An Exploration of the Reliability and Validity of Peer Assessment of Writing in Secondary Education

Elena Meletiadou and Dina Tsagari

University of Cyprus

eleni.meletiadou@gmail.com, dinatsa@ucy.ac.cy

Abstract

The current study investigates the reliability and validity of peer assessment (PA) of writing in secondary education in Cyprus since recent research provides few and mixed findings in this area (Sung, Chang, Chang & Yu 2010). Forty EFL students were involved in the PA of writing after receiving training. The results showed that the correlation between teacher and student marks was very high and that: (a) clear student-generated criteria, (b) adaptation of the instruments to meet the specific students' needs, (c) careful training, (d) guidance and continuous support can enhance the reliability and validity of PA of writing in the state school sector.

1. Introduction

Peer assessment (PA) is an educational arrangement where students judge a peer's performance quantitatively, e.g. by providing a peer with scores or grades, and/or qualitatively, e.g. by providing a peer with written or oral feedback (Topping 1998).

Strong justification for the use of PA is found in four theoretical stances: process writing, collaborative learning, Vygotskian learning theory and interactionist theories of L2 acquisition (Hansen & Liu 2005). Moreover, numerous studies underscore the role and value of PA in TESOL writing instruction, in terms of developing the learners' writing ability, writing performance and autonomy in learning (McIsaac & Sepe 1996; Plutsky & Wilson 2004).

Generally speaking, PA is said to maximize success by providing a number of benefits to learners such as:

- improvements in the effectiveness and quality of learning, at least as good as gains from teacher assessment, especially in relation to writing (Topping, Smith, Swanson & Elliot 2000);

- development of self-reliant and self-directed learners (Oldfield & Macalpine 1995);
- social, cognitive, affective and methodological benefits (Villamil & De Guerrero 1996), and
- creation of a strong link between instruction and assessment by forming part of a feedback loop that enables teachers to monitor and modify instruction according to results of student assessment (Tsagari 2004).

In summary, PA as a means of classroom assessment is student-centred. It allows learners to participate in the process of evaluation and provides them with rich opportunities for observation and modeling (Cho, Schunn & Wilson 2006).

2. Peer Assessment, Reliability and Validity

While PA is highly recommended by researchers (Shepard 2000; Topping 1998), the issue of the reliability and validity of PA needs clarification. A number of researchers have reported high correlations between student- and teacher-assessments (e.g. Rudy, Feijfar, Griffith & Wilson 2001; Pope 2005). For instance, Haaga (1993) and Cho et al. (2006) also reported relatively high reliability. Marcoulides and Simkin (1995) found that peer reviewers were consistent evaluators while Falchikov and Goldfinch (2000) after reviewing 48 studies concluded that peer ratings were highly correlated with teacher ratings.

However, there appear to be mixed findings related to the reliability and validity of PA. Some studies have shown low correlations between teacher and peer ratings (Cheng & Warren 1999; Swanson, Case & Van der Vleuten 1997). The question whether PA can be used as part of formal classroom assessment has also been a point of contention (Goldfinch & Raeside 1990). Moreover, judging dimensions such as appropriacy, fluency and clarity are subjective and are likely to affect the validity of assessments provided (Orsmond, Merry & Reiling 1997). A number of biases associated with marking and PA such as friendship bias (1997). A number of biases associated with marking and PA such as friendship bias, e.g. students tend to provide extra marks to their friends (Fletcher & Baldry 1999) might also operate.

The diverse validity and reliability results raise doubts in both teachers and learners who are often reluctant to use PA (Sluijsmans, Moerkerke, & van Merriënboer 2001; Orsmond et al. 1997). However, teacher marking has also been found to be problematic e.g. unreliable, inconsistent and/or biased (Falchikov & Magin 1997).

What one needs to keep in mind when employing PA minimize these biases and psychological resistance through various administrative strategies such as setting clear criteria, understanding goals and limits, and developing familiarity with the instruments (i.e. analytic rating scale) used in PA (Chapelle & Brindley 2002). Research has shown that, when assessment criteria are firmly set, the

reliability and validity of PA is enhanced since it enables students to judge the performance of their peers in a manner comparable to those of the teachers (Falchikov 2005).

3. Peer Assessment and Young Learners

A second concern of the present study is to determine the best type of participant in PA. Very little research has been conducted in the area of PA performed by adolescent learners (Tsivitanidou, Zacharia & Hovardas 2011). Although some studies consider PA as suitable for young learners (Shepard 2000), some researchers claim that PA is more suitable for older learners (Falchikov & Boud 1989; Jones & Fletcher 2002) and in-service staff (Jones & Fletcher 2002; Saavedra & Kwun 1993). Brown and Dove (1990) claim that it is a rather demanding method of assessment. Finally, significant differences between characteristics of adolescents and adults suggest that studies should specifically investigate whether PA is suitable for younger learners (Sung et al. 2010).

However, Oskay, Schallies and Morgil (2008) confirmed the suitability of PA for different settings in terms of educational levels and fields of study. Previous research in the Cypriot context has also indicated that secondary school students have 'the beginnings of PA skills and an understanding of what needs to be included in feedback' (Tsivitanidou et al. 2011: 517). Nevertheless, to implement PA effectively, students need explicit training in PA skills and techniques (Boud 1990; Hanrahan & Isaacs 2001; Sluijsmans 2002; Van Steendam, Rijlaarsdam, Sercu, & Van den Bergh 2010).

As a result, additional research needs to be conducted to investigate whether PA is suitable for adolescent learners. This study will expand on previous studies by investigating the reliability and validity of PA of writing with adolescent EFL learners.

4. The Educational and Assessment Context of the Current Study

Other than state schools, many Cypriot adolescent students learn English as a foreign language in State Institutes. These are situated in selected Cypriot high schools and run by the Cypriot Ministry of Education.

The curriculum of the State Institutes (Ministry of Education 2010: 4-6) promotes communicative and learner-centred language teaching and stresses the importance of informal assessment. Teachers are also advised to consider various types of alternative assessment such as self-assessment and portfolios (ibid: 16-18). However, despite the benefits of PA discussed in the literature (see Section 1), there is no reference to PA as such in the curriculum.

Moreover, according to local inspectors, teachers and headteachers of State Institutes, the assessment of EFL writing in such institutes is generally problematic. Apart from the fact that it is teacher-centered, no further guidelines, special seminars or training are provided. Adolescent learners continue to make the same mistakes, become more and more reliant on the teacher, and their writing does not necessarily improve as a result of teacher feedback. Consequently, the majority of these learners have poor writing skills, a negative attitude towards writing and the assessment of writing and face considerable problems in formal tests. Therefore, the Cypriot Ministry of Education needs to find new ways to improve the teaching and learning and in particular the assessment of EFL writing skills.

To respond to the relevant gaps in the literature (Sections 2 & 3), the problems regarding the assessment of writing skills in the Cypriot educational context and the lack of any research into PA of EFL writing in the particular context, at least to the knowledge of the authors of this paper, the present study set out to investigate PA of EFL writing of adolescent students. Two research questions guided this study:

- Is peer assessment of writing a reliable and valid assessment method?
- Can peer assessment of writing be employed successfully with adolescent learners?

5. The Design and Methodology of the Study

In January 2010, a study was conducted in order to explore the issue of reliability of PA and its suitability for adolescent learners. It extended over a period of four months. The time schedule of the study is seen in Table 1 below.

Table 1.
Time Schedule

January 2010	Week 2: Training of groups and the external assessor. Week 3: Piloting of instruments. Week 4: <i>Writing Task 1</i> - Writing the first draft of a narrative essay.
February 2010	Week 1: Feedback and remedial teaching. Week 2: Writing the second draft. Week 3: <i>Writing Task 2</i> - Feedback. Whole-class discussion. Writing the first draft of a descriptive essay. Week 4: Feedback and remedial teaching.
March 2010	Week 1: Writing the second draft. Week 2: <i>Writing Task 3</i> - Feedback. Whole-class discussion. Writing the first draft of an informal letter. Week 3: Feedback and remedial teaching. Week 4: Easter holiday.
April 2010	Week 1: Easter holiday. Week 2: Writing the second draft. Week 3: Feedback. Whole-class discussion.

The participants were forty 13-14 year old students. They had been attending EFL classes at a State Language Institute in Nicosia, the capital of the country, for the past five years. They were all native Greek-Cypriots and shared similar cultural and socio-economic background. Their teacher and an external assessor also took part in the study. They were both qualified EFL teachers with several years of experience and postgraduate degrees in TEFL. The teacher, who was also one of the researchers of this study, taught and trained the students and the external assessor in PA methods.

The learners were divided into two groups of twenty students (Table 2) and wrote three types of essays (a narrative essay, a descriptive essay and an informal letter), as part of the demands of their writing curriculum (Ministry of Education 2010: 27). The essays were submitted twice. Teacher and peer feedback were provided to students with a view to improving successive drafts and prompting more revision of the written essays.

Table 2.
Grouping of learners

Group B (Experimental group 1-Student/assessees)	Received teacher and peer feedback from group C
Group C (Experimental group 2-Student/assessors)	Received teacher feedback and provided peer feedback to group B

The aim was to introduce PA and examine its reliability by comparing teacher marks with student/assessors' marks. Anonymity of student/assessors ensured the reliability of the assessment process and assisted in avoiding conflicts and bitterness among the learners (Miller & Ng 1994). The only thing that varied was the type of feedback provided to students. Overall, learners were provided with:

- teacher feedback (grade and comments) for each draft
- teacher corrections only to second drafts so as not to interfere with the revision process, and
- one PA form (filled in by student/assessors - Group C - and given to student/assessees – Group B) for all drafts (Appendix I).

The external assessor also provided marks only for the second drafts of all essays to check the reliability of the teacher's marks.

6. The Training of the Learners and the External Assessor

Supporting learners in using peer-assessment is of paramount importance because PA is an activity for which learners need guidance and time to grow into. In other words, learners needed to build up a shared understanding of the nature,

the purposes and the requirements of the PA method (Stewart & Cheung 1989). According to previous research (Sluismans 2002; Van Steendam et al. 2010), substantial training and practice is required to develop competence in PA.

With this background in mind, a PA training session which lasted three teaching hours was offered to the students in January 2010 (Table 1). During the session, student/assessors designed a PA form (Appendix I) with the help of the teacher. The form was used by the teacher, external assessor and students to provide feedback for all drafts of all essays. The researchers decided to involve learners in the creation of the assessment criteria because, according to previous research (Falchikov & Goldfinch, 2000), this ensures students' active engagement in the PA procedure. Moreover, the research literature (Pond, Ul-haqa, & Wadea 1995; Stefani 1994) has shown that the reliability of PA which is supported by assessment rubric tables, checklists (cf. the present PA form in Appendix I), exemplification, teacher assistance and monitoring tends to be high.

Overall, the main purpose of the training session was to offer an introduction in PA methods. Student/assessees were taught how to revise their work using the PA form since they had never been asked to re-draft their work before. Student/assessors were also trained by rating and commenting on sample essays. Finally, the external assessor also received training (40 min) in relation to the assessment criteria employed before using them due to lack of any previous experience in PA. During the training, the external assessor also rated students' drafts using the PA form. The aim was to deepen understanding of the assessment criteria included in the form through exemplification and increase the reliability of the marks provided by the external assessor.

7. Findings

To answer the research questions, the researchers correlated teacher and student marks for each draft of all compositions in order to check the reliability of student-generated scores. A two-sided Pearson correlation test was conducted which produced the results presented in Table 3.

Table 3.
Reliability indices of student-generated scores

2-sided Pearson correlation test	T	f	p-value	Cor	P
Teacher & student/assessors' marks for the first draft of all essays	5.30	8	2.2e-16	0.89	<0.001
Teacher & student/assessors' marks for the second draft of all essays	.71	8	8.837e-14	0.78	<0.001

The correlation between teacher and student/assessors' marks for the first draft of all essays was very high (0.89, $p < 0.001$). The same kind of test was conducted with the marks of the second draft and the correlation was almost as high as the previous one (0.78, $p < 0.001$).

Moreover, in order to further explore the reliability of PA, the researchers compared teacher marks with those of the external assessor. The Pearson correlation test conducted (Table 4) indicated a very high correlation (0.93, $p < 0.001$) between teacher and external assessor's marks for the second drafts of all students' compositions. This confirmed that the teacher was not biased towards any of the groups.

Table 4.

Comparison of teacher and external assessor marks

Pearson correlation test	T	Df	p-value	Cor	P
Teacher and external assessor marks	34.35	178	< 2.2e-16	0.93	<0.001

8. Summary and Discussion of Results

The present study has shown high correlations between teacher, who is always considered as the expert (Falchikov 1995), and student marks (Table 3). This is evidence that PA can be a reliable and, valid assessment method.

More specifically the use of negotiated joint construction of the assessment criteria used for PA combined with the training of the learners and the external assessor, the use of checklists and exemplification during the training sessions deepened understanding of the assessment procedure, gave a greater sense of ownership to students, and increased the reliability of the PA method (also argued in MacArthur, Schwartz & Graham 1991). The above mentioned measures enabled the adolescents involved in the current study to provide reliable marks, that is marks close to the teacher marks (Table 3). These results also confirm previous research that has been conducted with adolescent learners (Sadler & Good 2006; Saito & Fujita 2004; Cho & MacArthur 2010) and older learners (Patri 2002; Rudy & Fejfar et al. 2001; Falchikov & Goldfinch 2000).

Careful training of the external assessor also increased rater consistency (Table 4). This implies that well-trained and experienced teachers can produce consistent marks. This might be extremely helpful to future researchers who attempt to use PA with larger populations and more teachers. Only when teachers provide consistent marks, can researchers test the reliability and validity of PA (by comparing teachers' and students' marks) effectively.

Additionally, the present study suggested the use of PA as a complementary practice to teacher assessment. This has also been supported by previous researchers (i.e. Tsui & Ng 2000). Substituting teacher assessment with PA would be daunting since the teacher is always the expert in the learners' eyes (Berg, 1999).

The researchers also avoided to place excessive emphasis providing lengthy and strict training to either students or external assessor as this might have outshone the advantages of mutual learning embedded in PA (Gibbs, 2006). The results proved that the evaluation process was not hindered in any way.

The current study also confirmed that the use of only one assessor per assessee can have a positive effect on the reliability and, in a way, the validity of PA as was indicated by previous research (Stefani 1994). Consequently, using multiple raters could be avoided since that increases the demand of time and effort which might compromise the efficiency of the execution (Sung et al. 2010).

In summary, careful training of the learners and the external assessor increased the reliability and validity of PA (Tables 3 & 4) in the current study and ensured the successful implementation of PA with adolescent learners.

9. Implications, Limitations and Suggestions for Future Research

The findings of this study have significant implications for both teachers and researchers. First of all, as several researchers (Pond et al. 1995; Ryan, Marshall, Porter & Jia 2007) claim, students' involvement in the formulation of the grading criteria improves the overall reliability and validity of PA because it allows them to better understand the scoring process. Teachers should keep in mind that students have to be given time, training and help to adapt to PA, in order to perform to the best of their ability and exploit its full potential (also in Berg 1999). Secondly, unlike previous researchers (e.g. Cho & MacArthur 2010), this study encourages teachers to use only one peer rater when they employ PA as part of their classroom-based assessment (Table 3 and Section 7).

Teachers are also encouraged to use rubrics with unambiguous scales and employ a small number of categories (five or fewer) since these are related with increased reliability (also in Sadler & Good 2006). This study showed that using a carefully designed analytic rating scale (Appendix I) increases the inter-rater reliability considerably (Table 4).

Moreover, future teachers/researchers should consider using specific guidelines, and blind review among peers because this increases student-teacher agreement on grades (reliability of PA) and presumably any resultant student learning outcome from this process (also in Sadler & Good 2006).

Furthermore, if the implementation of PA involves several teachers, these should receive sufficient and relevant in-service training (also in Brindley 1997) which will enable them to implement PA methods in their classrooms and support learners throughout this process. Finally, teachers and researchers should try to control the workload produced as a result of engaging in PA on students because this may reduce their willingness and grading accuracy (*ibid*).

At this point, it would be useful to refer to some limitations and concerns of this study. Firstly, PA has to become an integral part of the teaching programme throughout the school year on a national level in order to become effective and be accepted as part of teachers' assessment practices. (also in Black & William 1998). Moreover, this study was carried out with a small sample. Therefore, we must be cautious in generalizing more broadly. Other schools, teachers, and groups of students might behave differently and yield different results.

Furthermore, based on the findings of this research, future researchers should consider: (a) conducting a similar research with a larger sample of participants in order to increase the validity/reliability of the present research; (b) integrating peer-assessment into the EFL classroom by employing PA the beginning to the end of the school year in order to investigate its potential as a learning tool, and (d) extending peer-assessment implementation to lower grades or other educational levels, i.e. primary school.

10. Conclusion

The use of PA is an innovative way of classroom-based assessment. This study, making only a small contribution to the field of alternative assessment, has shown the potential of PA as a powerful alternative and learner-centred tool. It has shown that high levels of reliability (agreement between teacher and student marks) and validity are possible when students grade their peers' essays. Finally, it has indicated that all kinds of learners even the younger ones (e.g. adolescents), can get actively involved in PA and reliably assess the language proficiency of their peers as well as their own language skills and thus improve their writing performance.

References

- Berg, C.E. 1999. The effects of trained peer response on ESL students' revision types and writing quality. *Journal of Second Language Writing* 8(3): 215-241.
- Black, P. and D. William. 1998. Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice* 5: 7-74.
- Boud, D. 1990. Assessment and the promotion of academic values. *Studies in Higher Education* 15(1): 101-111.
- Brindley, G. 1997. Assessment and the language teacher: trends and transitions. *The Language Teacher*. Retrieved 20 May 2011 from: <http://langue.hyper.chubu.a.c.jp/jalt/pub/tlt/97/sep/brindley.html>
- Brown, S. and P. Dove. 1991. Self and peer assessment. *Standing Conference on Educational Development*. Birmingham: England.
- Chang, C.-C., K.-H. Tseng, P.-N. Chou and Y.-H. Chen. 2011. Reliability and validity of Web-based portfolio peer assessment: a case study for a senior high school's students taking computer course. *Computers and Education* 57: 1306-1316.
- Chapelle, C.A. and G. Brindley. 2002. Assessment. In N. Schmitt (ed.), *An introduction to Applied Linguistics*. London: Arnold, 268-288.
- Cheng, W. and M. Warren. 1999. Peer and teacher assessment of the oral and written tasks of a group project. *Assessment and Evaluation in Higher Education* 24(3): 301-314.
- Cheng, W. and M. Warren. 2005. Peer assessment of language proficiency. *Language Testing* 22(1): 93-121.

Cho, K. and C. MacArthur. 2010. Student revision with peer and expert reviewing. *Learning and Instruction* 20(4): 328-338.

Cho, K., C. Schunn and R.W. Wilson. 2006. Validity and reliability of scaffolded peer assessment of writing from the instructor and student perspectives. *Journal of Educational Psychology* 98: 891-901.

Falchikov, N. 1995. Peer feedback marking: development in peer assessment. *Innovations in Education and Training International* 32: 175-187.

Falchikov, N. 2005. *Improving assessment through student involvement: practical solutions for aiding learning in Higher and Further Education*. London: Routledge.

Falchikov, N. and D. Boud. 1989. Student self-assessment in higher education: a meta-analysis. *Review of Educational Research* 59(4): 395-430.

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

Falchikov, N. and D. Magin. 1997. Detecting gender bias in peer marking of students' group process work. *Assessment and Evaluation in Higher Education* 22(4): 393-404.

Fletcher, C. and C. Baldry. 1999. Multi-source feedback systems: a research perspective. *International review of Industrial and Organizational Psychology* 14: 149-193.

Gibbs, G. 1999. Using assessment strategically to change the way students learn. In S. Brown and A. Glasner (eds.), *Assessment Matters in Higher Education*. Buckingham: Open University Press, 41-53.

Goldfinch, J. and R. Raeside. 1990. Development of a peer assessment technique for obtaining individual marks on a group project. *Assessment and Evaluation in Higher Education* 15: 210-231.

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

Hanrahan, S.J. and G. Isaacs. 2001. Assessing self- and peer-assessment: the students' views. *Higher Education Research and Development* 20(1): 53-70.

- Hansen, J.G. and J. Liu. 2005. Guiding principles for effective peer response. *ELT Journal* 59(1): 31.
- Jones, L. and C. Fletcher. 2002. Self-assessment in a selective situation: an evaluation of different measurement approaches. *Journal of Occupational and Organizational Psychology* 75: 145-161.
- MacArthur, C.A., S.S. Schwartz and S. Graham. 1991. Effects of a reciprocal peer revision strategy in special education classrooms. *Learning Disabilities Research and Practice* 6: 201-210.
- Marcoulides, G.A. and M.G. Simkin. 1995. The consistency of peer review in student writing projects. *Journal of Education for Business* 70: 220-223.
- McIsaac, C.M. and J.F. Sepe. 1996. Improving the writing of accounting students: a cooperative venture. *Journal of Accounting Education* 14(4): 515-533.
- Meletiadou, E. 2011. *Peer assessment of writing in secondary education: its impact on learners' performance and attitudes*. M.A. dissertation, University of Cyprus.
- Miller, L. and R. Ng. 1994. Peer assessment of oral language proficiency. *Perspectives: Working papers of the department of English*. City Polytechnic of Hong Kong, 6: 41-56.
- Ministry of Education and Culture, Department of Secondary Education, State Institutes for Further Education. 2010. *Annual planning school year 2010-2011*. Nicosia: Ministry of Education.
- Oldfield, K.A. and J.M.K. Macalpine. 1995. Peer and self-assessment at tertiary level: an experiential report. *Assessment and Evaluation in Higher Education* 20: 125-132.
- Orsmond, P., S. Merry and K. Reiling. 1997. A study in self-assessment: tutor and students' perceptions of performance criteria. *Assessment and Evaluation in Higher Education* 22(4): 357-369.
- Oskay, O.O., M. Schallies and I. Morgil. 2008. A closer look at findings from recent publication. *Hacettepe University Journal of Education* 35: 263-272.
- Patri, M. 2002. The influence of peer feedback on self- and peer-assessment of oral skills. *Language Testing* 19(2): 109-131.

- Plutsky, S. and B.A. Wilson. 2004. Comparison of the three methods for teaching and evaluating writing: a quasi-experimental study. *The Delta Pi Epsilon Journal* 46(1): 50-61.
- Pond, K., R. Ul-haqa and W. Wadea. 1995. Peer-review: a precursor to peer assessment. *Innovations in Education and Training International* 32: 314-323.
- Pope, N.K.L. 2005. The impact of stress in self- and peer assessment. *Studies in Higher Education* 30(1): 51-63.
- Rudy, D.W., M.C. Fejfar, C.H. Griffith and J.F. Wilson. 2001. Self- and peer assessment in a first-year communication and interviewing course. *Evaluation and Health Professions* 24(4): 436-445.
- Ryan, G.J., L.L. Marshall, K. Porter and H. Jia. 2007. Peer, professor and self-evaluation of class participation. *Active learning in Higher Education* 8(1): 49-61.
- Saavedra, R. and S.K. Kwun. 1993. Peer evaluation in self-managing work groups. *Journal of Applied Psychology* 78(3): 450-462.
- Sadler, P.M. and E. Good. 2006. The impact of self and peer-grading on student learning. *Educational Assessment* 11: 1-31.
- Saito, H. and T. Fujita. 2004. Characteristics and user acceptance of peer rating in EFL writing classrooms. *Language Teaching Research* 8(1): 31-54.
- Sluijsmans, D.M.A. 2002. Student involvement in assessment, the training of peer-assessment skills. *Interuniversity Centre for Educational Research*.
- Sluijsmans, D.M.A., G. Moerkerke, J.J.G. van Merrknboer and F.J.R.C. Dochy. 2001. Peer assessment in problem based learning. *Studies in Educational Evaluation* 27(2): 153-173.
- Stefani, L. 1994. Peer, self and tutor assessment: relative reliabilities. *Studies in Higher Education* 19: 69-75.
- Stewart, M. and M. Cheung. 1989. Introducing a process approach in the teaching of writing in Hong Kong. *Institute of Language in Education Journal* 6: 41-48.
- Swanson, D.B., S.M. Case and C.P.M. Van der Vleuten. 1997. Strategies for student assessment. In D. Boud and G. Feletti. (eds.), *The challenge of problem based learning*. London: Kogan page, 269-283.

Sung, Y.-T., K.-E. Chang, T.-H. Chang and W.-C. Yu. 2010. How many heads are better than one? The reliability and validity of teenagers' self- and peer assessments. *Journal of Adolescence* 33: 135-145.

Topping, K.J. 1998. Peer assessment between students in college and university. *Review of Educational Research* 68: 249-276.

Topping, K.J. 2010. Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction* 20(4): 339-343.

Topping, K.J., E.F. Smith, I. Swanson and A. Elliot. 2000. Formative peer assessment of academic writing between postgraduate students. *Assessment and Evaluation in Higher Education* 25(2): 149-169.

Topping, K.J., D.J. Walker and S. Rondrigues. 2008. Student reflections on formative E-assessment: expectations and perceptions. *Learning, Media and Technology* 33(3): 221-234.

Tsagari, D. 2004. Is there life beyond language testing? An introduction to alternative language assessment. Retrieved 2 February 2011, from *CRILE Working Papers* 58 (Online). <http://www.ling.lancs.ac.uk/groups/crile/docs/crile58tsagari.pdf>

Tsivitanidou, O.E., Z.C. Zacharia and T. Hovardas. 2011. Investigating secondary school students' unmediated peer assessment skills. *Learning and Instruction* 21: 506-519.

Tsui, B.M.A. and Ng, M. 2000. Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing* 9(2): 147-170.

Van Steendam, E., G. Rijlaarsdam, L. Sercu and H. Van den Bergh. 2010. The effect of instruction type and dyadic or individual emulation on the quality of higher-order peer feedback in EFL. *Learning and Instruction* 20: 316-327.

Villamil, O.S. and M.C.M. De Guerrero. 1996. Peer revisions in the L2 classroom: social cognitive activities, mediating strategies, and aspects of social behavior. *Journal of Second Language Writing* 5(1): 51-75.

Appendix

The peer assessment form

Title of the essay: Group: No. of Comp. :.....

Criteria/Weighting	Excellent-Very Good	Good-Average	Fair-Poor	Very Poor
Content				
1. Are the main ideas clear and well-supported with helpful details?				
2. Are the ideas relevant to the topic?				
3. Is the text easy for the reader?				
4. Does the composition fulfill the task fully?				
Organization				
5. Is there thorough development through introduction, body and conclusion?				
6. Is there logical sequence of ideas and effective use of transition?				
7. Is there cohesion and are there unified paragraphs?				
8. Does the writer achieve coherence by using simple linking devices?				
C. Vocabulary and Language Usage				
9. Is the vocabulary sophisticated and varied?				
10. Is there effective word choice and usage? Is the meaning clear?				
11. Does the writer use simple/complex constructions effectively?				
12. Are there errors of tense and/or subject/verb agreement?				
13. Are there errors of number (singular /plural) and word order?				
14. Are there errors of articles, pronouns and prepositions?				
D. Mechanics				
15. Are there problems with spelling and handwriting?				
16. Are there errors of punctuation and capitalization?				

Analytic score:

Content: /5,

Organization: /4,

Vocabulary

and Language use: /6,

Mechanics: /5,

Total score: /20.