

Promise of positive outcomes from self- and peer-assessment

Ed Joyce, Carol Everhard and Niovi Antonopoulou

Aristotle University of Thessaloniki

Abstract

Research was undertaken to evaluate (a) the grading of written and oral assignments using peer-assessment and self-assessment compared with teacher-assessment and (b) the usefulness and effectiveness students felt this type of assessment had over more traditional forms. The research confirmed that (a) these methods of evaluation encouraged motivation and helped the students to understand course subject-matter better and (b) the checklists which were used helped the students not only to assess but to prepare their assignments more effectively. Thus, this assessment helped to make courses at university level more interesting for students and encouraged a greater amount of participation, which resulted in more effective learning. Results showed clearly that peer-assessment and self-assessment at tertiary level result in positive outcomes which could encourage lifelong learning.

Keywords: evaluation, checklist, criteria, peer-assessment, self-assessment

1. Introduction

Students entering the School of English of Aristotle University are, for the most part, products of a public or private education system which does not encourage them to take any learning initiative. Choices about curricula and syllabi lie, predominantly, in the hands of the Greek Ministry of Education, and this top-down approach to educational management straitjackets all instructors, who are required to test regularly what has been taught from the specified syllabus or adopted textbook. This puts the emphasis on testing rather than teaching from the very earliest stages of learning through to the final years of education provided by the State.

Private primary and secondary schools are obliged to follow the same provisos, leaving little room for deviations from the norm. In fact, very little time is devoted to foreign language education in the final years of secondary education when the emphasis, for the purposes of university entrance, is on History, Greek Philology, and the dead languages of Ancient Greek and Latin, as well as on Mathematics. The emphasis on test-taking rather than communication continues and this, in turn, encourages memorisation and rote-learning of language, with a view to taking and passing tests. Students in a test-driven and exam-based system lose any sense of

responsibility or independence, and have no choice but to go with the system, or either be rejected by it or reject it themselves. Indeed, most might choose the latter option if they were not driven by peer pressure, and, even more so, by parental pressure to join the university elite.

Even in the sector of private language tutoring, there is an emphasis on examination syllabi and exam-taking so that students gain certificates from public examination bodies as their main proof of learning and language competence. With a few exceptions, rather than empowering learners as language users, the approaches taken both in the state and in private sectors to language teaching and learning seem to result in students' "learned helplessness" (Wenden 1991: 57) and very little in terms of "ownership" (van Kraayenoord 1993, cited in Smolen *et al.* 1995: 22-27) of the language. The fact that students in their early twenties should still be 'helpless' is somewhat alarming and it was an attempt to counteract this situation that led to the undertaking of this project, which implements peer- and self-assessment at tertiary level.

2. Reasons for investigating assessment procedures

Instructors are always looking for ways to add vigour and interest to courses and engage their learners in the learning process and improve the quality and success of their teaching. It was with these views in mind that our project was set up. Being instructors in tertiary education, we were interested in investigating if involvement of students in the processes of peer- and self-assessment would (a) help students understand their courses and course content better, (b) help students see better how to plan their work and prepare assignments, (c) encourage students to develop a more critical eye with regard to assessments they are subjected to, (d) create more intrinsic forms of motivation, (e) increase clarity concerning the course subject-matter and, finally, (f) improve the accuracy of their abilities to judge their own performance and progress (Trim 1978: vii-xii). All of this would be achieved by involving the students much more in the assessment procedures used on them and by them, by making more transparent to students what the aims and objectives, of the course and of particular assignments, are through the use of checklists with criteria that reflect these objectives (Council for Cultural Cooperation Education Committee 2001, Swiss Conference of Cantonal Ministers of Education 2001). All the above, therefore, justify the reasons for investigating the efficacy of peer- and self-assessment.

2.1 Comparing the effects of assessment procedures

Given what has been stated previously, there is some value in comparing the effects of assessment procedures that are traditional with those that involve everyone concerned in the learning process. Traditional forms of assessment are usually conducted by the instructor choosing to use a tool handed down from and designed by a supposed testing expert. Thus all the authority and power as regards grading is overseen by this one individual. Such judgments are summative and supposedly objective and mostly based on criteria which are neither transparent to the assessor nor to the assessed, yet are accepted and condoned by all parties. While such scoring is accepted as valid and reliable for some tests, it is not guaranteed to be so for all tests. What is more, students are kept at a distance when such autocratic assessment is applied to them (Stickler, Lewis and Speight 1999: 271-289). It promotes passivity and the feedback students receive is negative since it puts emphasis on mistakes made and focuses on lack of progress, so increasing the students' sense of vulnerability and weakness. There is no positive energy or creative exploitation of students' learning from their own mistakes or those of others. Learning generally ends with the completion of the task and its submission by the learner.

With peer- and self-assessment, on the other hand, judgement is formative, with all parties i.e., the instructor, the producer of the assignment and peers, all contributing to assessment and playing an equal role. Through self-assessment, individuals are involved in a continuous process of self-awareness and reflection, during pre-production, production and post-production phases. The errors of other individuals as well as those made by themselves become apparent in a more democratic and revelatory way, adding to the self-awareness and reflection process (Dimitrova 1995: 91-101, Nachi 2003: 157-173). Mistakes are thus used as correction tools within the peer- and self-assessment cycle and reflection goes beyond the self and the production of a single assignment. With extended and extensive use, the validity and reliability of the method increases, as does the self-confidence, self-esteem and objectivity of its users. The processes involved are productive and active, as well as interactive and promote positive feelings about the value of knowledge and learning which are conducive to self-fulfilment and lifelong learning (Stickler *et al.* 1999, Wagner 2000: 25-27).

2.2 *Objections to moving away from tradition*

However unsatisfactory traditional procedures of testing and assessment might be, it is difficult both for instructors and students to break with pedagogical traditions. Many of the objections to peer- and self-assessment stem from the entrenched views of the status quo (Gardner 2000: 49-60). Concerns regarding handing over responsibility to learners and the encouragement of autonomous learning have not necessarily led to trusting individual students with assessment of their own work and that of other students. Suspicions still lurk among education authorities and pedagogues that scoring and assessment are not necessarily the kind of empowerment that should be given to learners. Rather, they feel that students would abuse the system (Stickler *et al.* 1999, Nachi 2003). In the academic setting of tertiary level education, in particular, the feeling prevails that the control and authority of the academic staff should predominate with regard to scoring and grading and there is the fear that anything new would require a great deal of time for training both the instructors and the students. Instructors themselves are not convinced of the need for change, and students, in turn, believe that they do not possess sufficient knowledge to assess either others or themselves. Also, because such a system of assessment requires a great deal of planning and collation of assessment grades, instructors claim that it is too time-consuming and stressful and that students also find it stressful. Most institutions also have a standard, rigid system of assessment that does not permit variation and deviation.

2.3 *Supporting the 'involving' approach*

As mentioned above, instructors usually strive for the betterment of conditions of learning and they should therefore attempt to use a system of evaluation which would both improve assessment and make it fairer and at the same time improve the teaching/learning relationship by creating conditions for effective learning through greater learner involvement and engagement with the learning process. By becoming engaged in evaluation and assessment, learners become more involved in the teaching/learning process overall. The responsibility involved in reflection over criteria and marking scales and on the quality of work produced would seem to encourage their powers of critical thinking and, at the same time, boost motivation. While training learners to peer-assess and self-assess seems to be a time-consuming process, everyone involved in such a process reaps great benefits. More importantly, students become much more active members of the learning community and the more democratic and

power-sharing processes involved in arriving at grades nurture positive feelings, while at the same time encouraging learner autonomy (Kohonen 2000: 1-6, Stickler *et al* 1999).

Students come to realise that it is not the grades that are of most significance, but the processes involved in arriving at those grades, which at the same time guide their learning, which are of most importance. Reflecting on the work which other students produce enables the students to take a more critical stance in relation to their own work. They become more aware of influences on their language such as that of transfer errors, they develop more patience and devote more time to producing accurate work, whether in assessment, using the same criteria and scales, or in production of their own work.

Traditional methods of assessment and evaluation offer only a grade by way of feedback and the only involvement of the learner is in the submission of assignments. Such assignments are usually produced in total ignorance of the criteria by which they will be judged. Using formative assessment, learners are involved at all stages and the feedback they receive is of much more significance to them when they are involved and their peers are involved. Each piece of work is viewed from at least three perspectives and thus offers feedback which is much more meaningful.

The results of this project would seem to support assessment procedures that involve students, and reveal that even when operating within a traditional and rigid approach to learning and assessment, it is possible to introduce a scheme of alternative assessment which results in improved preparation and understanding of assignments through the use of checklists which, as Oskarsson (1998) also claims, guide and improve assessment and encourage reflection on learning goals. What seems clear is that through time there is ever-stronger agreement between peer-, self- and instructor assessment.

While schemes which use continuous assessment are already practised by some instructors in conjunction with a score on a final examination, in order to increase fairness and to take political and ethical considerations about assessment into account, what we have done adds another dimension. Evidence suggests that this alternative approach can be combined with the traditional summative approach, to give greater clarity and achieve greater objectivity as regards assessment. This is true irregardless of the use the results are put to, whether that be to discriminate between students' performances or to guide them towards greater self-awareness and self-improvement. Comments gleaned from students using follow-up questionnaires and discussions as

feedback seem to support the view of Gardener (2000) that students feel that such a route to assessment definitely helps.

3. The Research

There were three stages to the research, with improvements being made at each stage. Stage 1, in the autumn semester of 2004 – 2005, was a pilot stage to introduce post-graduate students on the ‘Measuring Second Language Performance’ course to alternative assessment. After an introduction to the theory behind peer- and self-assessment to the group as a whole, 9 students made brief oral presentations on a topic taken from the course. They researched it and supported their positions with reference to a given bibliography. In non-class time the presenters-to-be received instructions about how to complete their projects and present them and learnt about the criteria of the checklists. They were assessed by peers and assessed themselves using checklists and criteria given in Appendix 1. The instructor also assessed. On completion of these procedures, a follow-up group discussion led by the instructor within a loose framework revealed attitudes and responses, and led to the conclusion that improved checklists were needed.

The participants in Stage 2, in the spring semester of 2004-2005, were 9 students from one of the 10 Language Mastery 2 (LM2) courses who worked with 5 written assignments and used checklists to peer-assess and self-assess, each having criteria that tightly fitted the course outline and objectives, at that point in the course, of the assignment (see Appendix 2 for an example). The instructor also assessed. Language Mastery courses aim to help 1st year students gain an overall command of the language and develop their language abilities. LM2 focuses on argumentative texts, whereas LM1 focuses on descriptive and narrative texts. On completion of this first application, after discussion, instructors felt another stage would remove doubts that the findings to date were not due to chance.

Stage 3, in the autumn semester of 2005-2006, was conducted, in the first instance, with 30 students (2 x 15) from the two large ‘Testing and Evaluation’ and ‘Assessment in the Classroom’ 4th year courses, who delivered oral presentations, based on topics from their course-books, to the others, and, in the second instance, with the students from 10 1st year Language Mastery 1 courses who completed both oral and written assignments. Instructors were well-informed beforehand about the kind of assessment to be conducted and about the improved checklists. Peer-, self- and instructor assessments

were carried out and a follow-up questionnaire (see Appendix 3.) was given to 97 of the students still available, drawn from both the 4th year and 1st year groups.

3.1 Summary of the findings

Interesting patterns emerged from the research. As one would expect, initially students were rather afraid of being critical both with others and themselves and tended to be over-generous in their scoring, so that, to begin with, the instructor assessment tended to be lower than both self-assessment scores and peer-assessment scores. In the initial stages, there tended to be quite strong agreement between peer-assessment scores and self-assessment scores.

Given that students find it more difficult to be critical of themselves, self-assessment scores tended to be higher than scores given by peers, a finding in agreement with Saito and Fujita (2004: 31-54). Generally, there seems to be slightly less agreement over scores for writing than there is compared to scores for oral work. This is not surprising since it is generally recognized that assessing writing is a more problematic task. The pleasing discovery was that with frequency of application, over time such assessment led to greater agreement and convergence of scoring between self, peer and instructor scores, showing a shift away from the subjectivity to which all were prone initially, to that where an understanding of the task purposes, the criteria on the checklists and the correspondence with course content, led to more consistent scoring and, hence, objectivity.

Information gleaned from questionnaires indicated that students were satisfied with this method of assessment and genuinely felt that they were more aware of the learning process. Students felt they had gained a greater understanding of how assessment procedures worked, and enjoyed the fact that they had been more actively engaged in the processes of learning and assessment than they would have been had traditional methods alone been used. As with Oskarsson (1984), students also reported that they felt much more strongly motivated and focused and that their learning was better-organised. They thought that the teaching of other subjects on their university programme could benefit from the application of such an approach to assessment and felt that they wanted to learn more about the purposes of assessment schemes and the thinking that underlies them, points brought out also by Brindley (1994). Another point that gives still greater weight to the promotion and use of such 'inclusive' and 'involving' methods of assessment and which helps to overcome any objections raised

to it is the fact that, just as with Nachi (2003) and Stickler, Lewis and Speight (1999), students also admitted that gradually they did not try to abuse the method and cheat.

4. Authentic methods of assessment in universities

The indications from this research reveal that peer- and self-assessment schemes are acceptable to students and instructors and as such can be applicable to higher education contexts. Their use, especially in the initial stages, does require patience and tolerance on the part of all involved, and the creation of an atmosphere that demonstrates a framework of impartiality within which everyone cooperates (Kohonen 2006: 28-29). With regard to writing, the criteria need to be quite 'tight' in order for the assignment and course objectives to be quite clear. Assessment of writing also seems to work better with smaller groups (Kohonen 2006), a consideration that should be taken into account by course planners, providers and administrators.

Time has to be made to explain procedures to students and build on feelings of trust and security (Wagner 2006). Time also has to be allowed for the training of instructors and learners in the method in order to avoid misunderstandings since a significant change in role is required (Gardner 2000, Kohonen 2006). What seems clear is that, with time, all parties learn to see this method of evaluation as one which not only provides a tool for assessment, but also as being valuable in the encouragement of more fulfilling teaching and learning experiences (Kohonen 2006). This research exercise seems to leave us with the impression and promise that using such assessment procedures in higher education contexts leads to more positive learning outcomes.

References

- Brindley G. (1994) *Assessing achievement in the learner-centred curriculum*. Sydney: NCELTR.
- Council for Cultural Cooperation Education Committee (2001) *Common European framework of reference for languages: learning, teaching, assessment*. Strasbourg/Cambridge: Cambridge University Press for the Council of Europe, 177-196.
- Dimitrova S.G. (1995) Peer and self-assessment – a step towards a more effective acquisition of essay-writing skills. *Working Papers in Applied Linguistics* 2: 91-101.
- Gardner D. (2000) Self-assessment for autonomous language learners. *Links and Letters* 7: 49-60.
- Kohonen V. (2006) A note on self-assessment and peer-assessment using the European Language Portfolio (ELP) in Finland. *Independence, Newsletter of the IATEFL Learner Autonomy SIG* 39:28-29.

- Nachi H.E. (2003) Self-assessment and learner strategy training in a coordinated program: Using student and teacher feedback to inform curriculum design. In A. Barfield and M. Nix (eds), *Autonomy you ask!* Tokyo: The Learner Development Special Interest Group of JALT, 157-174.
- Oskarsson M. (1984) Self-assessment of foreign language skills: A survey of research and development work. Strasbourg: Council for Cultural Cooperation, Council of Europe.
- Oskarsson M (1998) Learner self-assessment of language skills: A review of some of the issues. Gdansk: IATEFL Special Interest Group Symposium.
- Saito H. and T. Fujita (2004) Characteristics and user acceptance of peer rating in EFL writing classrooms. *Language Teaching Research* 8:31-54.
- Smolen L., C. Newman, T. Wathen and D. Lee (1995) Developing student self-assessment strategies. *TESOL Journal* 5:22-27.
- Stickler U., T. Lewis and R. Speight (1999) Taking students seriously: The use of self-assessment and peer-assessment on a university-wide language programme. In D. Bickerton and M. Gotti (eds), *Language centres: Integration through innovation. Papers from the 5th CERCLES Conference, Bergamo 17-19 Sep. 1998.* Plymouth: CERCLES, 271-290.
- Trim J.L.M. (1978) Preface and foreword. In M. Oskarsson (ed.), *Approaches to self-assessment in foreign language learning.* Oxford: Pergamon Press for Council of Europe, vii-xii.
- Wagner S. (2006) Working with the CEF. *Independence: Newsletter of the IATEFL Learner Autonomy SIG*, 39: 25-27.
- Wenden A. (1991) *Learning strategies for learner autonomy.* Hemel Hempstead: Prentice Hall International.

Appendix 1

The following (based on criteria from Holt, Rinehart and Winston at http://go.hrw.com/resources/go_ss/teacher99/rubrics/RUBRICS24.pdf) was used in the first and third stages where students made individual oral presentations. It is clear that the statements in it are designed to make students assess more general criteria. The audience and participants had access to a handout that described, in general terms, what the numbers meant.

| ORAL PRESENTATIONS | |
|--|-------------------------------|
| Topic: | |
| Presenter: | |
| Instructions: Use this form to assess the presentation you will attend. For each of the statements use the scale to indicate the extent to which you judge the presenter to have fulfilled the criterion it reflects. | |
| 1 = Weak 2 = Moderately weak 3 = Average 4 = Moderately strong 5 = Strong | |
| 1. The presenter has researched the topic well. | 1 2 3 4 5 |
| 2. The presentation was well organised and cohesive | 1 2 3 4 5 |
| 3. The presenter(s) had prepared the presentation well. | 1 2 3 4 5 |
| 4. Appropriate and accurate information was included in the presentation. | 1 2 3 4 5 |
| 5. The presenters understand the topic well. | 1 2 3 4 5 |
| 6. The presenters were able to present a presentation that was well put together. | 1 2 3 4 5 |
| 7. The presentation included everything that was needed to make it work successfully. | 1 2 3 4 5 |
| 8. The presentation was well supported by the visuals and graphics that were used. | 1 2 3 4 5 |
| 9. The presentation was pleasant to watch. | 1 2 3 4 5 |
| 10. The presenter did not need to rely heavily on notes to make the presentation clear. | 1 2 3 4 5 |
| 11. There was no problem hearing the presenter(s). | 1 2 3 4 5 |
| Additional comments:..... | |
| | |
| | |
| | |
| Total points: | |

Appendix 2

The following checklist was used with the 1st presentation made by the LM2 classes as part of the second stage. Different checklists were used for each of the following 4 assignments. Each of those checklists had a statement that asked the students to judge the statements on previous checklists as a whole, but also had new statements that were to do more with the aspects of writing argumentative texts that had been concentrated on in the relevant, preceding lessons.

As stated in the text, these classes aim to make students take command of the language. Instructors could set their own assignment titles and topics, but the idea was that they roughly all followed the syllabus and course outline that dealt with the features of producing argumentative texts in the same order, and so the criteria for marking was the same for each assignment no matter what the topic/title. Some discussion was held about the scale numbers, but the idea was that the judges/students worked out for themselves what they thought they meant. An example of an LM2 title was: 'Each person is responsible for his/her own happiness'. Assignments were prepared and completed as homework, or in class.

| WRITTEN ASSIGNMENTS |
|---|
| Author: Peer / Self / Instructor (circle) |
| Circle the numbers that represent the degrees to which you think the author's work satisfies the criteria. 1 = Nearly not at all 2 = Very little 3 = Not bad/not good 4 = Quite well 5 = Very well |
| 1. The purpose of the text became clear as you went through it. 1 2 3 4 5 2. The text was organised in a way that matched the purpose. 1 2 3 4 5 3. The text revealed what the author felt about the topic. 1 2 3 4 5 4. The content the author chose to put in the text was relevant to the author's purpose. 1 2 3 4 5 5. There was a feeling of unity about the text. 1 2 3 4 5 |
| Total points: |

Obviously there is only one instructor score and one self score for each individual student's assignment. The peer score was worked out using the average of the scores of however many students attended the class in the particular time slot that had been set aside for such work. Similar checklists were used in the third stage with those LM1 students who wrote 5 descriptive and narrative assignments.

Appendix 3

The following is part 'B' of the questionnaire used to gather feedback from students who took part in the third stage. A covering letter that explained the position of alternative assessment in applied linguistics, particularly that of self- and peer-assessment, and the reason for the research accompanied it. It also assured that confidentiality would be kept, and expressed the researchers' thanks. Part 'A' collected personal details about the students and their language learning experiences. Part 'C' asked for the students to comment on the assessment research in which they had been guinea-pigs, its strong and weak points and how to improve it.

| FEEDBACK QUESTIONNAIRE (Part 'B') | | | | | |
|---|--|---|---|---|---|
| Circle the numbers that represent how you feel about aspects of the assessment | | | | | |
| 1 = Not at all 2 = Very little 3 = Not bad/not good 4 = Quite well 5 = Very well | | | | | |
| 1. | I could read the checklists, statements and criteria. | | | | |
| | 1 | 2 | 3 | 4 | 5 |
| 2. | I find it an interesting way of conducting assessment. | | | | |
| | 1 | 2 | 3 | 4 | 5 |
| 3. | I find it easy to understand this way of conducting assessment. | | | | |
| | 1 | 2 | 3 | 4 | 5 |
| 4. | I understand the reasons for wanting to carry out such assessment. | | | | |
| | 1 | 2 | 3 | 4 | 5 |
| 5. | I found it easy to assess my own work. | | | | |
| | 1 | 2 | 3 | 4 | 5 |
| 6. | I found it easy to assess the work of my classmates. | | | | |
| | 1 | 2 | 3 | 4 | 5 |
| 7. | I prefer traditional forms of assessment. | | | | |
| | 1 | 2 | 3 | 4 | 5 |
| The checklists and criteria helped me to: | | | | | |
| 1. | participate more actively in the lesson. | | | | |
| | 1 | 2 | 3 | 4 | 5 |
| 2. | understand the aims of the course and its lessons. | | | | |
| | 1 | 2 | 3 | 4 | 5 |
| 3. | realise what exactly was being assessed. | | | | |
| | 1 | 2 | 3 | 4 | 5 |
| 4. | pay attention to areas and points of the lesson I would otherwise have thought unimportant | | | | |
| | 1 | 2 | 3 | 4 | 5 |
| 5. | develop the ability to think critically about my assignments. | | | | |
| | 1 | 2 | 3 | 4 | 5 |
| 6. | develop such a critical approach to my classmates' work. | | | | |
| | 1 | 2 | 3 | 4 | 5 |
| 7. | assess my work objectively. | | | | |
| | 1 | 2 | 3 | 4 | 5 |
| 8. | assess my classmates' work objectively. | | | | |
| | 1 | 2 | 3 | 4 | 5 |
| 9. | follow and judge the progress I made. | | | | |
| | 1 | 2 | 3 | 4 | 5 |

| |
|---|
| 10. know exactly how I was being assessed. 1 2 3 4 5 |
| 11. organise my work better. 1 2 3 4 5 |
| I would prefer: |
| 1. the existence of more statements and criteria. 1 2 3 4 5 |
| 2. to do more assignments. 1 2 3 4 5 |
| 3. to assess more assignments. 1 2 3 4 5 |
| 4. to be assessed in this new way in all my lessons. 1 2 3 4 5 |