Canada.*

Cheng, W. & Warren, M. (1999). Peer and teacher assessment of the oral and written tasks of a group project, *Assessment and Evaluation in Higher Education*, 24(3), 301-14.

Cheng, W. & Warren, M. (2005). Peer assessment of language proficiency, *Language Testing*, 22(1), 93-121.

Cho, K., Schunn, C. D., & Wilson, W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives, *Journal of Educational Psychology*, 98(4), 891-901.

Davidson, F. & Henning, G. (1985). A self-rating scale of English difficulty, *Language Testing*, 2, 164-69.

Davies, P. (2006). Peer assessment: judging the quality of students' work by comments rather than marks, *Innovations in Education and Teaching International*, 43(1), 69-82.

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, 331-50.

Fagot, R. (1991). Reliability of ratings for multiple judges: intraclass correlation and metric sales, *Applied Psychological Measurement*, 15(1), 1-11.

Falchikov, N. (1995). Peer feedback marking: developing peer assessment, *Innovations in Education and Training International*, 32(2), 175-87.

Falchikov, N. & Goldnch, J. (2000). Student peer assessment in higher education: a meta-analysis comparing peer and teacher marks, *Review of Educational Research*, 70, 287-322.

Falchikov, N. & Magin, D. (1997). Detecting gender bias in peer marking of students group process work, *Assessment and Evaluation in Higher Education,* 22, 385-96.

Ferguson, G. A. (1966). *Statistical analysis in psychology and education (2nd ed*.). New York: McGraw Hill.

Fletcher, C. & Baldry, C. (2000). A study of individual differences and self-awareness in the context of multi-source feedback, *Journal of Occupational and Organizational Psychology*, 73, 303-19.

Freeman, M. & McKenzie, J. (2002). SPARK, a confidential web-based template for self and peer assessment of student teamwork: benefits of evaluating across different subjects, *British Journal of Educational Technology*, 33, 551-69.

Guilford, W. H. (2001). Teaching peer review and the process of scientific writing, *Innovations and Ideas*, 25(3), 167-75.

Haaga, D. A. F. (1993). Peer review of term papers in graduate psychology course, *Teaching of Psychology*, 20(1), 28-32.

Heilenmann, K. (1990). Self-assessment of second language ability: the role of response effects, *Language Testing*, 7, 174-201.

Hughes, I. (2001) But isn't this what you're paid for? The pros and cons of peer-and self-assessment. *Planet*, National Subject Centre for Geography, Earth and Environmental Sciences, *Learning and Teaching Support Network*, Issue 2, 20-23.

Hughes, I. & Large, B. (1993). Staff and peer group assessment of oral communication skills, *Studies in Higher Education*, 18, 379-85.

Interagency Language Roundtable. Interagency Language Roundtable Language Skill Level Descriptions, Speaking.
Retrieved on 20 July 2008 from: *http://www.govtilr.org /Skills/ILRscale2.htm*

Janssen-van Dieten, A. (1989). The development of a attest of Dutch as a second language: the validity of self-assessment by inexperienced subjects, *Language Testing*, 6, 30-46.

Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis, *Review of Educational Research*, 75(1), 63-82.

Langan, A. M. & Wheater, C. P. (2003).Can students assess students effectively? Some insights into peer-assessment, *Learning and Teaching in ACTION*, 2, 9-13.

Lin, S. S. J., Liu, E. Z. F., & Yuan, S. M. (2001). Web-based peer assessment: feedback for students with various thinking-styles, *Journal of Computer Assisted Learning*, 17, 420-32.

Little, D. (2005). The common European framework and the European language portfolio: involving learners and their judgments in the assessment process, *Language Testing*, 22 (3), 321-36.

Lu, R. L. & Bol, L. (2007). A comparison of anonymous versus identifiable e-peer review on college student writing performance and the extent of critical feedback, *Journal of Interactive Online Learning*, 6(2), 100-15.

MacLeod, L. (1999). Computer-aided peer review of writing, *Business Communication Quarterly*, 62(3), 87-94.

Magin, D. & Helmore, P. (2001) Peer and teacher assessments of oral presentations: how reliable are they? *Studies in Higher Education,* 26, 287-98.

Magin, D. (1993). Should student peer ratings be used as part of summative assessment? *Higher Education Research and Development*, 16, 537-42.

Marcoulides, G. A. & Simkin, M. G. (1995). The consistency of peer review in student writing projects, *Journal of Education for Business*, 70, 220-23.

McIsasc, C. M. & Sepe, J. F. (1996). Improving the writing of accounting students: A cooperative venture, *Journal of Accounting Education*, 14(4), 515-33.

Mowl, G. & Pain, R. (1995). Using self and peer assessment to improve students' essay writing A case study from geography, *Innovations in Education and Training International*, 32(4), 324-35.

Newstead, S. & Dennis, I. (1990). Blind marking and sex bias in student assessment, *Assessment and Evaluation in Higher Education,* 15, 132-39.

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, 23-38.

Oscarson, M. (1989). Self-assessment of language proficiency: rationale and applications, *Language testing*, 6, 1-13.

Patri, M. (2002). The influence of peer feedback on self- and peer-assessment of oral skills, *Language Testing,* 19, 109-31.

Paulus, T. M. (1999). The effect of peer and teacher feedback on student writing, *Journal of Second Language Writing*, 8(3), 265-89.

Plutsky, S. & Wilson, B. A. (2004). Comparison of the three methods for teaching and evaluating writing: A quasi-experimental study, *The Delta Pi Epsilon Journal*, 46(1), 50-61.

Rada, R. & Hu, K. (2002). Patterns in student commenting, *IEEE Transactions on Education*, 45, 262-67.

Richer, D. L. (1992). The effects of two feedback systems on first year college student writing proficiency, *Dissertation Abstracts International*, 53, p. 2722.

Ross, S. (1998). Self assessment in second language testing: a meta-analysis and analysis of experiential factors*, Language Testing,* 15 (1), 1-20.

Saito, H. & Fujita, T. (2004). Characteristics and user acceptance of peer rating in EFL writing classrooms, *Language Teaching Research,* 8(1), 31-54.

Sluijsmans, D., Brand-Gruwel, S., & Van Merriënboer, J. J. G. (2002). Peer assessment training in teacher education: Effects on performance and perceptions, *Assessment and Education in Higher Education*, 27(5), 443-54.

Sluijsmans, D., Dochy, F., & Moerkerke, G. (1999). Creating a learning environment by using self-, peer- and co-assessment, *Learning Environment Research*, 1, 293-319.

Smith, H., Cooper, A., & Lancaster, L. (2002). Improving the quality of undergraduate peer assessment: a case study from psychology, *Innovations in Education and Teaching International*, 39, 71-81.

Stefani, L. A. J. (1994). Peer, self and tutor assessment: relative reliabilities, *Studies in Higher Education*, 19(1), 69-75.

Stefani, L. A. J. (1998). Assessment in partnership with learners, *Assessment and Evaluation in Higher Education*, 23(4), 339-50.

Sullivan, K. & Hall, C. (1997). Introducing students to self-assessment, *Assessment and Evaluation in Higher Education,* 22, 289-305.

Sung, Y. T., Chang, K. E., Chiou, S. K., & Hou, H.T. (2005). The design and application of a web-based self- and peer-assessment system, *Computers and Education*, 45(2), 187-202.

Swanson, D., Case, S., & van der Vlueten, C. (1991). Strategies for student assessment. In D. Boud & G. Feletti (Eds.), *The Challenge of Problem Based Learning* (pp 260-273). Kogan Page: London.

Topping, K. (1998). Peer assessment between students in colleges and universities, *Review of Educational Research*, 68 (3), 249-76.

Topping, K., Smith, F. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students, *Assessment and Evaluation in Higher Education*, 25(2), 149-69.

Tseng, S. C. & Tsai, C. C. (2007). On-line peer assessment and the role of the peer feedback: a study of high school computer course, *Computers and Education*, 49, 1161-74.

Tudor, I. (1996). *Learner-centeredness as language education*. Cambridge: Cambridge University Press.

Venables, A. & Summit, R. (2003). Enhancing scientific essay writing using peer assessment, *Innovations in Education and Teaching International*, 40(3), 281-90.

Wheater, P. C., Langan, M. A., & Dunleavy, P. J. (2005). Students assessing student: case studies on peer assessment, *Planet*, 15, 13-15.

Wesche, M., Morrison, F., Ready, D., & Pawley, C. (1990). French immersion: postsecondary consequence for individuals and universities, *Modern Canadian Language Review*, 46, 430-51.

Williams, E. (1992). Student attitude towards approaches to learning and assessment, *Assessment and Evaluation in Higher Education,* 17, 45-58.

Woolhouse, M. (1999). Peer assessment: the participants' perception of two activities on a further education teacher education course, *Journal of Further and Higher Education*, 23, 211-19.

Xiao, Y. & Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki

environment, *The Internet and Higher Education*, doi: 10.1016/j.iheduc.2008.06.005

Zhao, Y. (1998). The effects of anonymity on computer-mediated peer review, *International Journal of Educational Telecommunication*, 4(4), 311-45.

Appendix[†]

A: Oral presentation assessment criteria

0. No/memorized proficiency
1. Elementary proficiency
2. Limited working proficiency
3. General professional proficiency
4. Advanced professional proficiency
5. Functionally native proficiency

Speaking 0 (No Proficiency / Memorized Proficiency): Unable to function in the spoken language. Oral production is limited to occasional isolated words. Has essentially no communicative ability. Or Able to satisfy immediate needs using rehearsed utterances. Shows little real autonomy of expression, flexibility or spontaneity. Can ask questions or make statements with reasonable accuracy only with memorized utterances or formulae. Attempts at creating speech are usually unsuccessful.

Speaking 1 (Elementary Proficiency): This speaker has a functional, but limited proficiency. He is able to satisfy minimum courtesy requirements and maintain very simple face-to-face conversations on familiar topics. A native speaker must often use slowed speech, repetition, paraphrase, or a combination of these to be understood by this individual. Similarly, the native speaker must strain and employ real-world knowledge to understand even simple statements/questions from this individual. Misunderstandings are frequent, but the individual is able to ask for help and to verify comprehension of native speech in face-to-face interaction. The individual is unable to produce continuous discourse except with rehearsed material.

[†] **Adapted from Interagency Language Roundtable Language Skill Level Description: Speaking. Retrieved from http://www.govtilr.org/Skills/ILRscale2.htm**

Speaking 2 (Limited Working Proficiency): Able to satisfy routine social demands and limited work requirements. Can handle routine work-related interactions that are limited in scope. In more complex and sophisticated work-related tasks, language usage generally disturbs the native speaker. Can handle with confidence, but not with facility, most normal, high-frequency social conversational situations including extensive, but casual conversations about current events, as well as work, family, and autobiographical information. The individual can get the gist of most everyday conversations but has some difficulty understanding native speakers in situations that require specialized or sophisticated knowledge. The individual's utterances are minimally cohesive. Linguistic structure is usually not very elaborate and not thoroughly controlled; errors are frequent. Vocabulary use is appropriate for high-frequency utterances. but unusual or imprecise elsewhere.

Speaking 3 (General Professional Proficiency): Able to speak the language with sufficient structural accuracy and vocabulary to participate effectively in most formal and informal conversations in practical, social and professional topics. Nevertheless, the individual's limitations generally restrict the professional contexts of language use to matters of shared knowledge and/or international convention. Discourse is cohesive. The individual uses the language acceptably, but with some noticeable imperfections; yet, errors virtually never interfere with understanding and rarely disturb the native speaker. The individual can effectively combine structure and vocabulary to convey his/her meaning accurately. The individual speaks readily and fills pauses suitably. In face-to-face conversation with natives speaking the standard dialect at a normal rate of speech, comprehension is quite complete. Although cultural references, proverbs and the implications of nuances and idiom may not be fully understood, the individual can easily repair the conversation. Pronunciation may be obviously foreign. Individual sounds are accurate: but stress, intonation and pitch control may be faulty.

Speaking 4 (Advanced Professional Proficiency): Able to use the language fluently and accurately on all levels normally pertinent to professional needs. The individual's language usage and ability to function are fully successful. Organizes discourse well, using appropriate rhetorical speech devices, native cultural references and understanding. Language ability only rarely hinders him/her in performing any task requiring language; yet, the individual would seldom be perceived as a native. Speaks effortlessly and smoothly and is able to use the language with a high degree of effectiveness, reliability and precision for all representational purposes within the range of personal and professional experience and scope of responsibilities. Can serve as in informal interpreter in a range of unpredictable circumstances. Can perform extensive, sophisticated language tasks, encompassing most matters of interest to well-educated native speakers, including tasks which do not bear directly on a professional specialty.

Speaking 5 (Functionally Native Proficiency): Speaking proficiency is functionally equivalent to that of a highly articulate well-educated native speaker and reflects the cultural standards of the country where the language is natively spoken. The individual uses the language with complete flexibility and intuition, so that speech on all levels is fully accepted by well-educated native speakers in all of its features, including breadth of vocabulary and idiom, colloquialisms and pertinent cultural references. Pronunciation is typically consistent with that of well-educated native speakers of a non-stigmatized dialect.

B: Tables

Table 1
Correlation Coefficient for Self/peer Assessments in Midterm

Assessment Type	N	Mean	Std.	Pearson C	Sig
Self Assessment	30	3.00	0.64	00.38	0.05
Peer Assessment	30	2.18	0.77		

Table 2
Correlation Coefficient for Self/Peer Assessments at the End of the Term

Assessment Type	N	Mean	Std.	Pearson C	Sig
Self Assessment	30	3.47	0.86	00.31	0.09
Peer Assessment	30	2.15	0.71		

Table 3
Correlation Coefficient for Self Assessments

Assessment Type	N	Mean	Std.	Pearson C	Sig
Midterm Self Assessment	30	3.00	0.64	00.68	0.000
Final Self Assessment	30	3.47	0.86		

Table 4
Correlation Coefficient for Peer Assessments

Assessment Type	N	Mean	Std.	Pearson C	Sig
Midterm Peer Assessment	30	2.18	0.77	00.90	0.000
Final Peer Assessment	30	2.15	0.71		