Language testing and assessment issues in the Greek educational context

Guest editors:

Dina Tsagari

Spiros Papageorgiou

PATRAS

Research Papers in *Language Teaching and Learning*

Research Papers in Language Teaching and Learning

Table of Contents of Volume 3, Issue 1

Special Issue

Language Testing and Assessment Issues in the Greek Educational Context.

EDITORIAL

The links between teaching, assessment and testing, and their impact on learning, have always raised a particular interest for language educators. Different contexts present different challenges and it is rewarding – indeed, it is necessary – to learn from research studies of how these links fare under widely variable circumstances. In this special issue of *Research Papers in Language Teaching and Learning* we are particularly happy – and, indeed, honoured – to host an extended discussion of a great variety of concerns in the area of language testing and assessment. The focus here is on the language teaching and learning contexts of two countries of the Expanding Circle with a great interest and a significant tradition in assessment and testing practices, i.e., Greece and Cyprus.

There are many reasons why this special issue is unique. For one, it offers a comprehensive overview of the state of the art in language assessment and testing in Europe and in Greece. It assembles a number of research accounts that will help readers appreciate different aspects of the current situation regarding high-stakes examinations in Greece. What is more, there are papers on central concerns for language teachers interested in assessment and testing practices, such as the role of the teacher, the function of courseware, the status, reactions to and beliefs about high-stakes examinations, and classroom-based assessment practices in the state and private domain. Of particular interest are a series of innovative proposals in the form of case studies that touch upon alternative assessment, self-assessment, peer-assessment and ICT-enhanced assessment – all written by language teachers who have practised them.

It is also important to mention that all this information is accessible freely on the web.

I would personally like to thank the guest editors of this special issue, Dr Dina Tsagari and Dr Spiros Papageorgiou, two of the most prominent Greek colleagues specialising in language assessment and testing today, for their professionalism and painstaking diligence in supervising the production of this issue from start to finish. It has proved to be a demanding but rewarding experience for everyone involved. I am confident that this volume will be a reference point for researchers and educators for many years to come.

Nicos C. Sifakis
Editor-in-Chief

Special Issue
on Language Testing and Assessment
in the Greek Educational Context

Introduction

Dina Tsagari and Spiros Papageorgiou

Language Testing and Assessment (LTA) is of special interest in Greece and Cyprus, the two countries where Standard Modern Greek is the official language[1], for various reasons. First of all, LTA has permeated state school EFL education with teachers following specific testing requirements while trying to meet other instructional and administrative needs (see Tsagari and Pavlou, 2008; Tsagari, 2011a; Vogt *et al*, in press). As a result, EFL teachers experience various roles, e.g. teachers as 'supporters of language development' as well as 'examiners' and 'raters'. In trying to maintain these roles teachers have to keep a balance between the directives for summative assessment data of learner achievement (for bureaucratic reporting purposes) and need for formative assessment for language learning and instructional planning as recommended in the new school curricula[2]. This state of affairs has resulted in creating an imbalance in the range of assessment practices and methods used in EFL state school education affecting significantly the development of enhanced, student-oriented assessment practices. There is also evidence of the 'washback effect' (Alderson & Wall, 1993) in teachers' LTA practices that stems from classroom-based testing practices associated with external measures of language performance, as well as from the overreliance on textbook materials and its accompanying test booklets that are often used as sources of teachers' assessments (Tsagari, 2009).

The private language sector in both countries presents an equally interesting picture. For example, even though students are taught EFL in public schools, the majority of them attend foreign language classes in private institutions called "frontistiria" or receive tuition on a one-on-one basis (Tsagari, 2006; 2009). The motivation for attending additional foreign language classes is the drive to obtain foreign language certification offered by international and local testing agencies who are very active in both countries (for a list of such agencies see Papageorgiou, 2009, p. 199). Tsagari (2009, pp. 190-202) presents empirical evidence that this desire is not only because of future professional or educational plans but also because of personal reasons, in particular self-esteem, as a result of belonging to the group of "successful" students (i.e. those who pass a language exam). All these, along with the official recognition such

language exams receive by the State explain the value added to the various language certification systems that exist in both contexts. This state of affairs has often led to discussion about and research on the impact of such tests on language learning and teaching (see Tsagari, 2009; 2011b; in press).

However, even though foreign language testing and certification has been a mainstay in the present contexts for so many years, there is currently lack of a compendium of studies that can offer insights into large-scale foreign language testing and classroom-based assessment practices in the two educational contexts. The aim of this special issue of *Research Papers in Language Teaching and Learning* is to address this gap in the existing literature by putting together a corpus of both theoretical and research-oriented papers that present local and European trends and findings of research projects in LTA conducted within the Greek educational systems. These are thematically organized under three parts which are briefly described below.

1. *Language testing developments and issues in Europe and in Greece.* The first paper in the volume is written by a distinguished European testing scholar, Professor Sauli Takala, who has served as the President of the European Association for Language Testing and Assessment (EALTA) and has also been closely involved in the work of the Language Policy Division of the Council of Europe (see for example Takala, 2004). In his paper, Takala presents a detailed account of the developments and future challenges in the European testing and assessment context, which inevitably have a strong influence on the Greek educational system. The second paper focuses on similar issues but this time with specific reference to Greece. It is written by an active academic in the country, Professor Sofia Papaefthymiou-Lytra, who has taught and trained hundreds of pre-service and in-service teachers of English.

2. *Issues related to English-language examinations administered in Greece.* The next four papers by doctoral students (Papafilippou, Nteliou, Liontou and Delieza) present research related to testing agencies currently administering examinations in Greece. They demonstrate a very interesting mix of methods and language skills they investigate, as well as materials published by various testing agencies.

3. *Assessment issues in foreign language classrooms.* The remaining papers were submitted by in-service teachers, who obtained the M.Ed. in TESOL from the Hellenic Open University with an emphasis on language assessment. Vlandi explores teacher assessment practices in the language classroom of a state high school and the students' view of these practices and makes recommendations about the professional development of EFL teachers. Five papers, by Barabouti, Kouzouli, Daphni, Bompolou and Efthymiou investigate different aspects of the use of portfolios in state school classrooms and offer suggestions regarding this type of assessment. Two more papers discuss the application of self-assessment (Chalkia) and peer-assessment (Meletiadou) and argue for a more learner-centered approach in the classroom. Lastly, three papers by Karayianni, Daskalogiannaki and Baglatzi present innovative assessment methods using technology.

The inclusion of papers written by graduate students (the majority of whom are practicing EFL state school teachers), as authors of this special issue was not accidental. By incorporating their papers in this volume we aimed to demonstrate that local expertise in LTA does exist and that a growing number of people with knowledge in the area can support innovative LTA practices in the educational system. For example, at the school level, the growing body of educators with a sound knowledge and understanding of LTA principles and techniques can lead to assessment innovations that are student-oriented and can have a positive impact on language learning. Supporting LTA literacy, especially in state school

education, is imperative at present as valid and fair LTA practices and methods facilitate the alignment of local educational standards with basic and 'transversal' competences and other educational priorities stipulated by the Council of Europe[3]. In addition, local expertise in LTA helps users of tests (students, teachers, parents, educational policy makers and employers) draw valid inferences from scores of large-scale tests. This is also important for testing agencies, both local and international, because their testing 'products' can be used in appropriate ways.

Last but not least, we would like to highlight one additional outcome of this edited volume. The gradual compilation, reviewing and editing of the papers as well as the provision of feedback during the lifespan of this special issue was a truly rewarding experience for us, the guest editors, as well as the graduate student/authors. In the numerous exchanges of communication via emails and phonecalls, the student/authors admitted that they gained a better understanding of LTA principles and of improving their academic skills in writing a paper for a peer-reviewed publication (for most of them this special issue was their first attempt). We hope that this collection of papers will be useful to future practitioners and researchers, as well as graduate students interested in LTA in Greece, Cyprus and elsewhere.

Authors' email: D. Tsagari: tsagari.konstantia@ucy.ac.cy; Spiros Papageorgiou: spapageorgiou@ets.org

---

Notes

1. Greek is the official language in Greece, and one of the two official languages (the other language being Turkish) in the government-controlled areas of the Republic of Cyprus (see Papageorgiou, in press, for more information on the assessment of Greek as a first or foreign language). Because Greek is also the language of instruction in the vast majority of schools in both countries, we refer to the educational systems of both Greece and Cyprus as the "Greek educational system".
2. See http://www.pi-schools.gr/download/programs/depps/english/3rd_b.pdf and http://www.nap.pi.ac.cy/files/agglika_prod_dem/04_AGGLIKA_GIA_TO_DIMOTIKO.pdf.
3. See http://ec.europa.eu/education/school-education/doc830_en.htm.

References

Alderson, J. C., & Wall, D. (1993). 'Does washback exist?' *Applied Linguistics, 14*/2: 115-129.

Papageorgiou, S. (2009). *Setting performance standards in Europe: the judges' contribution to relating language examinations to the Common European Framework of Reference*. Frankfurt: Peter Lang.

Papageorgiou, S. (in press). 'Assessing Greek.' In A. Kunnan (Ed.), *The companion to language assessment.* Hoboken, NJ: Wiley-Blackwell.

Takala, S. (Ed.). (2004). *Reference Supplement to the preliminary version of the Manual for relating language examinations to the Common European Framework of Reference for languages: learning, teaching and assessment*. Strasbourg: Council of Europe.

Tsagari, C. (2006). 'Linking EFL textbook materials to exam specifications.' Paper presented at the Third Annual Conference of EALTA, Krakow, Poland, accessed at http://www.ealta.eu.org/conference/2006/docs/Tsagari_ealta2006.ppt on 28 Jan 2012.

Tsagari, D. (2009). *The complexity of test washback: an empirical study*. Frankfurt: Peter Lang.

Tsagari, D. (2011a). 'Investigating the 'assessment literacy' of EFL state school teachers in Greece'. In D. Tsagari & I. Csépes (Eds.), *Classroom-based language assessment.* Frankfurt: Peter Lang, pp. 169-190.

Tsagari, D. (2011b). 'The washback effect of language proficiency testing on education and learning processes', accessed at http://rcel.enl.uoa.gr/periodical/research3_en.htm on 28 Jan 2012.

Tsagari, D. (in press). 'FCE-exam preparation discourses: insights from an ethnographic study.' *Research Notes*. (Available from http://research.cambridgeesol.org/research-collaboration/research-notes)

Tsagari, D. and P. Pavlou. (2008). 'Assessment literacy of EFL teachers in Greece and Cyprus'. In J. Burston , M. Burston, E. Gabriel & P. Pavlou (Eds), *Languages for intercultural dialogue*. Nicosia: European Parliament Office in Cyprus and Ministry of Education and Culture of the Republic of Cyprus, pp. 243-256.

Vogt, K., Tsagari, D. Sahinkarakas, Pavlou P., Migdal, U., Afiri, Q. & Guerin, E. (in press). 'Assessment Literacy of Foreign Language Teachers: Findings of a European Study.' *Language Assessment Quarterly*.

The Landscape of Language Testing and Assessment in Europe: Developments and Challenges

[Γλωσσική Δοκιμασιολογία και Αξιολόγηση στην Ευρώπη: Εξελίξεις και Προκλήσεις]

Sauli Takala

The article opens with a short sketch of developments in language testing and assessment, and presents Spolsky´s tripartite categorization of major approaches in language testing/assessment. This is followed by an account of current developments in language testing and assessment in Europe. One prominent development in Europe has been a strong increase in cooperation in language testing/assessment. This was shown by the emergence of associations devoted to language testing/assessment, ALTE in 1990 and EALTA in 2004, both producing codes/guidelines for good practice. A major outcome of cooperation in the field of language education, the Common European Framework of Reference (CEFR; Council of Europe, 2001), brought about further challenges for language testing/assessment, immediately after its publication in 2001. Its potential was utilized by DIALANG, a pioneering internet-based system for diagnostic assessment and still apparently unrivalled. Most attention in the article is devoted to an analysis of future challenges. Justification of testing/assessment is increasingly challenged, and work is being done on analyzing assessment use argumentation. The problem of criterion, while not a new challenge, calls for attention both in terms of its logical status and various systems of standards published. Standard setting – setting one, or increasingly more often, multiple cut scores to indicate levels of proficiency is another current challenge, especially in efforts to relate tests, examinations etc to the CEFR levels. Following interest in international comparisons generated by the PISA programme, there is a trend to carry out international comparisons in the area of language competence (e.g. EILC).

ଔ

Το άρθρο ξεκινά με μια σύντομη επισκόπηση των εξελίξεων στο χώρο της γλωσσικής δοκιμασιολογίας και αξιολόγησης, και παρουσιάζει την τριμερή κατηγοριοποίηση των κυρίων προσεγγίσεων της γλωσσικής δοκιμασιολογίας και αξιολόγησης όπως προτείνεται από τον Spolsky. Στη συνέχεια ακολουθεί περιγραφή των τρεχουσών εξελίξεων της γλωσσικής δοκιμασιολογίας και αξιολόγησης στην Ευρώπη. Ένα σημαντικό χαρακτηριστικό της γλωσσικής δοκιμασιολογίας και αξιολόγησης στην Ευρώπη πρόσφατα είναι η ενδυνάμωση της συνεργασίας. Αυτό φαίνεται από την εμφάνιση οργανισμών γλωσσικής δοκιμασιολογίας και αξιολόγησης, όπως ο ALTE το 1990 και EALTA το 2004, οι οποίοι εξέδωσαν κωδικούς / οδηγίες για την ορθή πρακτική στο χώρο αυτό. Ένα σημαντικό αποτέλεσμα της συνεργασίας στον τομέα της γλωσσικής εκπαίδευσης, το Κοινό Ευρωπαϊκό Πλαίσιο Αναφοράς (ΚΕΠΑ - Council of Europe, 2001), επέφερε

περαιτέρω προκλήσεις στο χώρο της γλωσσικής δοκιμασιολογίας και αξιολόγησης, αμέσως μετά τη δημοσίευσή του το 2001. Το δυναμικό του ΚΕΠΑ χρησιμοποιήθηκε από το DIALANG, ένα πρωτοποριακό και ασυναγώνιστο σύστημα διαγνωστικής αξιολόγησης βασισμένο στο διαδίκτυο. Μεγάλο μέρος του παρόντος άρθρου είναι επίσης αφιερωμένο στην ανάλυση των μελλοντικών προκλήσεων. Η αιτιολόγηση της εγκυρότητας των δοκιμασιών και της αξιολόγησης όλο και περισσότερο αμφισβητείται, και περαιτέρω συζητήσεις περιστρέφονται γύρω από την ανάλυση της επιχειρηματολογίας της χρήσης των αξιολογητικών αποτελεσμάτων. Το πρόβλημα των κριτηρίων, ενώ δεν είναι μια νέα πρόκληση, απαιτεί προσοχή τόσο από την άποψη της λογικής κατάστασης του και τα διάφορα συστήματα κριτηρίων που δημοσιεύθηκαν. Η θέσπιση κριτηρίων που να ορίζουν το διαχωρισμό δύο ή περισσοτέρων επιπέδων γλωσσικής ικανότητας είναι μια άλλη σημερινή πρόκληση, ειδικά στο πλαίσιο των προσπαθειών να συνδεθούν τα διάφορα γλωσσικά κριτήρια, εξετάσεις, κλπ με τα επίπεδα του ΚΕΠΑ. Ακολουθώντας στενά τις εξελίξεις στο χώρο φαίνεται πως υπάρχει μια τάση για τη διενέργεια διεθνών συγκρίσεων στον τομέα της γλωσσικής ικανότητας (π.χ., EILC), η οποία δημιουργήθηκε από τις συγκρίσεις που υπέδειξε το πρόγραμμα PISA.

Key words: CEFR, standard setting, cut score, ALTE, EALTA, DIALANG, EILC, international assessment, criterion, IRT

Historical sketch

Spolsky (1995) presents a review of the history of language testing in his seminal work "Measured Words". He refers to the long history of testing and cites ancient China as a case where written high-stakes testing was used in civil service recruitment. Competitive examination made its way in various forms to European countries, and a modern variant is the competitive examination that the European Union arranges for all those who wish to become EU officials (linguists, administrators, assistants)[1]. Spolsky´s main focus was on language testing in Great Britain and the United States, comparing developments in language testing and their contexts. The main thread in his exploration and analysis is the development of objective language testing.

At the 1975 AILA conference in Stuttgart, Spolsky (1978) presented a much-quoted tripartite classification of periods or approaches in language testing and assessment: pre-scientific (later called "traditional"), psychometric-structuralist (later "modern") and psycholinguistic-sociolinguistic (later "post-modern"). He believes that there is a good deal of truth in the tripartite division but feels a bit uneasy as it was based more on impression than documented evidence. He suggests that it is, in fact, more useful to see the development of language testing as "an unresolved (and fundamentally unresolvable) tension between competing sets of forces" (Spolsky, 1995, 354). These forces are both practical and ideological.

There is also an early account of the development of language testing in the US by John B. Carroll, who by all accounts was a towering figure in language testing/assessment (and in measurement in general). His review covers the period from the late 1920s through to 1954, the year in which it was produced, but as it was never published, it is unfortunately little known[2]. The review covers 49 pages of texts, provides a comprehensive list of available tests and contains a bibliography of about 80 references.

At regular intervals – often at the entry of a new decade - there has been stock-taking in the form of congresses/seminars and related publications (conference proceeding, books). Some examples are: the papers presented at an international symposium in Hong Kong in 1982 (Lee et al., 1985); de Jong & Stevenson (1990); Huhta, Sajavaara & Takala (1993), which includes the plenary by Charles Alderson on the state of language testing in the 1990s, and Spolsky (2000), which is a review of articles on language testing/assessment published in The Modern Language Journal.

Current situation in language testing in Europe

Cooperation in language testing/assessment

No man is an island, and Europe is no island in language testing and assessment. Language testing/assessment has become an increasingly international domain of activity since the setting up of the Language Testing Research Colloquium, LTRC, (1979) and the International Language Testing Association, ILTA, (1992). Interest in specifically European co-operation in language testing and assessment is partly due to the influence of the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001), which was published in 2001 but was available for consultation and comments a few years before that. ALTE (Association of Language Testers in Europe) was founded in 1990, currently consisting of 34 members (examination providers) and covering 27 languages. EALTA (European Association for Language Testing and Assessment[3]) was created in 2004. In early 2011, it had 1232 Individual members, 145 Associate Members, 17 Expert Members and 54 Institutional Members. EALTA has intentionally aimed at a broad range of membership and inclusiveness, low costs and collegial support and co-operation.

Despite the growing interest in testing and assessment in Europe, American scholars have dominated the development of measurement theory, and they have also produced the most important tools and references, such as the handbook entitled "Educational Measurement" (Lindquist, 1951), which, since then, has appeared in four editions and the "Standards for Educational and Psychological Testing" (AERA, APA, NCME, 1999), which, since 1954, has appeared in six editions. Ethical and programme evaluation standards have also appeared. In these publications, concepts such as reliability, validity, fairness/bias have been addressed and continuously elaborated. However, it is Europe that has made one of the most important contributions to testing: in the early 1960s the Danish psychologist/mathematician Georg Rasch developed the powerful measurement model which goes under the name of Rasch modeling (Rasch, 1960; see also Verhelst, 2004).

The language testing associations have also contributed to awareness-raising about good practice. ILTA pioneered the work on ethical standards in language testing and assessment and both ALTE and EALTA have developed related guidelines. EALTA´s Guidelines for Good Practice in Language Testing and Assessment have been published in 34 languages.

The challenge posed by the CEFR for language testing and assessment

The Common European Framework of Reference (CEFR), drawing on decades of development work within the Council of Europe (CoE) and based on a decade of focused activity, was ready to be published in 2001. The approach to language teaching and learning promoted by the CoE, and largely subsequently adopted also by the European Union, was designed to be responsive to the needs of the increasingly cooperative political structures in our multilingual and multicultural continent. While the aim was communicative and intercultural competence, language projects were also expected to contribute to the basic CoE values of human rights, democratic citizenship, and rule of law. This meant, among other things, strengthening pluralistic participatory democracy, promotion of intensified international cooperation, understanding and tolerance of cultural and linguistic diversity as a source of mutual enrichment, and democratization of education, with languages for all (Trim, 2007). This orientation has even strengthened in recent times.

The CEFR was developed during the medium-term (1990-1997) project entitled "Language Learning for European Citizenship". It can be described in a number of ways depending on which of the rich content facets one wishes to stress. Trim (2007:39), one of the key architects of the CEFR, indicates its broad scope in a succinct characterization:

a) a descriptive scheme, presenting and exemplifying the parameters and categories needed to describe, first, what a language user has to *do* in order to communicate in its situational context, then the role of the *texts*, which carry the message from producer to receiver, then the underlying *competences* which enable a language user to perform acts of communication and finally the *strategies* which enable the language user to bring those competences to bear in action;

b) a survey of approaches to language learning and teaching, providing options for users to consider in relation to their existing practice;

c) a set of *scales* for describing proficiency in language use, both globally and in relation to the categories of the descriptive scheme as series of *levels*;

d) a discussion of the issues raised for curricular design in different educational contexts, with particular reference to the development of *plurilingualism* in the learner, and for the *assessment* of language proficiency and achievement.

While the broad scope of the CEFR is increasingly being recognised and appreciated, it is obvious that the proficiency scales and their use in testing, assessment and examinations have received most attention. The scales have been seen by decision-makers as a concrete means for defining learning targets and assessing learning outcomes. One aim has been to use them in comparing the performance of the national language teaching provision with other nations. As this appeared not to be an easy task, there were immediately calls for the CoE to undertake a validation or certificating function.

However, CoE´s mandate does not include such functions. Instead, in cooperation with the Finnish national authorities it organized a seminar in Helsinki in the summer of 2002, which led to the setting up of an international working group with the task of developing a manual to help in relating/aligning examinations and tests to the CEFR levels. A pilot manual was issued in late 2003 and a revised one, based on feedback received, in 2009. The manual presents five steps: familiarization with the CEFR, specification of the content of the tests/exams in accordance with the CEFR descriptive scheme, training and benchmarking with samples for oral and written performance, standardization of level judgments/ratings and validation of set cut-scores. A Reference Supplement edited by the present author was also produced to provide more technical information about the theoretical foundations of standard setting. The expanded version of the Reference Supplement is only available on the Council of Europe website (http://www.coe.int/t/dg4/linguistic/Manuel1_EN.asp).

DIALANG – pioneering and still unrivalled?

To my knowledge, utilizing the CEFR for testing and assessment started with the EU project DIALANG, which my home department, Center for Applied Language Studies, University of Jyväskylä, coordinated during the first phase (autumn 1996 - November 1999). During the early part of 1996, the idea of producing language tests took form within EU and this activity was originally planned to be a pilot project like a number of other assessment projects within EU. The project was, however, soon transferred to DG XXII/LINGUA. In this context the original idea of accreditation and the personal skills card was abandoned in favour of a diagnostically oriented assessment system, which was approved by the SOCRATES committee.

DIALANG[4] was in many ways a novel approach to language testing and assessment. It developed a transnational assessment system with a large range of languages covered. It is diagnostically oriented, with one of its goals to promote diversified language learning in Europe. It combines self-assessment and external assessment. It uses the Internet as the delivery system and reports the results in accordance with the Council of Europe proficiency scales. This linking was decided upon as the use of the scales was seen to promote comparability across languages. I played an active role in developing the blueprint and I recall seeing in the system an opportunity to "democratise" testing/assessment by trying to put the user "in the

driver's seat" and by serving his/her individual interests (to be "at his/her beck and call"). In this, DIALANG displayed a similar sense of "service mission" as EALTA.

In practice, it was the CEFR Draft 2 self-assessment scales and communicative activity scales that were used/adapted. We also reviewed and utilised the objectives and definitions in the CoE publications entitled Waystage (A2), Threshold (B1) and Vantage (B2). While we found these useful for test specification, we also noted that there was considerable overlap, and thus progression was not always clear-cut.

DIALANG faced many daunting challenges: how to write test specifications, self-assessment statements, feedback statements and relate all this to the CEFR (see Alderson 2005). Of course, relating the scores to the CEFR was a huge challenge (Kaftandjieva, Verhelst & Takala 2000) and it became a "hot" topic as soon as the CEFR had been published in 2001. It needs to be pointed out that standard setting in DIALANG required a new approach: from the usual task of setting one cut-score (failing/passing the standard), a situation which was then typical in the US, as many as five cut-scores were needed. This was done using the "modified Angoff" method as a starting point[5].

The results of a validation study (Kaftandjieva & Takala 2002), which was designed and conducted as a part of a pilot study of a standard setting procedure specifically designed for the purposes of DIALANG, provided strong support for the validity of the CoE scales for listening, reading and writing. These findings not only confirmed that the DIALANG assessment system was based on a solid ground but they also had a broader impact, supporting the view that any further development of the CEFR could be undertaken on a sound basis.

There has been intensive work on standard setting in language education, especially in Europe. Reference will only be made to a few recent major sources that standard setters should be aware of and consult: Figueras & Noijons (2009), Kaftandjieva (2010) and Martyniuk (2010).

Future Challenges

Justification of testing/assessment challenged

One of the challenges facing testing and assessment in the future is related to one of the main meanings of "challenge". As assessment literacy (i.e. awareness and competence in assessment) grows – even if one might wish to see a more rapid growth than is in evidence at the moment – it can be expected that the values underlying testing/assessment as well as its practices will be increasingly challenged. It is also probable that more openness and transparency will be demanded. Even if we Europeans often tend to criticize the excessive emphasis on testing and examinations in the US, it seems clear to me that they are ahead of Europe in accountability in testing. Testing/examination procedures can be, and are regularly, challenged in court. Thus testing/examination bodies know that they have to be able to present good evidence and arguments for their procedures, practices and decisions. Major evaluation/assessment studies are regularly analysed critically and challenged[6]. A good EALTA colleague, Dr. Felianka Kaftandjieva (who passed away in 2009), was deeply disturbed by what she saw as too common European "sloppiness" and lack of transparency in testing/examination accountability and she wished to see a challenging approach similar to the American one taking root in Europe. I am basically sympathetic to this view. Bachman and Palmer (2010) provide a good discussion of the need to elaborate a reasonable case for any assessment, that is, assessment use argument.

As a founding member of EALTA and its second President, I am very pleased that its Guidelines for Good Practice in Language Testing and Assessment, available in 34 languages, address a broad range of target groups, using simple and comprehensible language. In retrospect, I now believe that it would be useful to

add the decision makers to the groups addressed. Their actions influence all other groups and there is a great need for much better assessment literacy among them.

Closely related to the issue of accountability is the concern with the *ethics* of language testing. The consequences of assessment were singled out by Messick (1989, 1994) as an important aspect of the uniform concept of construct validity. Shohamy (2001) has provided a critical discussion of the power that tests exert in controlling people and institutions. International language testing and assessment associations have developed codes of ethics/guidelines of good practice.

The problem of the criterion

McNamara (2004) provides a clear and concise model which illustrates the relationship between the test, the construct and the criterion. In the model, the construct is placed in the middle, the test to the left and the criterion to the right. The construct represents the theory of the domain assessed and provides a description of the essential features of performance. It influences the operationalization of the construct, test development, and leads to observable responses to test items and to other kinds of performances. This observable data leads to inferences – via the theoretical model – about the real-world performance, about the testee's actual standing in relation to the domain. These are inferences about something that is unobservable as tests are always a sample from the overall domain.

The nature of the *criterion* is a perennial issue in language assessment. Davies (2003) has addressed the elusive but ubiquitous concept of the *native speaker* as a criterion in applied linguistics, SLA research and also in language assessment (cf. Abrahamson & Hyltenstam 2009). Davies concludes that the native speaker is a myth but a useful one. This criterion has, however, been increasingly questioned and the interest in English as a lingua franca (ELF) and "World Englishes" (Kachru, Kachru & Nelson, 2006) has further problematized this criterion.

A number of other criteria (standards) have been produced in language testing. Important contributions include the Foreign Service Institute's scales developed about 50 years ago for testing proficiency and subsequent scales of proficiency. The Common European Framework for Languages (CEFR; Council of Europe, 2001) developed in Europe has acquired a strong position, not only in Europe but also in some other parts of the world. Such a standards-based approach to assessment defines *content* as well as *performance standards*, usually called Performance Level Descriptors (PLD, see e.g. Cizek & Bunch, 2007: 44-47).

Self-assessment

The role of *self-assessment* has been increasingly recognized. Early cognitive psychology (e.g. Flower & Hayes, 1977) reinforced the view that effective learning consists of a number of processes of which important ones are the skills of planning and appraisal. An effective learner can plan how to approach the tasks, which requires an ability to evaluate one's current level in relation to the task, and the ability to monitor the process and to assess and – if need be – revise the output. The ability to assess one's language proficiency is seen as a powerful factor in language learning. Oscarson's (1980) early work on self-assessment within the CoE modern language project was seminal in promoting the concept. Recent research in Europe is reported by Dragemark-Oscarson (2009) and Huhta (2010) in their respective doctoral dissertations.

The European Language Portolio (ELP), closely linked with the CEFR, is an example of a tool making use of self-assessment. There are several versions of it ranging from early learning to higher education. It seems, however, that the promise of the portfolio has not been very extensively materialized.

Less attention, undeservedly, has been devoted to peer assessment, a potentially very useful approach to assessment.

Standard setting

Standard-setting (setting cut scores) is a challenge when assessments, tests and examinations increasingly are required to report the outcomes in terms of proficiency levels. Cizek and Bunch (2007) is a useful general reference. A Manual for Relating Language Examinations to the CEFR is also freely available (Council of Europe, 2009).

Jaeger (1989, p. 492) observed that "much early work on standard setting was based on the often unstated assumption that determination of a test standard parallels estimation of a population parameter. This means that there would be a correct value for the cut score". This would simplify things, but unfortunately, it is not true, as Zieky (2001, p. 45) observes: "there is general agreement now that cutscores are constructed, not found". Standard setting cannot be reduced to a problem of statistical estimation, to the proper use of the best psychometric methods. Zieky (2001, p. 46) notes that "clearly, the cutscore is what participants choose to make it. It is also now clear that what participants choose to make the cutscore depends on subjective values ... Setting a sensible cutscore requires a determination of which type of error is more harmful [masters fail to pass; non-masters pass]." This does not mean, as early critics of standard setting argued, that standard setting is arbitrary and a waste of time. It is subjective but it does not have to be arbitrary in the ordinary negative sense of the word. Camilli, Cizek & Lugg (2001, pp. 449-450) argue that standards can be considered acceptable if they follow a "psychometric due process", an analogy to legal practice.

Zieky (2001, pp. 29-30) lists a number of issues that, in his opinion, had not yet (in 2001) been clearly resolved despite some twenty years of research and development work[7].

- Which method of setting cuts cores will be the most defensible in a given situation?
- Exactly how should the participants in the standard setting job be trained? Does training have to be face to face? What is the minimum acceptable training time?
- What normative information, e.g. prior score distributions, should be given to participants, if any?
- When should the participants receive normative information?
- Should the participants receive information about item difficulty? If so, should they receive information about every item?
- Exactly what cognitive processes do the participants use (see e.g., Papageorgiou, 2010)?
- To what extent are the participants capable of making the required judgments?
- Should participants be told the likely effects of their judgments on the resulting cut score?
- Should any judgments be excluded from the calculation of the cut score? If so, what criteria should be used to exclude judgments?
- How should the standard error of measurement affect the cut score?
- How should variability in the participants' ratings affect the cut score?
- Should compromise methods be used that combine normative and absolute judgments? Which method is the most appropriate?

A monograph by Felianka Kaftandjieva (2010), which discusses general issues in standard setting and explores six methods in setting standards in relation to the CEFR, is probably the most comprehensive treatment on the topic so far. In addition to providing evidence of the quality of the methods, it gives useful recommendations for standard setting.

More general guidelines are also available. The highly influential *Standards for Educational and Psychological Testing* (AREA, APA & NCME, 1999) listed six standards applicable to setting cut scores in its 1985 edition. These had mainly to do with being aware of and reporting error rates (misclassifications) and

providing rationales and explanations. There were no clear standards about HOW to set cut scores. The 1999 *Standards* contain 10 standards, with obvious overlap with the previous standards, but also standards about the actual processes of setting cut scores.

The 1999 standards show a new emphasis on the *actual processes* of setting cut scores[8]:

- Conditional standard errors of measurement should be reported at several score levels if constancy cannot be assumed. Where cut scores are specified for selection or classification, the standard errors of measurement should be reported in the vicinity of each cut score (2.14).
- When a test or combination of measures is used to make categorical decisions, estimates should be provided of the percentage of examinees who would be classified in the same way on two applications of the procedure, using the same form or alternate forms of the instrument (2.15).
- When raw scores are intended to be directly interpretable, their meanings, intended interpretations, and limitations should be described and justified in the same manner as is done for derived score scales (4.4).
- When proposed score interpretations involve one or more cut scores, the rationale and procedures used for establishing cut scores should be clearly documented (4.19).
- When feasible, cut scores defining categories with distinct substantive interpretations should be established on the basis of sound empirical data concerning the relation of test performance to relevant criteria (4.20).
- When cut scores defining pass-fail or proficiency categories are based on direct judgments about the adequacy of item or test performances or performance levels, the judgmental process should be designed so that judges can bring their knowledge and experience to bear in a reasonable way (4.21).
- When statistical descriptions and analyses that provide evidence of the reliability of scores and the validity of their recommended interpretations are available, the information should be included in the test's documentation. When relevant for test interpretation, test documents ordinarily should include item level information, cut scores and configural rules, information about raw scores and derived scores, normative data, the standard errors of measurement, and a description of the procedures used to equate multiple forms (6.5).
- Publishers and scoring services that offer computer-generated interpretations of test scores should provide a summary of the evidence supporting the interpretations given (6.12).
- Students who must demonstrate mastery of certain skills or knowledge before being promoted or granted a diploma should have a reasonable number of opportunities to succeed on equivalent forms of the test or be provided with construct-equivalent testing alternatives of equal difficulty to demonstrate that the skills or knowledge. In most circumstances, when students are provided with multiple opportunities to demonstrate mastery, the time interval between the opportunities should allow for students to have the opportunity to obtain the relevant instructional experience (13.6).
- If tests are to be used to make job classification decisions (e.g., the pattern of predictor scores will be used to make differential job assignments), evidence that scores are linked to different levels or likelihoods of success among jobs or job groups is needed (14.7).

International comparisons

There is a long tradition of empirical comparative research in educational outcomes. The pioneer was the IEA (International Association for the Assessment of Educational Achievement), which was started as a cooperative venture more than fifty years ago by a group of internationally minded educational researchers (e.g., Torsten Husén from Sweden and Benjamin S. Bloom from the US). In addition to international studies of mathematics and sciences, the IEA has carried out studies of English and French as a foreign language, of reading and literature (published in the early 1970s), and of writing in the late 1980s (the present author was the international coordinator). Studies of reading have continued and they have focused on 10–11-year olds (PIRLS, Progress in International Reading Literacy Study) in 2001, 2006 and 2011. OECD

(Organisation for Economic Co-operation and Development) initiated its PISA programme (Progress for International Student Achievement) in the early 2000s and has carried out three cycles of much publicized assessments (2001, 2006 and 2010).

Foreign language proficiency being one of the main priorities in the multilingual and multicultural Europe, EU recently initiated a project to assess the levels of proficiency attained. This initiative went under the name of "Barcelona Indicator" during the several years of preparations. A project was launched to explore the implications of this initiative called EBAFLS (Building a European Bank of Anchor Items for Foreign Language Skills[9]). It explored several questions, and one practical outcome was the infeasibility of using the L1 as the language of comprehension items in international testing. The project also found that many test items displayed strong DIF (differential item functioning, see Takala & Kaftandjieva, 2000) such that item difficulties tended to vary considerably among the participating countries. DIF is a major challenge and it will be very interesting to see the results of the EU project (EILC[10]) when they will become available in 2012.

 In international comprehension tests (such as PISA), it is important to control the effect of the different L1 translations and to ascertain that the level of text and item difficulty is comparable across the contexts. In PISA reading comprehension (L1) the translated versions are based on a source text – usually either in English or French – which is translated into several L1 languages. There is a set of procedures to guarantee that the texts and the items are not only equivalent translations but also of equal difficulty. However, this is a great challenge, and the results have occasionally been challenged on this account and on account of suspected DIF. Hambleton and de Jong (2003) have addressed the issues and report on progress made. In two doctoral dissertations, colleagues at the University of Jyväskylä have addressed the issue of text authenticity in international reading literacy assessment (Sulkunen 2007) and the problem of translated text equivalence (Arffman 2007).

With the availability of DIF and other analysis options, the EU study is the first one facing the necessity of dealing with problems of international comparisons in L2. It is to be hoped that researchers will take an interest in carrying out more detailed analyses after the main results have been published. Past experience has indicated that there will be such a time pressure to publish the first results that more in-depth analyses cannot be reported at that stage.

Some psychometric challenges

This is a vast area and only a few points will be made.

There is a common wish to aim at *simple solutions* and usually it makes sense to start with simplified models. However, *language ability is a complex phenomenon*. In a monumental analysis of human cognitive abilities, Carroll (1993) identified 16 different abilities in the domain of language. We are justified to seek simplicity (practicality) in language assessment but we should be aware of our simplifications. Reckase (2010) notes that the traditional true-score theory (Classical Test Theory - CTT) and the more recent unidimensional Item Response Theory (IRT) give good approximations for some test situations. However, in other cases, more complex models are needed to accurately represent the relationships in the test data. McDonald (1999) was probably the first to present a general introduction to test theory by introducing *items* as the starting point. This indicates the difference between the modern test theory (known as Item Response Theory) and Classical Test Theory, which focuses on test *scores*. Reckase (2010) considers this a good approach and he suggests that *items* are complicated. He also points out that they deserve careful attention as the quality of assessment is crucially dependent on the quality of items. This is a view that the present author has also espoused for quite some time and has found teaching courses about item writing very enjoyable. Training in item writing needs to be provided for teachers as well as for item writers who are commissioned to write items for examinations.

IRT-based psychometric models are being developed continuously. As far as I can judge from the literature, there is not full agreement on the soundness of their conceptual basis. Not having the competence to assess the merits of the arguments, I am happy to note that a recent contribution by Thomas Eckes (2010) on the many-facet Rasch measurement (MFRM) has been used to study the rating of composition. It is available in the Reference Supplement on the Council of Europe website. Those working on rating issues will benefit greatly from consulting the article.

There is also an increasing trend to use structural equation modeling, SEM, (including confirmatory factor analysis, CFA), e.g. in the study of the structure of motivation. There are some examples of the use of this approach also in assessment. A good example is Åberg-Bengtsson and Erickson (2006), which, among other things, is an interesting and sophisticated analysis of the internal structure of a Swedish national Grade 9 tests.

Conclusion

A perennial challenge (see Masters and Foster 2003) is to guard against placing undue emphasis on a limited range of easily-measured skills at the expense of students' abilities to apply their learning, to reflect on and think about what they are learning, and to engage in higher-order activities such as critical analysis and problem solving. One concrete aspect of this in language testing and assessment is to reflect whether we should, for instance, focus only on testing comprehension of text or also give some attention to learning from text. What would such a test look like?

A similarly persistent challenge is to ascertain where all students are in their learning. If a broad range (say, three CEFR levels) should be distinguished in a test/exam, the assessment tasks need to provide a challenge to, and yield useful information about an equally broad range of proficiency. Valued learning outcomes may require the use of assessment methods which are not common in large-scale assessments/examinations. In principle, this approach is relatively easy to implement/carry out in classroom assessment: direct observations and judgments of students' work and performances over time (e.g. in ELP). Masters and Forster (2003) suggest that having a sufficiently broad coverage of valued outcomes may involve greater use of open-ended tasks that allow students to respond at a variety of levels, or tests that do not require all students to attempt exactly the same items (e.g., tailored tests in which students take items of different difficulty). All this sets high demands on task construction and also on reliability (for two cut scores a reliability of at least .941 is required; Kaftandjieva 2010).

In a recent book, which provides an interesting discussion of assessment use argumentation, Bachman and Palmer (2010) note that there are frequent *misconceptions and unrealistic expectations* about what language assessments can do and what they should be like. They list (a) a misguided belief in one "best" way to test language ability for any given situation; (b) a belief that language test development depends on highly technical procedures and should be left to experts; (c) a belief that a test is either "good" or "bad", depending on whether it satisfies *one particular* quality instead of a number of requisite qualities.

Socio-cultural theory (SCT) is probably a very good candidate to contest views of what is good and bad in current practices in language testing and assessment. In a recent book, Swain and her colleagues (Swain, Kinnear & Steinman 2011) devote a chapter to assessment. They list the following as SCT tenets related to second/foreign language assessment:
- assessment is social and cultural activity
- language performance is co-constructed
- language instruction and language assessment form a dialectical unity of language development

- fairness and equity in language assessment occur during ZPD (Zone of Proximal Development, a concept introduced by the Russian psychologist Vygotsky to indicate the developmental stage that the learner is approaching)

The authors rightly suggest that SCT-inspired assessment will challenge traditional assessment approaches – and create controversies – about validity, reliability, scoring and fairness. Much work is needed to deal adequately with the controversies that are likely to emerge. Kane (2006) will be an indispensible source in this work.

In the European assessment context it is necessary to refer to the extensive and noteworthy work carried out by the Assessment Reform Group (e.g., Harlen 2007, Stobart, 2008).

It also goes without saying that computer adaptive testing will develop and will offer new opportunities as well as challenges.

In conclusion, I wish to cite Davies (2003) who has cautioned about "heresies" of languages testing research, resulting from too enthusiastic embracing of new approaches and leading to loss of proper balance. Therefore, developing *assessment literacy* in a wide sense is a permanent challenge in language testing and assessment. Davies cautions about heresies but he also welcomes them as an antidote to moribund orthodoxy. This is reminiscent of a dictum by Alfred North Whitehead, Russel's teacher and co-author and an endless source of challenging quotes, to the effect that "wherever there is a creed, there is a heretic round the corner or in his grave" (Whitehead, 1993, p. 52).

Author's e-mail: sjtakala@hotmail.com

Notes
1. See e.g. http://europe.eu/epso/index_en.htm.
2. Another little known but very useful early review of the profession is by Stern, Wesche & Harley (1978).
3. www.ealta.eu.org.
4. This section on DIALANG is based on my presentation at the joint IATEFL-EALTA conference on "Putting the CEFR to Good Use", Barcelona, October 29-30, 2010.
5. Actually three different modifications of the modified two-choice Angoff method as well as three different modifications of the contrasting group-method were applied to the standard setting procedure. Multiple matrix sampling with incomplete equal-sized linked design was used to pilot the items. Item response theory was applied to item calibration. The One Parameter Logistic Model (OPLM) was chosen, because it combines the desirable statistical characteristics of the Rasch model with the attractive features of the two-parameter logistic model. Moreover, the OPLM computer program allows application of incomplete test design, which at that time was not possible with most of the other computer programs that applied the IRT approach to test development and analysis. The adaptive test construction design was based on the two-stage multilevel adaptive testing approach. The role of the routing test (pre-estimation) is played by the Vocabulary Size Placement Test and the self-assessment tools. The second-stage language test has three overlapping levels of difficulty. For standard setting, see Cizek & Bunch (2007).
6. See e.g. National Education Policy Center, http://nepc.colorado.edu.
7. It can be claimed that, in spite ten more years of R & R on standard setting, Zieky´s questions have not received a definitive answer and new issues actually emerge all the time.
8. These are cited in full (but leaving out the annotated comments) as standard setting in language testing and assessment is such a topical question in Europe but there seems to be too limited awareness of the *general standards* that apply also in language testing. Living up to these standards is a huge challenge in language testing and assessment.
9. http://cito.com/research_and_development/participation_international_research/ebafls.aspx.
10. http://europa.eu/legislation_summaries/education_training_youth/lifelong_learning/c11083_en.htm.

References

Åberg-Begtsson, L. & Erickson, G. (2006). ´Dimensions of national test performance: A two-level approach.´ *Educational Research and Evaluation*, 12(5), 469-488.

Abrahamson, N. and Hyltenstam, K. (2009). 'Age of acquisition and nativelikeness in a second language: listener perception versus linguistic scrutiny', *Language Learning,* 59/2: 249-306.

AERA, APA, NCME, (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education) (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Alderson, J.C. (2005*). Diagnosing foreign language proficiency. the interface between learning and assessment.* London: Continuum.

Arffman, I. (2007). *The problem of equivalence in translating texts in international reading literacy studies. a text analytic study of three English and Finnish texts used in the PISA 2000 reading* test. Jyväskylä: University of Jyväskylä Institute for Educational Research.

Bachman, L. and Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University Press.

Camilli, G., Cizek, G.J. & Lugg, C. A. (2001). ´Psychometric theory and the validation of performance standards: history and future perspectives.´ In G. J. Cizek (Ed.), *Setting performance standards: concepts, methods, and perspectives*. Mahwah. N.J.: Erlbaum, 445-476.

Carroll, J. B. (1993). *Human cognitive abilities: a survey of factor-analytic studies.* Cambridge: Cambridge University Press.

Cizek, G. J. & Bunch, M. B. (2007)*. Standard setting. a guide to establishing and evaluating performance standards on tests.* Thousand Oaks: Sage.

Council of Europe (2001). *Common European framework of reference for languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.

Council of Europe (2009). 'Relating language examinations to the Common European Framework of Reference for anguages: learning, teaching, assessment: a manual.' Accessed at http://www.coe.int/T/DG4/Portfolio/documents/Manual%20Revision%20-%20proofread%20-%20FINAL.pdf on 23 Nov 2011.

Davies, A. (2003). *The native speaker: myth and reality*. Clevedon: Multilingual Matters.

de Jong, J. H. A. L. & Stevenson, D. K. (Eds.) (1990)*. Individualizing the assessment of language abilities*. Clevedon: Multilingual Matters.

Douglas, D. & Chapelle, C. (Eds.) (1993). *A new decade of language testing research*. Alexandria: TESOL.

Dragemark-Oscarson, A. (2009). *Self-assessment of writing in learning English as a foreign language: a study at the upper secondary school*. Gothenburg: University of Gothenburg.

Eckes, T. (2010). 'Many-facet Rasch measurement. Section H of the reference supplement to the preliminary version of the manual for relating language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment', accessed at http://www.coe.int/t/dg4/linguistic/Source/CEF-refSupp-SectionH.pdf on 11 Nov 2011.

Lindquist, E. F. (Ed) (1951). *Educational measurement*. Washington, D.C.: American Council on Education.

Figueras, N. Noijons, J. (Eds) (2009). *Linking to the CEFR levels: research perspectives.* Arnhem: Cito/EALTA.

Flower, L. & Hayes, J.B. (1977). ´Problem-solving strategies and the writing process.´ *College English* 39/4: 449-461.

Hambleton, R. K. & de Jong, J.H.A.L. (2003). ´Advances in translating and adapting educational and psychological tests.´ *Language Testing*, 20/2: 127-134.

Harlen, W. (2007). *Assessment of learning*. Los Angeles: Sage.

Hughes, A. & Porter, D. (Eds.) (1983). *Current developments in language testing*. London: Academic Press.

Huhta, A., Sajavaara, K. & Takala, S. (Eds) (1993). *Language testing: new openings.* University of Jyväskylä: Institute for Educational Research.

Huhta, A. (2010). 'Innovations in diagnostic assessment and feedback: An analysis of the usefulness of the DIALANG language assessment system.' Unpublished PhD thesis, University of Jyväskylä.

Jaeger, R.M. (1989). ´Certification of student competence.´ In R. I. Linn (Ed.), *Educational measurement. (3rd ed.)*. Washington, D.C.: American Council on Education, 485-514.

Kachru,B. Kachru, Y & Nelson, C. (Eds) (2006). *The handbook of World Englishes*. Malden, MA: Blackwell Publishing.

Kaftandjieva, F. (2010). 'Methods for setting cut scores in criterion-referenced achievement tests: a comparative analysis of six recent methods with an application to tests of reading in EFL.' Accessed at www.ealta-eu.org/resources on 23 Nov 2011.

Kaftandjieva, F., & Takala, S. (2002). Council of Europe scales of language proficiency: A validation study. In J. C. Alderson (Ed.), *Common European Framework of Reference for languages: learning, teaching, assessment: case studies.* Strasbourg: Council of Europe, 105-129.

Kaftandjieva, F., Verhelst, N. & Takala, S. (2000). 'Standard setting procedure'. Unpublished manuscript.

Kane, M. T. (2006). 'Validation´. In R. L. Brennan (Ed*.*), *Educational measurement (4th ed)*. Westport, CT: American Council on Education, 17-64.

Lee, Y. P., Fok, A. C. Y. Y., Lord, R. & Low G. (Eds.) (1985). *New directions in language testing.* Oxford: Pergamon Press.

Masters, G. & Forster, M. (2003). 'The assessments we need.' Accessed at http://cunningham.acer.edu.au/inted/theassessmentsweneed.pdf on 11 Nov 2011.

McDonald, R. M. (1999). *Test theory: a unified treatment.* Mahwah, N.J.: Erlbaum.

Martyniuk, W. (Ed.) (2010). *Aligning tests with the CEFR: reflections on using the Council of Europe's draft manual*. Cambridge: Cambridge University Press.

McNamara, T. (2004). ´Language testing.´ In A. Davies and C. Elder (Eds.), *The handbook of applied linguistics*. Oxford: Blackwell, 763-783.

Messick, S. (1989). 'Validity'. In R. L. Linn (Ed.), *Educational measurement (3rd ed.)*. New York: American Council on Education & MacMillan, 13-103.

Messick, S. (1994). ´The interplay of evidence and consequences in the validation of performance assessments.´ *Educational Researcher*, 23: 13-23

Oscarson, M. (1980). *Approaches to self-assessment in foreign language learning*. Oxford: Pergamon Press.

Papageorgiou, S. (2010). 'Investigating the decision-making process of standard setting participants'. *Language Testing,* 27/2: 261-282*.*

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: The Danish Institute of Educational Research. (Expanded edition, 1980. Chicago: The University of Chicago Press.)

Reckase, M. D. (2010). ´NCME 2009 presidential address: what I think I know.´ *Educational Measurement: Issues and Practice*, 29/3: 3-7.

Shohamy, E. (2001). *The power of tests. a critical perspective on the uses of language tests*. Harlow: Pearson Education.

Spolsky, B. (1978). ´Language testing as art and science.´ In G. Nickel (Ed.), *Proceedings of the fourth international congress of applied linguistics, vol. 3*. Stuttgart: Hochshulverlag, 9-28.

Spolsky, B. (1995). *Measured words*. Oxford: Oxford University Press.

Spolsky, B. (2000). ´Language testing in the Modern Language Journal.´ *Modern Language Journal*, 84/4: 536-552.

Stern, H. H., Wesche, M. B. & Harley. B. (1978). The impact of the language sciences on second-language education. In P. Suppes (Ed.), *Impact of research on education: some case studies*. Washington, D.C.: National Academy of Education, 397-475.

Stobart, G. (2008). *Testing times. the uses and abuses of assessment*. London: Routledge.

Sulkunen, S. (2007*). Text authenticity in international reading literacy assessment: focusing on PISA 2000*. Jyväskylä: University of Jyväskylä.

Swain, M., Kinnear, P. & Steinman, L. (2011). *Sociocultural theory in second language education: an introduction through narratives.* Bristol: Multilingual Matters.

Takala, S., & Kaftandjieva, F. (2000). 'Test fairness: A DIF analysis of an L2 vocabulary test'. *Language Testing*, 17/3: 323-340.

Trim, J.L.M. (2007). 'Modern Languages in the Council of Europe 1954-1997: International co-operation in support of lifelong language learning for effective communication, mutual cultural enrichment and democratic citizenship in Europe', accessed at http://www.coe.int/t/dg4/linguistic/Source/TRIM_21janv2007_%20EN.doc on 11 Nov 2009.

Verhelst, N. (2004). 'Item response theory. Section G in the Reference Supplement to the Pilot version of the Manual for relating language examinations to the Common European Framework of Reference for languages: learning, teaching, assessment', accessed at http://www.coe.int/t/dg4/linguistic/CEF-ref-supp-SectionG.pdf on 23 Nov 2011.

Whitehead, A. N. (1933). *Adventures of ideas.* New York: New American.

Zieky, M. J. (2001). ´So much has changed: how the setting of cut scores has evolved since the 1980s'. In G. J. Cizek (Ed.) *Setting performance standards: concepts, methods, and perspectives*. Mahwah, N.J.: Erlbaum, 19-51.

Foreign Language Testing and Assessment in Greece: An overview and appraisal

[Επισκόπηση και Εκτίμηση της Γλωσσικής Δοκιμασιολογίας και Αξιολόγησης στην Ελλάδα]

Sophia Papaefthymiou-Lytra

The aim of this paper is twofold. First, I will briefly review practices in testing and assessment in Greece; I will refer to current practices, their prospects and long term influence inside and outside school contexts. In particular, I will make brief reference to foreign language (FL) demands outside the school context, namely, the job market and appraise the consequences of the choices made so far for FL education in the primary and secondary school context. Second, I will advocate a way forward by advancing the principle 'learn a foreign language(s) for lifelong use' rather than 'get-trained to get a certificate as early in life as possible'.

ॐ

Στην παρούσα μελέτη αρχικά παρουσιάζω μια επισκόπηση της γλωσσικής δοκιμασιολογίας και αξιολόγησης στην Ελλάδα, κάνοντας αναφορά σε παρούσες πρακτικές και την επιρροή τους εντός και εκτός εκπαιδευτικού περιβάλλοντος. Επίσης, αναφέρομαι στις γλωσσικές ανάγκες εκτός εκπαιδευτικού περιβάλλοντος, συγκεκριμένα στην αγορά εργασίας και στις επιπτώσεις των μέχρι τώρα επιλογών στην ξενόγλωσση διδασκαλία στην πρωτοβάθμια και δευτεροβάθμια εκπαίδευση. Στη συνέχεια προτείνω την επικράτηση της αρχής «εκμάθηση της ξένης γλώσσας για δια βίου χρήση» έναντι της «απόκτησης ενός πιστοποιητικού γλωσσομάθειας όσο το δυνατόν νωρίτερα στη ζωή κάποιου».

Key words: language testing and assessment, performance, Greece, Foreign Language Learning, school context

The background story

In Greece, foreign language (FL) learning was introduced in state secondary education after the independence of the country in the 19th century. In particular, French had been the compulsory foreign language taught up until 1945 when English was also introduced as the second foreign language in secondary education gaining equal status with French.[1] As a result, French was taught almost in half of the schools and English in the other half. In the 1980s, German as a FL was also introduced in secondary education in a selected number of schools. Since 1989, English has become the compulsory FL in primary education too, whereas other languages such as French and German as FL are taught on an optional basis besides English. After the country joined the European Union (EU) in 1979, due to the language policies of the EU, the tendency has been to teach two foreign languages in compulsory education besides the mother tongue.[2]

Ever since FL learning was introduced as a school subject there has always been a need to assess learners' performance. Indeed, in the school context[3], language teachers are currently expected to assess their students. They may use quizzes focusing on the lesson of the day or progress tests that concern a longer teaching/learning period. This period usually coincides with three or more teaching units of the textbook. At the end of the school year teachers administer an attainment test to learners in order to allocate a summative grade to each student in accordance with the assessment scale adopted by the Ministry of Education and Life Long Learning. The same numerical assessment scale is used for all subjects taught in the curriculum of primary or secondary education.[4] This FL grade that appears on school certificates demonstrates that students have successfully passed the subject of English as a FL in the class they had been attending.

Teachers are expected to allocate one summative grade incorporating all skills and abilities while taking into account learners' performance in the classroom; in other words, there is no indication of the actual level learners have reached at that point concerning their skills and competences in the FL. Common practices of testing and assessment by skill, knowledge and ability recommended in the CEFR (Council of Europe, 2001), for instance, have not yet been introduced in the state school system. What is more, the Ministry recommends that teachers in secondary education, in particular, use a specific type of test format, specific types of test items and follow specific guidelines for how the test is to be carried out. The test modes advocated emphasize use of English, reading and writing, in particular. In other words, testing and assessment in schools is strictly regulated by Ministry Decrees. Therefore, language testing and assessment in state school has become a bureaucratic exercise of grade allocation in accordance with prescribed regulations rather than real assessment of learners' skills, abilities and knowledge. This practice has never meant much to stakeholders such as students and parents alike.

The vacuum in language testing and assessment in the country was filled in by foreign institutions, such as the British Council in collaboration with the University of Cambridge Syndicate (UCLES) now called University of Cambridge ESOL Examinations, Alliance Française, the Goethe Institute, the Thervantes Institute, Scuola Italiana to mention but a few, which have administered proficiency tests and issued language certificates for their national languages at various levels of proficiency for a very long time.[5] For a variety of geo-political, diplomatic, economic, socio-political, administrative and other reasons, that are not within the scope of the present paper to explore, the State has recognized these certificates as valid language

certificates, even though it has had no saying in their philosophy, construction and assessment practices. Moreover, it has recognized the C2 level language certificates they issue as demonstrating their holders' ability to teach the respective languages. In this way, the State has equated mere language certificate holders with University degree holders, graduates of the Faculties of Foreign Languages and Literatures who are especially trained at a pre-service level to teach foreign languages! Indeed, as late as 1985, the State allowed University degree holders and language certificate holders with no teacher training qualifications at all to compete for positions in state secondary education. It was Law 1566/85 that made this practice obsolete in state school education.[6]

These certificates issued outside the official educational context have been long used for professional purposes by adults, either as language certificates to seek employment, particularly, in the state sector or as a teaching qualification to seek employment as a teacher of a FL in the private sector. Consequently, getting a language certificate has become a must in the broader Greek society, since it can eventually function as a life long professional qualification.

This recognition very quickly led to the development of a flourishing language learning industry outside the school system all over the country, the so-called 'Language Centres' or 'Frontistiria'. The sole aim of this industry and its main attraction has been to prepare learners to successfully sit FL examinations offered by the above mentioned foreign examination bodies in order to obtain a language certificate for future professional purposes in the shortest possible time, in the case of adult learners, or as early as possible during their school life, if their learners are school age students. As a matter of fact, the practice promoted among school-age students and their parents has been for students to acquire a B2 level certificate in their lower secondary school years in order to secure a language certificate for life and 'get done with foreign language learning for good' as the popular saying goes. As a result, students as young as 12 years old may sit a B2 level exam in English, in particular, which is the compulsory foreign language in the Greek primary school system. Thus, by lowering the age entry level for such adult certificates, students can sit for the C1 or even C2 level English certificate as young as 14 or 15 years old. It is worth noting, however, that all language certificates Greek school students opt to sit for, are General Language Certificates, in spite of the fact that they aim to make use of them later in life as adults for professional purposes.

The success language certification has had among language learners attending courses in language centres has showed the way to the Pan-Hellenic Association of Language School Owners (PALSO)[7] to demand that the State should run its own certification system rather than relying solely on foreign institutions. In the 1980s, PALSO introduced its own four-level examination system for its members hoping that the State would recognize their HIGHER certificate (their highest level exam to be certified) as a 'B2' level language certificate as it has done with the relevant certification systems of external examination boards. However, their aspiration never materialized.

Finally, in 1999 the State officially recognized the need to run its own language certification system. This initiative was funded by the EU to help member states improve foreign language learning, teaching and assessment with the purpose of developing multilingualism in the European Union (Ν 2740/99 ΦΕΚ 186 τ.Α΄). The state-owned certification system abbreviated as "ΚΠγ" (KPG in English, State Certificate of Language Proficiency) was launched in 2003[8]. Unfortunately, the State was not forward thinking enough and did not take the opportunity to

abolish the use of C2 level foreign language certificates as official teaching qualifications. Due to the continuous pressure from University Departments of Foreign Languages and Literatures in Greece to introduce a state owned certification system, which will provide language certificate holders with a teaching qualification rather than a mere language qualification, the KPG has not as yet run its own C2 level examinations in any of the languages it examines and certifies so far. The KGP certifies language proficiency levels (C2 excluded) in accordance with CEFR provisions and reflects the ΑΣΕΠ (ASEP in English, Higher Council for the Selection of Personnel for the Public Sector) requirements set by the Ministry of Decentralisation and E-Government for adults interested to join the civil service but need proof of their FL competence.[9] Nevertheless, the State continues to grant permission to C2 level language certificate holders, issued by other certification bodies as well as to any person who holds a University degree from a FL speaking country, to teach foreign languages in the private sector.

Current practices and prospects

State practices such as those described in the previous section along with parental and student attitudes have had severe consequences on FL education in state schools. As a matter of fact, learners have grown indifferent to the FL work carried out in state school classrooms; instead, students work hard, often under tremendous social and parental pressure, to prepare themselves for a language certificate in the language centres. As mentioned earlier and as teachers attest, students typically aim to obtain a language certificate, preferably at B2 level, before entering upper secondary education where they will need to concentrate on preparing for the University Entrance Examinations. As also mentioned earlier, any of these language certificates, regardless of when they were acquired, are a life-long proof of knowledge of the respective foreign language even if certificate holders have hardly ever used the foreign language for school, academic or professional purposes after obtaining the certificate. As a result, the focus of FL teaching in this country has been to train young and adolescent learners to obtain a certificate rather than to help them learn a language for life long use. Equally importantly, B2 level language certificates and beyond address the needs and interests of adults rather than young learners and adolescents.

It is, therefore, high time we reconsidered FL learning in state schools in the light of current educational research and needs. We need to remind ourselves how foreign languages are learnt, at what age and why they are learnt. As of 2010, FL learning has been introduced in the primary school from Grade 1, albeit in a small number of schools for experimental purposes; the compulsory foreign language taught is English.[10] According to the cross-curricular approach purported in the Cross-thematic Curriculum Framework for Modern Foreign Languages[11], learning a foreign language aims at developing learners' abilities and skills to communicate effectively with others who come from different linguistic and cultural backgrounds as well as to manage information that is derived from various cognitive and scientific fields. In other words, FL learning in the European context aims at promoting literacy, multilingualism and multiculturalism.

Apart from the basic concepts, principles and the ideological orientation, a curriculum also makes reference to "what pupils learn, ... how they learn it, how teachers help them learn, using what supporting materials, styles and methods of assessment, and in what kind of facilities" as Richards argues (2005, p. 39). Assessment, therefore, is one of the important parameters to consider in language education. It functions as an aid to teachers and learners aiming to

promote learning, awareness as well as self-assessment and other-assessment (cf. West and Tsagari, 2004).

On this issue, in a recent paper titled 'Communicating and learning strategies: Two faces of the same coin' (Papaefthymiou-Lytra, 2009), I have argued that foreign language learning should take into account the developing cognitive, linguistic and social capacities of learners as they grow older. Teachers should make use of these developing capacities and encourage learners to become involved in higher order tasks in accordance with their age and general maturity using the foreign language as a means of communication. After all, what language users are able to do with language is related to the language users' age, cognitive, linguistic, emotional and social development. In other words, for successful language learning, content and strategy use in FL learning tasks should respect the learners' age, cognitive, linguistic, social and emotional development (see also Vanderplank, 2008; Veenman *et al.*, 2004 among others).

Moreover, it is important to remember that FL learning is not meant to be merely a school subject for learners with no specific use later on in their lives. In Greece in particular, learners aspire to make good use of a foreign language and the accompanied certification at a later stage in their professional life. In this sense, Greek society is instrumentally orientated towards foreign language learning leaving emotions, such as enjoyment, fun, satisfaction, curiosity and so on, that may derive from learning a foreign language aside.

These assumptions, however, concerning the work-related benefits of FL learning request that we re-examine the kind of English learners are actually required to use in adult life and the kind of communicative competence(s) they need to be able to demonstrate in the foreign language particularly at work. This raises the following question: Is a general certificate of language acquired by holders at the age of 14 or even younger a good enough proof of their ability, skills and competencies later on in life? In my opinion, this is an important question that all stakeholders involved in language teaching and testing should ask themselves and reflect on.

It is common knowledge that mastering a foreign language in the school context takes time and effort; learners are able to demonstrate various competences over time and acquire different sets of knowledge in accordance with their age and overall maturity. This view is also supported by research carried out in bilingual classrooms where the participants are migrant students. Indeed, as Cummins (2001, cited in Lyons 2011, p. 34) argues it takes five to nine years for students to become proficient in English. These migrant students may be fluent in everyday use of language in a couple of years but it takes them longer and requires hard work to master academic English for use in schools and, I would add here, good enough English for the job market later. As students grow up, they are understandably expected not only to deal with more cognitively demanding tasks, but also to use appropriate 'linguistic skills to access higher-order thinking, to interpret, infer, and synthesise information; to pick out the main idea; to relate ideas and information to their background experiences; to recognize the conventions of different genres; and to recognise text structure' (*ibid*.). Therefore it could be argued that Greek learners, who strive to acquire language certificates at such an early age, lack the knowledge, experience, skills and competences of mature foreign language users who can make successful use of the foreign language(s) at work.

Here I would like to add some anecdotal evidence to illustrate my point. A few years back, I received a telephone call by a high ranking official from ASEP. The caller wanted my advice to

address a recurring problem ASEP was facing. As I explained earlier in this paper, those interested to join the ranks of the civil service must take certain examinations organized by ASEP. Examinations vary depending on the duties prospective civil servants will undertake. The requirements are set by the various Ministries or other state-affiliated bodies. Knowledge of a foreign language is usually a requirement. Candidates do not sit exams administered by ASEP in the foreign language(s) they claim they know, but they merely attach the language certificates they once gained at school age to their application as valid proof of foreign language learning and use. However, ministerial officials among others repeatedly complained to the caller that newly appointed civil servants did not use foreign language(s) well enough as had been required. I explained to the caller that this was to be expected. A language certificate gained eight or ten years before entering the civil service or even much earlier reflects the holders' language skills, abilities and knowledge of the foreign language(s) at the time they were examined. If holders had not practiced/used the language ever since, they have lost many of the language skills they once possessed. The only solution to the problem, I suggested to the caller, was for ASEP to declare that certificates need to have been issued over or during the last two or three years at the most before submitting them as evidence of good knowledge of the foreign language.

Assessment, testing and certification: ways forward

Given the current state of language learning and assessment in Greece today, it is high time that we went back to basics. It is necessary to make a clear distinction between assessment, testing and testing for certification as three separate entities that serve different purposes for stakeholders and the society at large. Taking this line of thought as a point of departure, I will also attempt to reposition the role and the status of the FL teacher in the primary and secondary school context.

In the literature, assessment, and in particular alternative assessment, is defined by West and Tsagari, (2004, p. 12), for instance, as the means "to assess and understand student performance in class, identify the specific needs of individual students, tailor instruction to meet these needs, monitor the effectiveness of instruction, and make decisions about advancement or promotion of individual students to the next level of instruction". Testing, on the other hand, West and Tsagari (2004, p. 13) claim "is generally associated with more formal measurement procedures which are carried out at specific times of a school year when all students are usually tested on the same content." Last but not least, testing for certification purposes involve standardized tests often administered country-wide in a specific country or world wide. Assessment and testing as defined above are geared primarily towards educational contexts, be it primary, secondary or tertiary education. Testing for certification purposes, however, is usually the function of external bodies that can serve adults aged 18 and beyond who require language certificates for professional purposes.

In the school context, primary and secondary, the purpose of assessment and, in particular, alternative assessment should be to help learners understand the learning process, overcome the difficulties they face by becoming aware of alternative solutions to remedy problems of production and reception. By doing so, learners become autonomous and independent of teacher intervention. They develop strategies for dealing with problems, but also furthering their possibilities and advancing their knowledge, skills, abilities and competences in the foreign language drawing on self-assessment and other-assessment processes and practices to suit their

lifelong learning goals. In other words, during school years, emphasis needs to be placed on *foreign language learning and language development for lifelong use* rather than on *training students to take an examination*, which often focuses on content and strategy beyond the learners' capacities in accordance with their age. In this way, learners' knowledge, abilities, skills and competences in the foreign language will grow and mature along with their cognitive, linguistic, social and emotional maturity, which comes with age.

As I have argued in Papaefthymiou-Lytra (2011), primary school learners are '*activity*' motivated (see also Papaefthymiou-Lytra, 1987). They want to be actively involved in doing things in the classroom rather than passively receiving information delivered by the teacher or other media. They love to play, move around and have fun with their classmates, explore the world around them and discover new, interesting and challenging ways for doing things, which I have called *activity* motivation.

Similarly, secondary school students, and in particular, the so-called gymnasio students aged 12 to 14/15 of age, are in a transitional stage between childhood and early teens. They notice changes in their bodies and in their mood. They are not always easy to handle, they often become rebellious and disobedient. What they are interested in mostly and are motivated by is to explore and understand themselves, their relationships with others and the world around them, which I have called '*exploring thyself and the world at large*' motivation (cf. Papaefthymiou-Lytra, 2011). Still it is at that crucial age when parents and the society at large puts pressure on them to work hard not to learn foreign languages but to get trained in order to obtain a language certificate that they will be able to use some time in the future. As a result, during those crucial years of their emotional and personality development, emphasis is not placed on learning foreign language(s) for lifelong use as an alternative way to handling personal matters, understanding oneself, discovering and understanding others and the world at large. Instead emphasis is placed on training students to pass a language examination offered by external bodies that will secure them a certificate for future use. Thus, learning a foreign language has become devoid of the fun aspect of learning. It has become a series of repetitive 'drills': mostly working on past papers in a language centre or with a tutor so as they can master the necessary testing strategies to pass B2, C1 or C2 level examinations that are intended for adults and not for their age and interests. No wonder as soon as they manage to obtain a language certificate, they are not interested at all in advancing their learning of the language any more.

Pupil and parental attitudes that prevail among low secondary school students create similar problems in upper secondary education too. Teachers claim that Upper Secondary or Lykio students are not interested in FL learning any more as they are preoccupied with University Entrance Exams. What is more, the students who have secured a language certificate usually at a B2 level feel that they have completed the study of FLs (*teacher attestations-personal communication*). As a result, upper secondary school students are not engaged in advancing and refining their knowledge and use of the foreign language(s) they have been learning in understanding differences in attitudes, stances and beliefs of writers or speakers, for instance (cf. Garidi, 2011).

Rather than merely getting trained to pass a certificate examination, it is more conducive to successfully learning a foreign language for lifelong use if students, depending on their age, become involved in such activities as project work, role playing and one-act plays, sing-along

and pair work tasks that can also involve use of the Internet and other multi-media which allow for collaborative learning (cf. Vlachos and Papaefthymiou-Lytra, 2008). The present generation of students is much more knowledgeable concerning the use of the Internet and multi-media than their teachers. Here teachers are presented with the opportunity to learn from their students as their learners learn from them; teachers can allocate them duties and responsibilities that will make them proud of and willing to work hard. Adopting learning-teaching approaches, such as task-based learning and CLIL (Content and Language Integrated Learning), will help teachers come up with their own repertoire of learning activities that are appropriate for their own students and respond well to local circumstances and needs.

Besides strategies for learning and advancing the language as well as coping with difficulties in perception and production of the FL, a good way to expand foreign language for lifelong use is through extensive reading and listening. This practice has been undervalued in or deprived from schools for quite some time now – most probably because school libraries are an exception rather than the rule in most Greek state schools. It is important, therefore, that teachers try to instill in learners good extensive reading and listening habits (cf. Papaefthymiou-Lytra, 1984).

More specifically, teachers can start reading to young learners, for instance, original children's books as well as simplified books appropriate for their age that they can find in abundance in bookshops. Children's stories, such as Aesop's fables, Andersen' stories, Greek myths etc. that learners are familiar with already are ideal to start with. It will be a good idea to try to build up a class 'library' with English books of this kind.[12] A good way to go about it is by asking students to bring an English book they have read and would like to share with others. The next step is for students to start borrowing books from the class 'library' – this can be a project run by learners themselves while the teacher supervises the whole process. Students are asked to keep a record of the title of the book and the date they borrowed it. Eventually, they can add information concerning what the book was about and how much they liked it and why. The next big leap forward is to ask learners to write a summary of the story. Similar procedures can be followed for listening practice. The Internet can provide learners and teachers alike with a plethora of authentic listening material that can be exploited in the classroom and out of the classroom.

As for secondary school learners, emphasis should be placed on learning practices similar to those mentioned above crossing over to other subjects of the curriculum too. Content-wise the learning material should reflect the students' interests, needs, cognitive, linguistic and emotional maturity among others rather than training-learners-to-learn certificate practices. In this way, schoolteachers become learners' collaborators and facilitators for learning purposes rather than trainers for testing purposes. The different kinds of learning activities briefly outlined for school learners need to be combined with alternative assessment practices such as self- and peer-assessment and other testing practices such as, quizzes, short progress tests, attainment tests etc. that aim to make learners become proactive in taking responsibility for their own learning.[13]

In short, FL education needs to refocus on FL learning from a learning and assessment perspective, whereas the FL learning process should evoke emotions of fun, satisfaction, curiosity, fulfillment, anticipation and happiness among others. Following Sternberg (1999), positive emotions release learners' imagination, creativity and inventiveness, which may lead learners to express in the foreign language "a mental entity that has never been represented before, an idea that has never been expressed before" (Vega Moreno, 2007, p. 5). This feeling of

pleasant surprise, fulfillment and joy makes the communication of thoughts in the L2 worthwhile for older and younger learners alike.

Conclusions

In this paper I have argued that foreign language learning for lifelong use has been subsided by a 'get-trained-to-get-the-certificate' ideology. Inordinate weight has been placed on FL testing for certification purposes by all stakeholders rather than learning for lifelong use. Similarly, assessment and testing practices in schools in the form of placement, progress, attainment tests and so on have followed a testing rather than a learning orientation. As a result, teachers and learners alike have been primarily interested in the product of learning rather than the process of learning, which will allow students to become autonomous and independent learners for life. I have also presented some of the ways this situation may be remedied. The longer learners study a foreign language the better lifelong users they will be.

This is particularly true for English which learners start learning at a very early age. We must give learners time and space so that the foreign language grows as they grow cognitively, linguistically, socially and emotionally. In schools, we must highlight foreign language learning for life long use rather than the short-sighted testing for certification purposes. This attitude will also reinstate teachers as educators rather than mere trainers. And once our students have learnt the language they can easily train themselves - with minimal help from a teacher if need be - in the mechanics of test taking. The important thing is to enjoy using the foreign language(s) for life.

After all, foreign language certification, and, in particular, language certification for specific purposes, can be a professional qualification for adults provided that they have acquired it close to the time of employment and forthcoming use as an indication of what competences they have currently acquired in the foreign language. After all, all higher level certificates, namely, B2, C1 or C2 level certificates are attuned to the cognitive development as well as the needs and interests of adults rather than adolescents or young learners.

Author's email: spapaef@enl.uoa.gr

Notes

1. For a historical account about English as a FL in Greek education, see Μαλιβίτση (2004). For a survey of the geopolitical and other reasons that have helped English as a *lingua franca* to gain ground in Greek education and society, see Σηφάκις (2012).
2. Concerning EU and plurilingualism see the Language Policy Division of the Council of Europe: a Platform of resources and references for plurilingual and intercultural education – language across curriculum (http://www.coe.int/t/dg4/linguistic/Schoollang_EN.asp).
3. English is a compulsory subject taught in lower secondary education, the so-called 'gymnasio' attended by learners aged 12-15 as well as the first two years of upper secondary school the so-called 'lykio' attended by learners aged 15-17. A more detailed description of the situation is beyond the scope of the present paper.
4. The scale that has been used in upper and lower secondary education ranges from 01 to 20. The highest grade is 20 whereas the passing grade is 10. Any grade below 10 is a fail. In primary education for the 3rd and the 4th grade a letter assessment scale has been introduced ranging from A to D, D is a

fail. For the 5th and the 6th grade the scale ranges from 01-10. Any grade below 5 is a fail. A fail grade is scarcely ever granted in primary education unless there are severe learning problems. There is no grading scale of either type for Grades 1 and 2 in the primary school.

5. Ever since the introduction of the CEFR (Council of Europe, 2001), these institutions have conformed to its requirements and language proficiency levels. For a comparison of these certificates across languages and their stakeholders, see http://www.ALTE.org.

6. Use of these certificates as teaching qualifications originate in the decrees AN 1100/1938 and A.N. 752/1945. The latter, in particular, states that as soon as there would be trained university graduates to teach foreign languages in the school system, this law would become obsolete. Although the Faculties of English Language and Literature in Athens and Thessaloniki were established in 1951 (cf. Παπαευθυμίου-Λύτρα *et al.* 2008) the aforementioned law became obsolete in 1985 (Law 1566/85). However, certificate holders can still teach in the private sector in the so-called Language Centres or Frontistiria where they can compete for positions with University graduates.

7. For more information about PALSO visit www.palso.gr

8. For more information about the 'ΚΠγ' visit http://www.kpg.ypepth.gr/

9. (ASEP) ΑΣΕΠ - Ανώτατο Συμβούλιο Επιλογής Προσωπικού. For more information about ASEP visit http://www.asep.gr/

10. English as a FL was introduced from the 3rd grade of the primary school onwards in 1989; for experimental purposes the teaching of English has been introduced to selected schools from the 1st grade in 2010.

11. Διαθεματικό Εννιαίο Πλαίσιο Προγραμμάτων Σπουδών Ξένων Γλωσσών (Cross-thematic Curriculum Framework for Modern Foreign Languages). For more information visit http://www-pi-schools.gr Also see ΦΕΚ 303 και 304/13-3/2003 *Προγράμματα Σπουδών για την Αγγλική Γλώσσα στην Α/θμια και Β/θμια Εκπαίδευση.*

12. As mentioned, school libraries are usually the missing link in Greek schools. So it is advisable that teachers start their own class library for ELT books. They can invite their pupils to bring English books to class for the class library which they can share with their classmates during the school year. Pupils take their books back home at the end of the year.

13. Although it is beyond the scope of this paper, it is worth noting that the situation in tertiary education is equally bleak. As the limited research on FL learning in tertiary education shows, students are expected to attend language courses for special purpose (LSP) or academic purposes (LAP). Wherever in higher education FL learning is mandatory, a summative grade is allocated, as is the case in secondary education, slotting in all skills and abilities. Tests usually assess the reading and writing modes only. As a result, tertiary education students who have acquired a language certificate(s) have very little interest in advancing their knowledge, skills and ability of the respective foreign language(s) (cf. Psaltou-Joycey and Kantaridou, 2009; Sifakis, 2006; also Παπαευθυμίου-Λύτρα, 1990a, 1990b). After all, it is the language certificate(s) they may already hold that constitutes proof of knowledge rather than their having studied foreign languages at university level. Of course, this is the case particularly for state sector employment and less so for private sector employment. It is true that small business may rely on language certification when they look for employees. However, corporations, banks etc. invite candidates for an interview in the foreign language they are expected to be able to use efficiently later on at work.

References

Cummins, J. (2001). *Negotiating identities: education for empowerment in a diverse society (2nd ed.).* Los Angeles, CA: Association for Bilingual Education.

Council of Europe, (2001). *Common European Framework of References for languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.

Garidi, E. (2011). 'News articles and critical discourse analysis: a case study'. Unpublished MA Dissertation, University of Athens.

Lyons, Z. (2011). 'The English language support programme materials for secondary schools'. *NALDIC Quarterly*, 8/2: 33-40.

Μαλιβίτση, Z. (2004). 'Το μάθημα της αγγλικής γλώσσας στην ελληνική δημόσια εκπαίδευση: η θέση του μαθήματος και οι διδακτικές προσεγγίσεις του, όπως παρουσιάζονται στη σχολική νομοθεσία.' Αδημοσίευτη Μεταπτυχιακή Εργασία, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης.

Papaefthymiou-Lytra, S. (1984). 'Reading to children, reading by children'. In S. Efstathiadis *et al.* (Eds.), *Communicative language teaching and teaching methodology, vol. 3*. Thessaloniki: University Studio Press, 235-246.

Papaefthymiou-Lytra, S. (1987) *Language, language awareness and foreign language learning.* Athens: The University of Athens Press.

Papaefthymiou-Lytra, S. (2009). 'Communicating and learning strategies: Two faces of the same coin'. *Journal of Applied Linguistics,* 25: 27-46.

Papaefthymiou-Lytra. S. (2011). 'Creativity and conformity in language teaching: the teacher's perspective'. *Aspects Today*, 32: 8-13.

Παπαευθυμίου-Λύτρα, Σ. (Επιμ.) (1990α). *Οι ξένες γλώσσες στην τριτοβάθμια εκπαίδευση: ΑΕΙ – ΤΕΙ.* Αθήνα: Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών.

Παπαευθυμίου-Λύτρα, Σ. (1990β). ΄Ξένες γλώσσες στο πανεπιστήμιο – προβλήματα και προτάσεις.' Στα Πρακτικά του συνεδρίου *Το Πανεπιστήμιο στην Ελλάδα Σήμερα: Οικονομικές, Κοινωνικές και Πολιτικές Διαστάσεις.* Αθήνα: Πάντειο Πανεπιστήμιο - Ιδρυμα Σάκη Καράγιωργα, 238-248.

Παπαευθυμίου-Λύτρα, Σ. *et al.* (2008). *Τμήμα Αγγλικής Γλώσσας και Φιλολογίας, 56 χρόνια παρουσίας.* Αθήνα: Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών.

Psaltou-Joycey A. & Kantaridou, Z. (2009). 'Plurilinguialism, language learning strategy use and learning style preferences'. *International Journal of Multilingualism*, 6/4: 460-474.

Richards, J. (2005). *Curriculum development in language teaching*. New York: Cambridge University Press.

Sifakis, N. C. (2006). 'Fundamental principles, historical perspective and classification of ESP'. In N. C. Sifakis & V. Rizomilioti, *The teaching of English for specific purposes: vol. 1—the field, needs analysis and language issues*. Patras: HOU, 19-140.

Sternberg, R. (Ed.) (1999). *Handbook of creativity.* New York: Cambridge University Press.

Σηφάκις Ν. Κ. (2012). *Αγγλική γλώσσα και παγκοσμιοποίηση: όψεις της σύγχρονης πραγματικότητας στην Ελλάδα, στην Ευρώπη και στον υπόλοιπο κόσμο.* Αθήνα: Ηρόδοτος.

Vanderplank, R. (2008). 'The significance of first language development in five to nine year old children for second and foreign language learning'. *Applied Linguistics,* 29/4: 717-722.

Veenman, M. V. J., Wilhelm, P. & Beishuizen, J. J. (2004). 'The relation between intellectual and metacognitive skills from a developmental perspective'. *Learning and Instruction*, 14: 89-109.

Vega Moreno, E. R. (2007). *Creativity and convention: the pragmatics of everyday figurative speech*. Amsterdam/Philadelphia: John Benjamins.

Vlachos, K. & Papaefthymiou-Lytra, S. (2008). 'Collaboration through asynchronous online networking and the learning of foreign languages'. In L. A. Tomei, (Ed.). *Encyclopedia of information technology curriculum integration, vol. A.* Hershey, PA: IGI GLOBAL, 63-70.

West, R. & Tsagari, D. (2004). *Testing and assessment in language learning. (vols. I, II, III).* Patras: Hellenic Open University.

Bullets in their (heads) CV: The construction of the 'subject' by English language examination boards and Greek private language schools

[Σφαίρες στα (κεφάλια) CV τους: Η κατασκευή του 'υποκειμένου' από τις εξεταστικές επιτροπές Αγγλικών και τα φροντιστήρια]

Vanda Papafilippou

This paper explores the ways that the subjectivity of the Greek test-taker is constructed through the discourses employed by examination boards and private language schools. By employing Critical Discourse Analysis I examine the online texts uploaded on the websites of the examination boards that administer exams in Greece recognised by ASEP and advertising material of these examination boards and private language schools. The analysis illuminates that the dominant discourse operating is this of knowledge economy and English language tests appear to be dressed with neoliberal ideology. Therefore, these tests form and promote certain subject positions not only regarding ourselves as learners but also as citizens. The subjects seem to be constituted by their 'achievements' and their main aspiration appears to become successful employees, who have all the skills and qualifications employers need. Hence, this paper is an attempt to encourage all stake-holders to develop a critical view of tests as well as to question and critique the values that are inherent in them.

<p align="center">ဆ</p>

Η εργασία αυτή διερευνά τους τρόπους με τους οποίους ο Έλληνας υποψήφιος σε εξετάσεις Αγγλικών κατασκευάζεται ως υποκείμενο μέσω του λόγου που χρησιμοποιείται από τους εξεταστικούς οργανισμούς και τα φροντιστήρια στην Ελλάδα. Χρησιμοποιώντας την κριτική ανάλυση του λόγου, εξετάζω τα κείμενα στις ιστοσελίδες όλων των εξεταστικών οργανισμών που παρέχουν πιστοποιήσεις αναγνωρισμένες από το ΑΣΕΠ, το διαφημιστικό υλικό των εξεταστικών οργανισμών, καθώς και αυτό από διάφορα φροντιστήρια. Η ανάλυση υποδεικνύει ότι ο επικρατών λόγος είναι αυτός της κοινωνίας της γνώσης και ότι οι εξετάσεις των Αγγλικών είναι επενδυμένες με τη νεοφιλελεύθερη ιδεολογία. Έτσι, οι εξετάσεις αυτές φαίνεται να κατασκευάζουν και να προωθούν συγκεκριμένες υποκειμενικότητες, που έχουν να κάνουν όχι μόνο με το άτομο ως μαθητή, αλλά επίσης και με το άτομο ως πολίτη. Τα υποκείμενα εμφανίζονται να αποτελούνται τα «επιτεύγματά» τους και η κύρια φιλοδοξία τους φαίνεται να είναι να επιτύχουν ως εργαζόμενοι οι οποίοι

έχουν όλες τις ικανότητες και τα προσόντα που χρειάζονται οι εργοδότες. Ως εκ τούτου, το παρόν άρθρο είναι μια προσπάθεια να ενθαρρύνει όλα τα ενδιαφερόμενα μέρη να αναπτύξουν μια κριτική άποψη για τις εξετάσεις των Αγγλικών καθώς και να ενεργήσουν βάσει αυτής, αμφισβητώντας και κρίνοντας τις αξίες που είναι συνυφασμένες με αυτές.

Key words: Critical language testing, impact, subjectivity, knowledge economy, discourse, neoliberalism

---

Introduction

Hanson (2000, p. 68) argued that "the individual in the contemporary society is not so much described by tests as constructed by them" as the subjectivity[1] and social identity of the test-taker is realised only in the test itself (Hanson, 2000; McNamara and Roever, 2006). Tests, then, construct certain pictures of learners/test-takers (McNamara, 2001), and we, by allowing ourselves to be subjected or by subjecting others to them, we validate them (Torrance, 2000). English language tests in particular combine two sources of power: the English language, with its colonial past and linguistically imperialist present (Holborow, 1999; Pennycook, 1994; Phillipson, 1994; Phillipson and Skutnabb-Kangas, 1996) and tests, thus bringing together two extremely powerful institutions, creating a new institution (Shohamy, 2001), with significant symbolic power, that is, "a power of constructing reality" (Bourdieu, 1991, p. 166). So, through their symbolic power, English language tests impose certain categories of thought and perception, present the current social order as legitimate, while at the same time they constitute a major criterion of worth, quality and value (McNamara and Shohamy, 2008).

Greece is considered one of the biggest markets for English language teaching and testing enterprise mostly for three reasons: the great demand for English language certificates from the job market (Hamp-Lyons, 2000; Tsagari, 2009) the promotional role of international institutions, examination boards and private language; and the European Council's recommendations for 'mother-tongue plus two languages' (Little, 2007).

This article explores how the subjectivity of the test-taker is constructed through the discourses employed by examination boards and frontistiria. By employing Critical Discourse Analysis (CDA), I examine the texts on the websites and advertising material of private English language schools and all the examination boards recognised by ASEP[2], exploring the discourses *around* and *promoted by* them and the respective subjectivities these discourses form, in order to reveal the ideological and political character of these tests. Following Critical Language Testing as introduced by Shohamy (2001), I challenge the 'knowledge' tests are based upon as well as the 'knowledge' that is produced by them and interrogate the social character and political functions of language tests.

To explore the effects of discourse and the ideologies entailed on the construction of the *subjectivity* of the test-taker, I conceive the 'self' as constructed, validated and offered through discourses that are available to individuals at a particular time and place (Pavlenko and Blackledge, 2004). Through the hegemonic discourses -which are mediated via institutions- people are classified into categories, into pre-existing identity-options (Butler, 1993) thus being rendered 'subjects' (Foucault, 1977a).

Institutions, discourse and ideology

For Foucault, institutional structures are *a means* that power uses and contribute to the production of normalised and docile subjects (Caputo and Yount, 1993). Althusser's (1971, p. 7 italics in original) view about institutions coincides with Foucault's, as he claimed that institutions –or in his terms Ideological State Apparatuses- like school, or in our case, educational institutions such as examination boards and frontistiria, "ensure *subjection to the ruling ideology* or the mastery of its 'practice'. So, in order to research how subjects acquire a certain perception of the world and of themselves, a certain ideology, we need to examine the role of institutions that "organise, manage or propagate such cognitions" (van Dijk, 1998, p. 186), as different organisations appear to have vested interests in different constructions of certain identities (Ainsworth and Hardy, 2004).

Thus, the questions that this paper addresses are: Who are we as people for whom English language tests have become so important? What are the emerging subjectivities that are constructed by examination boards and private English language schools? What subjection do they ensure?

Methodology

Discourse plays an important role in the processes of 'making up' people (Ainsworth and Hardy, 2004) as it imposes the existing rules of subjectification and thus a certain way of thinking and conceiving the world (Femia, 1981). Discourse analysis involves an interest in how the various classifications are brought into being and what the effects are "for those targeted by these categories as well as those involved in their construction" (Ainsworth and Hardy, 2004, p. 240), thus challenging the existing 'natural' and 'obvious' categories, such as the test-taker. One of the 'fields' that discourses operate are texts. Fairclough (1995) sees texts as *social spaces* in which cognition, and social interaction occur simultaneously. Thus, texts have causal effects both immediate, as they can change our knowledge, values, attitudes and beliefs, and longer-term, as they can contribute to (re)shaping our identities (Fairclough, 2003). CDA in particular, attempts to detect how elite groups define 'reality', 'objects' and 'subjects', in our case test-takers, and the role of discourse in the (re)production of dominance (van Dijk, 1993). Hence, CDA will allow us to understand the process of construction of the category of the test-taker as well as the meanings, ideological and cultural assumptions that are attached to it. And as Ainsworth and Hardy (2004, p. 243) argue, by understanding the processes of identity construction "we are better placed to understand the way in which constructions of identity constrain individuals, as well as the prospects for resistance".

My approach to CDA does not claim 'objective' findings and as de Certeau (1984, p. 170) wrote "the text has a meaning only through its readers; it changes along them", thus, in textual analysis subjectivity is inevitable. I also believe that if we are to pursue social justice we should do it without imposing our (predetermined) notions of emancipation. The approach evolves cyclically and entails eleven 'steps' that are the fruit of a combined reading of the work on discourse of Foucault (1977b; 1977c; 1981; 1984a; 1984b; 2000), Gramsci (1971), Barthes (1975; 1981), de Certeau (de Certeau 1984) and Fairclough (2003):
1. Visual analysis
2. Identifying the discourses operating in the text (Which are the main parts of the world/areas of social life that are represented? Which are the main themes? From which perspective are they represented?)

3. Ideology/hegemonic values (Which words/phrases carry a particular ideological content? What propositions are neutralised? What conception of the world is presented? What is presented as 'common sense'?)
4. Modes of existence of the discourse(s) (the specific socioeconomic structures that gave birth and developed this discourse, the relationship of 'meaning' and power, the different groups that are implicated, where these discourses have been used, how they can circulate, who can appropriate them and their impact on Greece)
5. Delimitation of 'objects' and 'subjects' (How are 'objects' and 'subjects' defined? What are their characteristics?)
6. Credibility of the discourse(s) (How credible is this discourse? How is this credibility achieved?)
7. The emerging relations of power
8. The regime of truth this discourse belongs to and the apparatus(es) of power this discourse fits
9. The macrostructure (historical background and the current socio-economic background)
10. 'Silences' (What is silenced?)
11. Signifiance[3]

However, due to space limitations I will focus only on the findings relevant to the construction of the test-taker as a subject, namely, visual analysis; discourses operating; ideology; credibility of discourse; delimitation of the 'subject', and the emerging power relations.

Materials

In order to explain the 'logic' behind English language tests as presented by the institutions that administer these tests as well as identify the discourses that operate around and are promoted by these tests and the English languages within the Greek society, I composed a corpus of texts (size: 40,534 words). These texts came from the websites of all the English language tests acknowledged by ASEP[4] as well as from brochures and the websites of Greek English language schools (language school chains, such as Stratigakis Ison, as well as local ones). The reader can see the detailed list in Appendix I.

Analysis and discussion

The documents analysed are apparently directed at the wider public, and most directly to test-takers, teachers, parents and employers.

Visual analysis

Both the leaflets and the websites use eye catching colours, headings and layout. The images employed are these of happy parents and students, but also of happy and successful employees and devoted students.

Hence, the subjectivity of the test taker begins to take a certain form: the test-taker is presented as a (future) high knowledge-skilled, successful *entrepreneur*, albeit implicitly *classed* and *raced*, as people appear to be predominantly white and to have a high position in an enterprise. A subjectivity that is further fortified by the texts.

Image 1 Hellenic American Union website Image 2 TOEIC leaflet in Greek
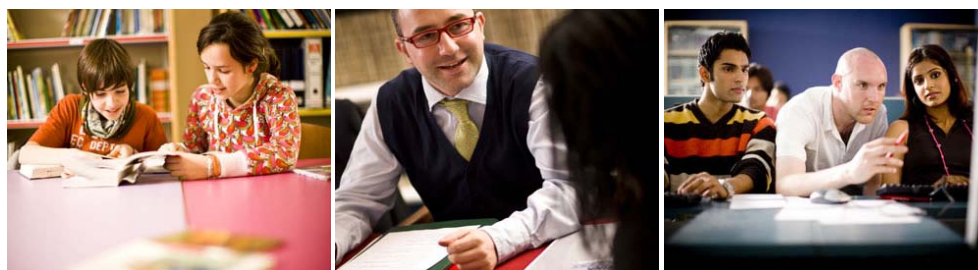
Images 3, 4, 5 British Council, Greece website

Discourses operating in the text

The discourse that appeared to be operating in all the texts was no other but this of *globalised knowledge economy* operated in all the texts. In knowledge economy, individuals are viewed primarily as knowledge workers who, through education or training, have to acquire the necessary *skills* in order to raise their productivity and thus their future income (Becker, 1993). Therefore, knowledge is believed to be equal with prosperity (Graham, 2002). Knowledge functions as a central capital and the crucial means of production in society, and the emerging 'knowledge workers' are of major importance in an economy which is mostly knowledge-based (Drucker, 1969). For this reason, "[i]n a knowledge society, school and life can no longer separate" (*ibid.*, p. 24), and as a consequence, there is increased emphasis on life-long learning. In all the texts analysed, the job market and tertiary education are presented as the main areas of social life and English language tests are presented as the key to being successful in these domains:

> Cambridge ESOL offers the world's leading range of certificates for learners and teachers of English - taken by over 3 million people in 130 countries. They help people gain entrance to university or college, improve job prospects or measure progress in English. More than 11,000 employers, universities and government bodies around the

> world recognise Cambridge ESOL qualifications. (Cambridge ESOL website: http://recognition.cambridgeesol.org/our-exams)

> IELTS is the world's proven English test. Over 1.4 million candidates take the test each year to start their journeys into international education and employment. (IELTS website: http://www.ielts.org/)

The main themes that emerged were how to find a job and enhance one's career prospects, further education, life-long learning and labour mobility.

> Win the key of success in your career with the acquisition of a qualification in Business English. (Easy learning, local frontistirio, leaflet)

> Learners who struggle with communicating in English are unlikely to be able to achieve their true potential in their studies or at work. By improving their English skills, these learners stand a much better chance of progressing. (Edexcel website: http://www.edexcel.com/quals/esol/esol-life/Pages/default.aspx)

> It has been specifically designed for students who wish to study in a third level institution and for professionals who wish to enhance their professional profile by demonstrating a competence in English (Tie website: http://www.tie.ie/)
> The certificate of language proficiency is viewed as:
> - work qualification and professional credential
> - attestation to literacy
> - instrument for lifelong learning
> - passport for educational and professional mobility inside and outside the European Union (E.U.)
> (KPG, Rationale:
> http://rcel.enl.uoa.gr/rcel/texts/Rationale_and_Ideology_of_the_KPG_Exams.pdf)

Finally, all the texts appear to adopt mainly the employer's perspective, even when the test discussed does not assess English in the workplace (for example TOEIC) but is promoted by the exam provider as a general proficiency test.

> Employers and educational institutions are familiar with the level of language skills and personal achievement that they represent and a Cambridge ESOL Certificate can help you to find work in Greece or study abroad. (British Council website, Greece: http://www.britishcouncil.org/greece-exams-cambridge-recognition.htm)
> Test content reflects real-world tasks and provides you with the information you need to easily:
> - Recruit, place and promote the most qualified employees
> - Identify job-training requirements
> - Assign employees to positions overseas
> (TOEIC website: http://www.ets.org/toeic/listening_reading/about)

> A City & Guilds qualification is proof that you have the right skills to do your job well- which is why so many employers look for people who have one. You can be confident your qualification is well respected within your industry. (City & Guilds website: http://www.cityandguilds.com/48281.html)

Credibility of discourse

De Certeau (1984, p. 148) wrote that "the *credibility* of a discourse is what first makes believers act in accord with it", in other words, it produces practitioners. One 'device' that

renders a discourse credible is a *story*, as in this way the text makes people believe that "this text has been dictated for you by Reality itself" (*ibid*.). Such a story is this told by the following test-takers:

> 'Preparing for the CAE was a pleasant process through which we can take an acknowledged *ptychio*[5] which will help us to have a better job. In addition, we will have an improved level of studies with prospects to go abroad.'
> Mary Linardou, Greece (Cambridge CAE leaflet in Greece)
>
> 'I am working in an international environment which requires me to continuously improve my English. To pass the FCE at Grade B is certainly a commitment to that. After the exam I got the motivation to study English more and then decided to enrol in an MBA conducted in English.'
> Phan Hoang Hoa (Cambridge ESOL website:
> http://www.cambridgeesol.org/exams/fce/index.html)
>
> 'The TOEIC test helps us quickly and efficiently to find elite candidates who meet our language requirements and helps our employees establish a good foundation for follow-up training.'
> Danny Dong, Senior Recruitment Manager Motorola, Beijing Branch (TOEIC website:
> http://www.ets.org/toeic/succeed)

As we can see in Mary's story, a student's story, is an expression of what is considered *common sense* at least in Greece, that a 'ptychio' in English will later enable you to find a better job (hence my question: Better than others?) and to go abroad (hence another question: Is abroad the Promise Land?). The story of Phan Hoang, an employee's story, on the other hand, verifies the globalisation and competitiveness of the labour market, as it is the market that demands workers to *constantly* improve themselves and actually, *be committed to* that. And lastly, Danny Dong's story, a manager's story, the voice of authority, simply confirms that if test takers want to be considered as "elite candidates", who meet the requirements of the labour market, they should have certified competency in English; in other words, one of those who control the labour market confirms what is considered common sense.

Ideology

As far as the ideological content of the texts is concerned, the analysis pointed clearly at neoliberal ideologies. Rose (1999, p. 141) argues that within neoliberal regimes "all aspects of social behaviour are now reconceptualised along economic lines". For Treanor (2001), the idea of *employability* is characteristically neoliberal, as "neoliberals see it as a moral duty of human beings, to arrange their lives to maximise their advantage on the labour market", to conform and respond to the needs of the market forces (Gordon, 1991). Therefore, it is the *subject's responsibility* to respond to the needs of the job market. However, it is not any kind of employment that is considered as 'good', as the *entrepreneur* has a particularly privileged status (Rose, 1999; Simons and Masschelein, 2008; Treanor, 2001) as "[i]nnovation is expected to follow spontaneously from the liberation of the animal spirit of individual entrepreneurs" (Jessop, 2002, p. 260). Lastly, according to Fotopoulos (2001, p. 240) neoliberalism views globalisation as "the inevitable result of technological and economic changes" and as beneficial because it allows "healthy competition develop".

Indeed, by employing nominalisation that is, the representation of processes as entities (Fairclough, 2003), the globalised and competitive job market appears to be the omnipotent

force, an uncontrolled phenomenon acting in its own right, controlling our desires, our needs and in the long term, our lives (Jarvis, 2007).

> In today's competitive job market, certification of knowledge and skills plays an important role for a person's personal and professional development. (Hellenic American Union website: http://www.hau.gr/?i=examinations.en.exam-overview)

> The general manager of the group, G. Stratigakis, greeted those who were present and congratulated students, parents and teachers while he underlined the importance of the competence of foreign languages in the competitive world of nowadays. (Info magazine, Stratigakis group)

As a consequence of this globalised world, English has become absolutely necessary not only in order to communicate but also to be able to participate in the –globalised knowledge-society and above all, to be employed: a need of the market.

> In the 21st century, a good level of spoken English will be an important part of any career progression and personal development. (Trinity College London website: http://www.trinitycollege.co.uk/site/?id=368)

> In the recent years there have been made great changes due to economic globalisation and internationalisation of the labour market. These changes created an environment of increasing demands for proficiency in foreign languages, particularly English. (Lingua+com, local frontistirio, website: http://www.lingua.com.gr/pub/cmsserv/default.aspx?cid=3)

> A qualification in English can open doors to jobs all over the world, in trade, tourism, communications and with international companies (City & Guilds website: http://www.cityandguilds.com/3446.html)

> Edexcel's easy-to-administer ESOL qualifications develop the literacy and communication skills that speakers of other languages need to take part fully in work and society. (Edexcel website:http://www.edexcel.com/quals/esol/Pages/default.aspx)

The only goal that a person appears to have, or is constrained to have, is to study in order to be later employed preferably in a "leading" organisation[6], an international company, to materialise in other words the neoliberal dream of becoming 'somebody' (Walkerdine, 2003), or the Greek dream of finding a job in the public sector, as all frontistiria mainly stressed that they are preparing their students for certificates that are acknowledged by ASEP. However, from the analysis, education, or rather skills acquisition, appears to be the employee's responsibility, an *investment* that one has to make in order to be rewarded later by the job market.

Delimitation of 'subjects'

The best rendition of the emerging subject is perhaps the collage employed by the City & Guilds leaflet in Greek (Image 6).

Image 6 City & Guilds leaflet in Greek

Test-takers seem to be products of their 'achievements', as it is only those with a certificate that will eventually be employed as these certificates appear to "open the door to a world of opportunities"[7] in the public or the private sector. These certificates are presented to be our 'right hand', as shown in the image 6, as without them we will not be able to do anything. People seem to take these exams mainly in order to "get a good education or [work] in another country to fast-track [their] career"[8]. Hence, the emerging subject is a highly educated and ambitious person, who wants to study, travel, work and live abroad -and most probably abroad is defined as an English-speaking country, unless the whole world is perceived to exclusively speak English. For this reason perhaps, the collage uses a travel pillow as head, in order to fully illustrate that the workplace has indeed become international. However, all these prestigious qualifications and skills we gain seem to be aiming only at a prosperous career:

> Proving your business English abilities can open the door to career opportunities with a new employer, or can make your ambitions for promotion or career development within your current organisation a reality. (BEC, Cambridge website: http://www.cambridgeesol.org/exams/bec-vantage/index.html#tab2)

> Learners who struggle with communicating in English are unlikely to be able to achieve their true potential in their studies or at work. By improving their English skills, these learners stand a much better chance of progressing. (Edexcel website: http://www.edexcel.com/quals/esol/esol-life/Pages/default.aspx)

We base our lives on money, as our legs, as the collage shows, are made of coins, money that is gained by travelling around the world (the other leg is depicted as a newspaper and a compass).

Furthermore, as Brine (2006) argues, knowledge economy discourses construct two types of – lifelong – learners: those with high knowledge skills and those with low or no knowledge skills. In the case of English language tests, learners are divided into those who can communicate in English and those who do not. However, the divisions do not stop here: those who speak English are further divided into three categories: those who have a

certificate, those who do not have a certificate and those who do not have the 'right' certificate that will enable them to be *internationally competitive*:

> I trust TOEIC so as my qualifications will be acknowledged by EVERYBODY and FOREVER. It is my passport for the global job market. I don't want a test that is only acknowledged in Greece.
> I want TOEIC for REAL and INTERNATIONAL RECOGNITION**.** (TOEIC leaflet in Greek, capitals in original)

In neoliberal economies, competition –between individuals, firms or nations- is held to be a primary virtue (Harvey, 2005) and students, as future workers, "must be given the requisite skills and dispositions to compete efficiently and effectively" (Apple, 2006, p. 32).

However, the test-taker is not devoid of ethnic characteristics. Subjects appear to be divided into two categories: native speakers of English and non-native speakers of English, as, if one is not a native speaker of English, s/he has to be particularly competent in this language.

> Learners who struggle with communicating in English are unlikely to be unable to achieve their true potential in their studies or at work. (Edexcel, website: http://www.edexcel.com/quals/esol/esol-life/Pages/default.aspx)

Hence, the discourses of these tests confirm the discursive binary division between native and non-native speakers, as all these people who do not have English as their 'native language' are automatically labelled as the 'Other', the 'non-native speaker' (Pennycook, 1994). But the issue of race and discrimination does not stop here.

Pennycook (1994, p. 14) has argued that the English language, and in particular English language certificates (Hamp-Lyons, 2000), function as a gatekeeping device to positions of prestige in a society, thus becoming "one of the most powerful means of inclusion into or exclusion from further education, employment, or social positions". After all, Bourdieu and Passeron (1990, p. 162) characterised the examination as unequalled in imposing and legitimating certain social hierarchies as "behind its technical function of producing qualifications", it serves a social function of legitimating *class differences*. However, in all the texts analysed, these tests appear to be available to all, as if everybody has the –financial- ability to learn English and take these tests. As Harvey (2005, p. 68) notes though, in neoliberalism, "[T]here are presumed to be no asymmetries of power or of information that interfere with the capacity of individuals to make rational economic decisions in their own interests". Nevertheless, in reality, the hazards for certain social groups (e.g. immigrant and working class children, children with special educational needs) are high (Madaus and Horn, 2000). In the case of Greece in particular though, where students rely exclusively on frontisteria for their language education, we can assume that automatically groups such as immigrants or people that belong to low socioeconomic classes, cannot afford it and thus they are automatically excluded by the *appearingly* meritocratic knowledge society. Therefore, at least in the case of the Greek test-taker, there are three additional characteristics: s/he is not an immigrant, s/he has no learning disability and comes from a middle or higher socioeconomic background.

Relations of power

Regarding power relations, there seems to be a top-down approach to power, in the sense that what 'commands' us is the market, then come the institutions which 'give us the

opportunity' to correspond to the needs of the market, and at the bottom, us, the test-takers. But even at test-takers' level, the nature of power relations is again hierarchical: those who succeeded (οι επιτυχόντες) and those who fail, those whose names will be put outside the private English language school as an indication of success and those who will be in no such list (see example of such a list below):

Image 7 A list of the students who succeeded in English language exams at the entrance of a private language school

But the hierarchisation of people according to their qualifications continues after their school years, as the more skills and qualifications you have, the better and more prestigious and prosperous job you will have, the higher you are in the neoliberal social hierarchy. So we can conclude that the relation-to-self associated with this subject is this of *self-government*, as learning and professional success appears to be individual responsibility, a commitment, and one has to 'work upon the self' and invest in human capital and to add value to the self so as to be included (Simons and Masschelein, 2008). To this contribute also the disciplinary discourses that are integrated in the examination procedure, as certain institutions explain clearly how to prepare for the exams, how to take the stress out, and clear regulations of how the examinees should behave (examples in Appendix II). Thus, the emerging subject is no other but the *entrepreneurial subject*, a subject that is situated in an environment to which s/he has adapt proactively and creatively in order to satisfy her/his needs and gain an income (*ibid*.).

Concluding remarks

The aim of this article has been to explore how the subjectivity of the test-taker is being constructed by examination boards and private English language schools. What we are told is that test-takers are mostly characterised by their 'achievements', their qualifications; that those who possess the 'right' qualifications will later become successful entrepreneurs, as they will have all the skills that the labour market demands; that they can compete with the rest of the world for a job in a 'leading' organisation in any country of an Anglophone world; that the test-taker is no one else but the self-governed entrepreneurial subject. However, the CDA has also showed that the discourse adopted was that of *knowledge economy* and English language tests appear to be dressed with *neoliberal* ideology. Thus, market exchanges become an ethic in itself "capable of acting as a guide to all human action, and substituting for all previously held ethical beliefs" (Harvey, 2005, p. 3), ensuring in this way

our subjection to the needs of the labour market. Neoliberalism locates subjects in a fiercely competitive globalised world rendering them responsible for their success or failure, as it assumes that in all our behaviour we act as self-interested individuals (Peters 1996) who have the same access to the same information and resources, thus precluding "asymmetries of power or information that interfere with the capacity of individuals to make rational economic decisions in their own interests" (Harvey, 2005, p. 68).

However, the analysis presented in this article suggested that the binary classifications of the test-taker into those who can communicate in English and those who cannot, into those who have the appropriate certificate and those who do not, is *classed* and *raced*. Hence, behind the neutral and meritocratic façade, the discourse employed, this of knowledge economy, or as Brine (2006, p. 663) put it 'lifelong learning'

> is a discourse of competition, of personal striving, of constant becoming, of inclusion and exclusion, of stratification that continues to (re)construct educational and labour market power relations based on gender, class and race, and on disability, age and migrant/citizen status.

Neoliberalism appears to have become "hegemonic as a mode of discourse", as it is incorporated in the common-sense, in the way we interpret, live in and understand the world (Harvey, 2005, p. 3), thus granting consent "to the general direction imposed on social life by the dominant fundamental group" (Gramsci, 1971, p. 12). But by consenting to the hegemonic ideology, we *subject to* the needs of the market forces. Market forces though, are perceived to be intensified by intensifying assessment, as even within a contract period, an employee is subject to continuous assessment (Treanor, 2001) like in the case of Phan Hoang. Thus, the entrepreneurial subject becomes *perpetually candidate* for a 'good life' as defined in neoliberal terms (O'Flynn and Petersen, 2007), taking into consideration that the present socio-economic and political system urges us to think of ourself as *perpetually* inadequate for the market, *perpetually* in need to prove our 'value', our skills and be committed to it.

But is there a way out of this discourse? Are there prospects of resistance? Can we define ourselves outside the realm of names, of qualities and imposed desires given to us by social and educational institutions and the interests of power (Lemke, 2008)? Of course, the solution is not to change the texts on the websites of examination boards or the advertising material of frontistiria, as the repression of discourse will actually lead to its multiplication, to an explosion of different discursivities (Foucault, 1998). As Foucault (1977b, p. 14) has noted, "[T]he problem is not one of changing people's 'consciousness' or what's in their heads, but the political, economic, institutional regime of the production of truth". Therefore, if we want to change the present situation, we should start to question and critique English language tests and the values that are inherent in them, the way we perceive others and ourselves *in relation to* these tests, and the socio-economic and political system that has established these tests as necessary. After all, according to Elana Shohamy (March 2009, personal communication), tests are not like the moon and the sun; we are not born with them. And neither are we born with the neoliberal labour market and its demands on individuals and societies.

Author's email: edxvp@bristol.ac.uk

Notes

1. By subjectivity/subject I mean the subject positions (re)produced within a particular discourse. These subject positions are seen not only as a product of a particular discourse and at the same time subjected to its meanings, power and regulations.
2. ASEP (ΑΣΕΠ) stands for the Supreme Council for Civil Personnel Selection (in Greek: Ανώτατο Συμβούλιο Επιλογής Προσωπικού).
3. Signifiance is a term coined by Barthes (1981) and refers to how the text relates to our personal concept of 'self', to how a text's language, its discourses- work us and undo us as the reader, when we 'enter' it, enter as opposed to 'observe' it. The term invokes the idea of an "infinite labour" of the signifier upon itself, the endless play of the signifiers, thus opening the way for other approaches and interpretations. In my analysis, in order to show how I felt when I 'entered' the text, I wrote a free-verse poem that can be found in Appendix C.
4. For a detailed list of the exams recognised by ASEP, see: http://www.asep.gr/asep/site/home/LC+Menu/FORIS/Ipodigmata/prok.csp
5. Ptychio (πτυχίο) in Greek means 'university degree'. However, that is how the majority of Greeks call English language certificates, something that indicates the importance that is attributed to these certificates by the Greek society. It is really interesting to note though, that the same word (ptychio) is used in the advertising leaflet of Cambridge ESOL as this shows that Cambridge ESOL is promoting indirectly its certificates as 'ptychia'.
6. Cambridge ESOL website: http://www.aul.org.uk/listings?os=4&b=all&br=1&ns=0&lr=1
7. Cambridge ESOL website: http://www.cambridgeesol.org/institutions/index.html
8. Cambridge ESOL website: http://www.cambridgeesol.org/sector/study-work-abroad/index.html.

References

Ainsworth, S. & Hardy, C. (2004). "Critical discourse analysis and identity: why bother?" *Critical Discourse Studies*, 1/2: 225-259.

Althusser, L. (1971). *Lenin and philosophy and other essays*, New York: Monthly Review Press

Apple, M. W. (2006). *Educating the "right" way: markets, standards, god, and inequality*. New York: Routledge.

Barthes, R. (1975). *The pleasure of the text*. New York: Hill and Wang.

Barthes, R. (1981). "Theory of the text", in R. Young, (Ed.), *Untying the text: a post-structuralist reader*. Boston: Routledge & Kegan Paul, 31-47.

Becker, G. S. (1993). *Human capital: a theoretical and empirical analysis, with special reference to education*, Chicago: Chicago University PRess.

Bourdieu, P. (1991). *Language and symbolic power*, Oxford: Polity Press.

Bourdieu, P., and Passeron, J.-C. (1990). *Reproduction in education, society and culture*. London: Sage Publications.

Brine, J. (2006). "Lifelong learning and the knowledge economy: those that know and those that do not-the discourse of the European Union." *British Educational Research Journal*, 32/5: 649-665.

Butler, J. (1993). *Bodies that matter: on the discursive limits of "sex"*. New York: Routledge.

Caputo, J., and Yount, M. (1993). "Institutions, normalization and power". In J. Caputo & M. Yount, (Eds.), *Foucault and the critique of institutions*. University Park, Pennsylvania: The Pennsylvania State University Press, 3-26.

de Certeau, M. (1984). *The practice of everyday life*. Berkeley and Los Angeles: University of California Press.

Drucker, P. (1969). *The age of discontinuity: guidelines to our changing societ.* New York: Harper and Row.

Fairclough, N. (1995). *Critical discourse analysis: the critical study of language*. Harlow: Pearson Education Limited.

Fairclough, N. (2003). *Analysing discourse: textual analysis for social sciences*, London: Routledge.

Femia, J. V. (1981). *Gramsci's political thought*. Oxford: Clarendon Press.

Fotopoulos, T. (2001). "Globalisation, the reformist Left and the anti-globalisation 'movement'." *Democracy & Nature: The International Journal of Inclusive Democracy*, 7/2: 233-280.

Foucault, M. (1977a). *Discipline and punish: the birth of prison*. London: Penguin Books.

Foucault, M. (1977b). "The political function of the intellectual." *Radical Philosophy*, 17: 12-14.

Foucault, M. (1977c). *Michel Foucault: language, counter-memory, practice*. Oxford: Basil Blackwell.

Foucault, M. (1981). "The order of discourse". In R. Young, (Ed.), *Untying the text: a post-structuralist reader*. Boston: Routledge & Kegan Paul, 48-78.

Foucault, M. (1984a). "Truth and power", in P. Rabinow, (Ed.), *The Foucault reader: an introduction to Foucault's thought*. London: Penguin, 51-75.

Foucault, M. (1984b). "What Is an author?". In P. Rabinow, (Ed.), *The Foucault reader*. London: Penguin, 101-120.

Foucault, M. (1998). *The history of sexuality: the will to knowledge*. London: Penguin.

Foucault, M. (2000). "Interview with Michel Foucault". In J. D. Faubion, (Ed.), *Michel Foucault: power / essential works of Foucault, vol III*. London: Penguin, 239-297.

Gordon, C. (1991). "Governmental rationality: an introduction". In G. Burchell, C. Gordon, & P. Miller, (Eds.), *The Foucault effect: studies in governmentality*. London: Harvester Wheatsheaf, 1-51.

Graham, P. (2002). "Hypercapitalism: language, new media and social perceptions of value." *Discourse and Society*, 13/2/: 227-249.

Gramsci, A. (1971). *Selections from the prison notebooks (edited and translated by Quintin Hoare and Geoffrey Nowell-Smith)*. New York: International Publishers.

Hamp-Lyons, L. (2000). "Social, professional and individual responsibility in language testing." *System*, 28: 579-591.

Hanson, F. A. (2000). "How tests create what they are intended to measure". In A. Filer, (Ed.), *Assessment: social practice and social product*. London: Routledge/Falmer, 67-81.

Harvey, D. (2005). *A brief history of neoliberalism*. Oxford: Oxford University Press.

Holborow, M. (1999). *The politics of English*, London: Sage.

Jarvis, P. (2007). *Globalisation, lifelong learning and the learning society: sociological perspectives*. London: Routledge.

Jessop, B. (2002). *The future of the capitalist state*. Cambridge: Polity.

Lemke, J. L. (2008). "Identity, development and desire: critical questions". In C. R. Caldas-Coulthard & R. Iedema, (Eds.), *Identity Trouble: critical discourse and contested identities*. Houndmills: Palgrave Macmillan, 17-42.

Little, D. (2007). "The Common European Framework of Reference for languages: perspectives on the making of supranational language policy." *Modern Language Journal*, 91/4: 645-655.

Madaus, G. F. & Horn, C. (2000). "Testing technology: the need for oversight". In A. Filer, (Ed.), *Assessment: social practice and social product*. London: Routledge/Falmer, 47-66.

McNamara, T. (2001). "Language assessment as social practice: Challenges for research." *Language Testing*, 18/4: 333-349.

McNamara, T. & Roever, C. (2006). *Language testing: the social dimension*, Malden: Blackwell Publishing.

McNamara, T. & Shohamy, E. (2008). "Language tests and human rights." *Journal of Applied Linguistics*, 18/1: 89-95.

O'Flynn, G. & Petersen, E. B. (2007). "The 'good life' and the 'rich portfolio': young women, schooling and neoliberal subjectification." *British Journal of Sociology of Education*, 28/4: 459-472.

Pavlenko, A. & Blackledge, A. (2004). "Introduction: new theoretical approaches to the study of negotiation of identities in multilingual contexts". Clevedon: Multilingual Matters Ltd, 1-33.

Pennycook, A. (1994). *The culture politics of English as an international language*, London: Longman.

Peters, M. (1996). *Poststructuralism, politics and education*, Westport, Connecticut: Bergin & Garvey.

Phillipson, R. (1994). "English language spread policy." *International Journal of Sociology of Language*, 107(1): 7-24.

Phillipson, R. & Skutnabb-Kangas, T. (1996). "English only worldwide or language ecology?" *TESOL Quarterly*, 30/3: 429-451.

Rose, N. (1999). *Powers of freedom: reframing political thought*, Cambridge: Cambridge University Press.

Shohamy, E. (2001). *The power of tests: a critical perspective on the uses of language tests*, Harlow: Pearson Education Limited.

Simons, M. & Masschelein, J. (2008). "The governmentalization of learning and the assemblage of a learning apparatus." *Educational Theory*, 58/4: 391-415.

Torrance, H. (2000). "Postmodernism and educational assessment" In A. Filer, (Ed.), *Assessment: social practice and social product*. London: Routledge/Falmer, 173-188.

Treanor, P. (2001). "Neoliberalism: origins, theory, definition" accessed at http://web.inter.nl.net/users/Paul.Treanor/neoliberalism.html on 15 Dec 2011.

Tsagari, K. (2009). *The complexity of test washback,* London: Peter Lang.

van Dijk, T. (1993). "Principles of critical discourse analysis." *Discourse & Society*, 4/2: 249-283.

van Dijk, T. A. (1998). *Ideology: A multidisciplinary approach*, London: Sage.

Walkerdine, V. (2003). "Reclassifying upward mobility: femininity and the neo-liberal subject." *Gender and Education*, 15/3: 237-248.

Appendix I

The texts came from the following sources:
Tests accredited by ASEP

- Cambridge ESOL, CPE, CAE, FCE, PET, KET:
  http://www.cambridgeesol.org/index.html
- Cambridge ESOL, BULATS:
  http://www.bulats.org/
- Cambridge Business English Certificate:
  http://www.cambridgeesol.org/exams/bec-higher/index.html
- IELTS:
  http://www.ielts.org/default.aspx
- MICHIGAN ELI:
  http://www.cambridgemichigan.org/
- London Tests of English:
  http://www.londonexams.gr/
- EDEXCEL:
  http://www.edexcel.com/quals/esol/esol-life/Pages/default.aspx
- Integrated skills Trinity College:
  http://www.trinitycollege.co.uk/site/?id=369
- City & Guilds ESOL international:
  http://www.cityandguilds.com/46753.html?search_term=esol
- Ascentis (former OCNW) Certificate in ESOL:
  http://www.ascentis.co.uk/
- ESB Certificate in ESOL international all modes:
  http://www.esbuk.org//content/Home.aspx
- Michigan State University (MSU-CELP):
  http://www.msu-exams.gr/swift.jsp?CMCCode=1002&extLang=
- Test of Interactive English:
  http://www.tie.ie/index.htm
- TOEIC:
  http://www.ets.org/toeic
- EDI (JETSET):
  http://www.edigreece.gr/edigreece/
- KPG:
  http://www.kpg.minedu.gov.gr/%CE%91%CF%81%CF%87%CE%B9%CE%BA%CE%AE
  %CE%A3%CE%B5%CE%BB%CE%AF%CE%B4%CE%B1.aspx
- Advanced Level Certificate in English (ALCE) Hellenic American University:
  http://www.hauniv.us/?i=hau-uni.en.alce
- UCLAN:
  http://www.uclan.ac.uk/information/services/sas/examinations_and_awards/examinations.php

Frontistiria (advertising material from brochures, websites and magazines published by these language schools)

- Easy learning
- Lingua com
- Estia
- Dana Languages
- Barbara Costa
- Gnosi
- Tachtatzis-Chatzaki
- Stratigakis Ison

Appendix II

From the Hellenic American Union website
(http://www.hau.gr/?i=examinations.en.home):
How you can prepare for the exams(http://www.hau.gr/?i=examinations.en.how-you-prepare)

- Become familiar with the structure of the test. Learn more about its sections and the number of questions that you will be asked to answer.
- Become familiar with the duration of the test. Practice without breaks and follow closely the guidelines and instructions provided.
- Follow the instructions, directions and suggestions of your teachers.

Be prepared for the actual test conditions:

- Mark your answers with pencil only.
- Mark all your answers on the answersheet, not in the test booklet.
- Do not make any other marks on your answer sheet.
- If you change your mind about an answer, erase your first mark completely.
- Fill in one circle for each problem.
- If you are not sure about an answer, you may guess.
- In the test booklet you can only keep notes or underline the parts you believe might help you answer correctly the questions posed.

Remember:

- Working on past papers or mock tests does not replace systematic work on language.
- Succeeding in past papers or mock tests does not guarantee success on the actual test.

Examination Regulations (http://www.hau.gr/?i=examinations.en.examination-regulations)

Examinees' conduct

- Any effort on the part of examinees to claim or gain an advantageous position or receive preferential treatment over other examinees before, during, or after the examination is strictly prohibited.
- The behavior and conduct of examinees during the examinations should be appropriate and under no circumstances should it disturb the rest of the examinees or the examiner or proctors in the examination room.
- On the day of examination, examinees are required to have a valid Greek or Cypriot ID, or passport, or a temporary ID issued by the Municipal or Police Authorities bearing a stamped photograph. Candidates who report to the test center without one of the above documents will not be allowed to enter the examination room.
- Examinees are required to report to the test center at least 30 minutes prior to the beginning of the examination.
- Examinees are not allowed to choose their seat. Seating will be assigned by the examiners. The Hellenic American Union Examinations Department staff in charge of the examination room reserve the right to change examinees' seating arrangement during the examination.
- Examinees are required to remain in the examination room throughout the examination.
  Examinees may not leave the examination room without the Examiner's permission. Examinees that have the Examiner's permission to leave the

examination room will be escorted by one of the proctors, who will remain with them for the duration of their absence and will escort them back into the room. The test materials of these examinees will be submitted to the Examiner and will be returned to them once they are back in the examination room.

- At the end of the examination, examinees must submit their completed answer sheets to the proctors, along with all test materials.

Appendix III

Doors

Some big people told me
That a certificate opens doors
To a world of opportunities

So I dreamt that I could manage to open these doors
-If I managed to find them-
And explore this new world
Of chance
Of prospects of success

So I got the certificates
The keys
I have a whole bunch of them
(I'd rather have a bunch of flowers instead)

But no doors were to be found

Only more bullets in my CV

So big people could compare me with others
With more
Or less
Bullets
In their
(heads)
CVs

And I competed
With the whole world
Or so they told me
And I got a job
And tried to please my bosses
With my skills
The skills they asked for

But when I clock off
I go for walks

To find the doors

KPG oral exams: task design considerations and actual performance

[Προφορικές εξετάσεις στο ΚΠΓ: προβληματισμοί σχετικά με το σχεδιασμό και την εκτέλεση των δραστηριοτήτων]

Eleftheria Nteliou

This paper presents a task analysis model that has been designed for the oral tasks of the Greek State Exams for foreign language certification (known as KPG). The theoretical background of the model is based on the systemic functional approach to language use as well as on the notion of genre as both process and product. The purpose of this paper is to show how the model has been applied to describe and analyze the tasks designed for the KPG oral exams at levels B1 and B2. The analysis of the oral task rubrics shows that the generic process specified by the task is likely to lead test takers to particular lexicogrammatical choices, which, in turn, cause differences in the expected language performance. Implications for future research, language teaching and testing are mentioned.

⸎

Στο παρόν άρθρο παρουσιάζεται ένα μοντέλο γλωσσολογικής περιγραφής και ανάλυσης των προφορικών δραστηριοτήτων που χρησιμοποιούνται στις κρατικές εξετάσεις πιστοποίησης γλωσσομάθειας (γνωστές, εν συντομία, ως ΚΠΓ). Το θεωρητικό υπόβαθρο του μοντέλου βασίζεται στη συστημική λειτουργική προσέγγιση χρήσης της γλώσσας, καθώς επίσης και στην αντίληψη ότι το κειμενικό είδος δεν είναι μόνο τελικό αποτέλεσμα αλλά και διαδικασία. Ο σκοπός του άρθρου είναι να παρουσιάσει την εφαρμογή του μοντέλου στην περιγραφή και ανάλυση των προφορικών δραστηριοτήτων στις εξετάσεις του ΚΠΓ, στα επίπεδα Β1 και Β2. Η ανάλυση των εκφωνήσεων των προφορικών δραστηριοτήτων δείχνει ότι η κειμενική διεργασία είναι πιθανό να οδηγήσει τους υποψηφίους σε συγκεκριμένες λεξικογραμματικές επιλογές, οι οποίες, με τη σειρά τους, μπορεί να προκαλέσουν διαφορές στην αναμενόμενη παραγωγή λόγου. Σχολιάζεται, επίσης, η σημασία της παρούσας μελέτης για τον τομέα της διδασκαλίας και εξέτασης των γλωσσών καθώς και για μελλοντική έρευνα.

Key words: task demands, generic process, lexicogrammar

Introduction: oral task characteristics and test performance

In the construct-based approach to assessment, the theoretical description of the language ability to be measured is the first important step to test design, because it normally influences task requirements and the assessment criteria on which actual performance is evaluated (Luoma, 2004, p. 42). In fact, the match between what is intended to be tested (i.e. the construct) and what is actually produced and measured during test performance, constitutes a 'validity argument', which needs sufficient evidence to be sustained (Fulcher and Davidson, 2007). Therefore, in test-task validation procedures, tasks play a central role, because the language output they generate constitutes evidence that the test actually measures what it intends to measure.

In oral testing, the central role of 'task' and its direct influence on language performance can be seen in Graph 1 which presents a conceptual framework for performance testing. This visual representation was initially proposed by Milanovic and Saville (1996) and later reproduced by O'Sullivan et al (2002, p.35). The framework portrays the various factors (or facets) that should be taken into consideration when designing a test, because they are likely to affect performance and threaten test reliability and validity. According to the framework, the task is directly related to test specifications and the theoretical description of the language intended to be tested (i.e. the construct). Moreover, the task affects the sample of language produced by the candidates, which is evaluated by the examiners and leads to a final score. Based on O'Sullivan et al's framework, Graph 1 schematically presents how the task relates to the test construct as well as to language performance and assessment.
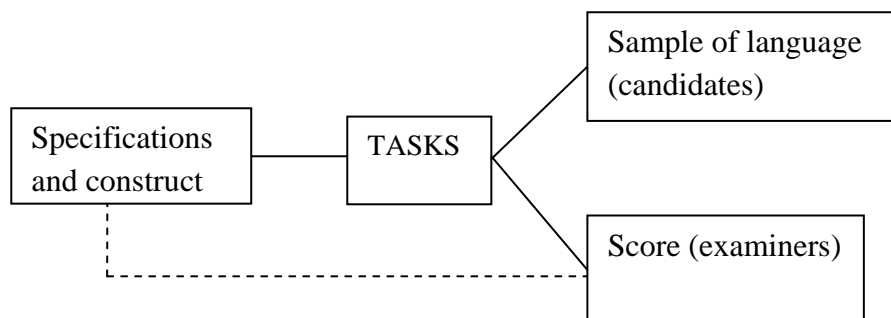
Graph 1. Tasks and their influence on performance.

This paper focuses on the oral tasks designed for the B1 and B2 level speaking module in the KPG[1] exams in English and presents a model that was used to describe and analyze the speaking tasks in terms of their linguistic demands at the level of oral language production. Apart from their influence on language performance, what is also worth examining is how the task demands relate to the theoretical construct of language ability, as described in the KPG exams specifications (ΥΠΕΠΘ, 2003). The focus of the research is on levels B1 and B2 because exams at these two levels attract a great number of candidates, given that language knowledge certification can also be used as a job qualification, especially at level B2. Moreover, if there is a clear picture of what a B1 and B2 level speaker is expected do with language, we can also make sound inferences about task design and performance at lower and higher levels (i.e. A1-A2 and C1-C2). The next section presents the research background, which relates to tasks and their influence on language production, as well as the theoretical linguistic background on which task design in the KPG exam system is based.

Literature review

The assumption that language output is greatly affected by task design has been widely investigated in the oral testing literature. Studies have shown that different types of tasks (e.g. role play, narration) lead to the elicitation of different language characteristics during oral performance (e.g. Bygate, 1999; Kormos, 1999; Pavlou, 1997; Shohamy, 1988; Young, 1995). O'Sullivan (2008) offers a comprehensive review of the literature on oral tasks and their influence on language production, from a number of perspectives that do not only relate to task types but also to test methods (e.g. 'live' versus 'tape mediated' oral tests) and specific intra-task characteristics, such as planning time and level of difficulty. Because tasks greatly influence language output on which a score is based, Bygate (2009, p. 414) concludes that 'to appraise students' language, we first need to understand the linguistic demands of our tasks'.

The analysis of task demands is likely to lead to expectations or predictions relating to actual performance. O'Sullivan et al (2002) note that there is scarce research on relating test performance to the test designers' predictions or expectations, which are based on their definition of the language construct. For that reason, they first analyzed the construct of spoken ability in the Cambridge ESOL Main Suite speaking tests in terms of a list of language functions. Based on this analysis, they then developed observation checklists, which were used in the *a priori* and *a posteriori* analysis of speaking task output. Using this method, they tried to validate the match between intended and actual test taker language. One of the implications of the study was that the analysis of task content and prediction of the actual performances elicited by test tasks define the construct tested. The prediction of linguistic responses can greatly help in the process of task design, because it can "lead to a greater understanding of how task and task formats can be manipulated to result in specific language use" (O'Sullivan et al., 2002, p. 47).

In order to explore the relationship between expected and actual task performance, we first need to be aware of the construct of language ability, which forms the basis for task design and determines the demands regarding language production. In the case of the KPG exams, the definition of language ability is based on a functional approach to language use, whereby language is viewed as a social phenomenon, taking place within a particular context of communication, which determines the language choices participants make to achieve certain communication purposes (ΥΠΕΠΘ, 2003, p. 31). As Karavas states, the main aim behind test design in the KPG exams is "to evaluate socially purposeful language knowledge and literacy" (2009, p.24).

The functional approach to language use, which is adopted by the KPG exam system, is mainly described in the work of Halliday (Halliday, 2002; Halliday and Hasan, 1976, 1985; Halliday and Mathiessen, 1999, 2004), who has laid the foundation of what is known as 'Systemic Functional Linguistics' (SFL). SFL is a theory which emphasizes the importance of context in making meaningful interactions. In contrast to traditional grammar, according to which language is strictly realized through a set of rules of form and structure, SFL underscores the importance of the context of communication, which should determine language choices. The communication context, which refers to 'the linguistic, physical and psychological dimensions of the situation in which language is used' (Karavas, 2009, p. 25), determines the types of functions that the participants should perform (e.g. to inform, to advise, to persuade, etc.) as well as the production of different text types, which are called 'genres'. In her systemic functional approach to language use, Eggins (1994, p. 30)

distinguishes two types of context, which influence linguistic behaviour: the 'context of situation', which actually determines the register variables of the text (e.g. topic, participant roles, the role of language) and the 'context of culture' that describes the overall purpose of the text with which a genre is associated.

Several SFL applications can be found in the educational context, especially in literacy instruction in English as a second language, most notably in the work of "Sydney School" in Australia (Martin, 1998). However, its relation to assessment procedures has been scarcely investigated. There are some studies that have shown how the SFL principles can effectively be applied in second language learning for classroom assessment procedures (e.g. Huang and Mohan, 2009; Hughes, 2009). Nevertheless, no relation between SFL and the linguistic theory behind high stakes exams has ever been reported. The KPG examination system must be the first and unique to base its description of the language construct and test specifications on the principles of SFL. In particular, this study intends to show how the theoretical approach to SFL can influence task design and lead to specific expectations regarding oral production, focusing on the KPG speaking tests.

Research context: The KPG B1 and B2 speaking modules

The Greek certification system of foreign language proficiency was introduced in April 2003. Tests are designed for A1-A2, B1-B2 and C1 levels, as they are set by the Council of Europe and described by the Common European Framework of Languages (CEFR; Council of Europe, 2001). Regarding the KPG oral exams in English, level B1 and level B2, which both have exactly the same format and structure, were introduced in November 2003 and in May 2007 respectively[2]. Test takers are examined in pairs, but they do not interact with each other. There are two Examiners, one of whom acts as an Interlocutor (i.e. he assigns the tasks to each candidate and marks their performance) and the other is an Examiner-Assessor, who is silent during the oral exam and just marks performance.

Candidates are examined in their ability to respond to personal questions (Activity 1), to talk about something based on one or more visual prompts (Activity 2) and to relay in English information from a Greek text (Activity 3). What differs from level to level are the task demands and the complexity of test input (e.g. the Greek texts used as input in the mediation activities). Moreover, the description of the illustrative descriptors[3] for oral production presents differences in the quality of the expected language output between these two levels. In general, the B1 level tasks and questions are intended to be cognitively and linguistically less demanding than the ones at level B2 and, therefore, the quality of language production at these two levels is likely to differ, at least in terms of complexity.

In particular, KPG oral task designers take into account the expected differences in the B1 and B2 level speakers' speaking skills, which are described in the KPG specifications for these levels (Dendrinos, 2007). The KPG specifications for oral ability at levels B1 and B2 are slightly adapted from the description of oral language ability in the respective levels, provided by the CEFR. Moreover, oral task design incorporates characteristics of the SFL approach to language use. Since, according to SFL, language is a means to make meaning with words within a specific contextual frame, the task designers in the KPG exams should be careful in providing a clear context and purpose of communication as well as specific participant roles. This type of information is usually provided in the input material (e.g. photos, texts) as well as in the task rubrics. In order to investigate the link of task design with task performance in the KPG exams, a tool has been designed, whose purpose is to describe

the linguistic features of expected performance, on the basis of the theoretical considerations that govern task design in the KPG exams.

Relating KPG task demands to performance: methodology adopted

The University of Athens launched a large-scale research project in 2007, whose aim was to linguistically describe and analyse the test tasks and texts that had been used in the KPG exams of all levels and modules in the four languages (i.e. English, French, German, Italian) that were being tested up to that moment[4]. For that purpose, a group of experts specified the linguistic parameters on the basis of which researchers from various areas in the field of Linguistics (e.g. applied and theoretical linguistics, computational linguistics, psycholinguistics, etc.) were asked to linguistically describe the KPG test tasks. The present researcher participated in the group that described and analysed the expectations of the oral tasks designed for the KPG oral exams in English. The project coordinator was responsible for checking all the descriptions and providing feedback. However, the initial task descriptions differed a lot, because the researchers came from very different research and educational backgrounds, which influenced their view of the linguistic terms on which description was based. Consequently, there was a need for clearer description guidelines as well as for a clearly defined theoretical basis on which task description would be conducted.

The theoretical background of the Task Description Model (TDM) that was finally determined mainly adheres to the systemic functional approach to language use, proposed by Halliday (1985, 2002) as well as to a model of language proposed by Knapp and Watkins (1995), which describes grammar in relation to genre and text. Although Knapp and Watkin's model was developed to assist the instruction and assessment of writing skills within the Australian educational context, it was very helpful in the process of associating text purpose with specific lexicogrammatical features in all the tasks used in the KPG exam system. Besides, this model is also based on the principle that language is a socially constructed phenomenon, mainly expressed through the Hallidayan functional approach to language use, which generally maintains that genres relate to specific linguistic structures. However, Knapp and Watkins move one step forward. What makes their model special is the fact that genres are not seen as products or text types, but as sets of generic processes with a specific purpose, which is achieved by the application of "relevant structural and grammatical knowledge to produce appropriate texts" (Knapp and Watkins, 1995, p.26). In particular, in their model, there are five types of generic processes (*description*, *explanation*, *instruction* giving, *argumentation* and *narration*), each of which is realized through specific lexicogrammatical, structural and cohesive elements.

Based on the theoretical frame described above and after several discussions and revisions (see Κονδύλη & Λύκου, 2009), the task description model (TDM) finally included six categories of analysis: *topic*, *genre*, *generic process*, *speaker (or writer) - audience roles* and *lexicogrammar*. These categories were also used in the analysis of the KPG speaking tasks designed for levels B1 and B2, the results of which are presented in this study.

The TDM and the B1-B2 KPG speaking tasks: research aims

The six categories of the TDM were used to analyze the task rubrics designed for Activities 2 and 3, in the B1 and B2 KPG speaking modules and administered over nine consecutive exam administrations from April 2005 to May 2009. *Topic* and *speaker-audience roles* were determined by the information provided in the task rubrics. Regarding *genre*, two types

were analyzed: the 'monologic talk, based on visual prompts (Activity 2 tasks) and the 'monologic talk based on Greek text(s)' (tasks of Activity 3[5]). Concerning 'generic process', the features on which description was based came from Knapp and Watkins' categorization of five types of generic processes, on the basis of their lexicogrammatical, structural and cohesive characteristics:
• Description
• Narration
• Explanation
• Instructions
• Argumentation

Regarding the final category (Lexicogrammar), the description included the following characteristics, most of which were considered within the frame of the Hallidayan approach to functional grammar:
• *vocabulary* (every-day / subject specific)
• *tense* and *aspect*
• *modality* (deontic / epistemic)[6]
• *clause types* (verbal, mental, relational, material, existential)
• *cohesive devices* (additive, temporal, causal, consequential, pronouns etc)
• *impersonal syntax*
• *direct/reported speech*
• *grammatical metaphors* and *idioms*[7].

An example of B2 task analysis is provided in Table 1.

Activity 2: Having Fun (May 2006) – Task 2				
Look at photos 14 and 16 and tell us who you think the people in them are in each case, what their relationship is and what they enjoy doing in their free time.				
Topic:	Genre:	Generic process:	Speaker - Audience roles:	Lexicogrammar
Having fun	Monologic talk (based on visual prompts)	Description (of people, feelings, likes and dislikes)	Test taker - Examiners	Present tenses Epistemic modality (modal verbs and adverbial phrases expressing probability) Types of clauses: material, relational, mental (cognitive and emotive) Type of cohesion: through pronouns

Table 1: Using the TDM categories to analyze an example of B2 level speaking task.

The analysis initially focused on two of the five categories in the TDM: the generic process and the lexicogrammar. Given that Knapp and Watkins associate their five types of generic processes with written production, it would be interesting to find out what types of generic processes are more frequently generated by oral production tasks in the KPG exams. Moreover, the interaction between generic processes and lexicogrammatical choices in spoken performance has been relatively unexplored. Therefore, the aims of the present study were:

• to detect whether there were any types of generic processes or any lexicogrammatical characteristics that tended to appear more frequently than others in the oral task rubrics of Activities 2 and 3, at levels B1 and B2, and

• to examine these characteristics in combination, in order to find out whether there were any systematic patterns of item combinations at each level (e.g. whether a particular type of generic process was expected to systematically lead to specific lexicogrammatical characteristics).

For data analysis purposes, two separate Excel worksheets were created (one for each level of proficiency under examination) in order to count the frequency of the items in each category. The task rubrics of Activities 2 and 3 were also analyzed separately, in order to detect any differences in the frequency of the analytical categories across activities. In this study, the findings focus on the interaction between generic process and its lexicogrammatical realization in terms of modality and cohesion in the oral tasks designed for the B1 and B2 KPG oral exams.

Findings and Discussion

Based on the analysis of the task rubrics designed for Activity 2 at levels B1 and B2, Table 2 shows in how many tasks test takers are expected to get involved with each generic process. According to Table 2, in most tasks test takers are asked to deal with *Description*. *Explanation* and *Narration* appear to be the next most frequent types of generic processes at levels B2 and B1. However, Table 2 fails to show that in Activity 2, there are some tasks which involve test takers in a combination of generic processes. The most common of these is '*Description and Explanation*'. This combination of generic processes is very frequent, both at level B1 (60%) and at level B2 (94%). For example, the test takers are asked to describe something (e.g. a picture, a situation, their feelings, etc) and explain something, in relation to the situation portrayed in the visual prompts. Furthermore, '*Instructions and Explanation*' as well as '*Narration and Explanation*' are found to be two, less frequent, combinations. The task rubric analysis also revealed that the generic process of *Description* should be more subtly defined because it may include the description of people, events, thoughts, feelings or even the function of providing information about something.

Generic Process	B1 level tasks (total number of tasks: 84)	B2 level tasks (total number of tasks: 190)
Description	59	152
Explanation	18	86
Instructions	7	3
Narration	18	7
Argumentation	2	3

Table 2: Number of Activity 2 tasks with expected types of generic processes

The analysis of the task generic processes could lead to certain assumptions regarding their realization at the lexicogrammatical level. In particular, since test takers at levels B1 and B2 are very likely to be assigned tasks in which they have to describe and explain something, oral text cohesion will probably be achieved through *referential*, *causal* and *temporal* cohesive devices. For example, the coherent description of a situation involves the use of referential pronouns that link the participants together, whereas when an explanation is provided, the use of causal connectives (e.g. because, since, etc.) is necessary. Moreover,

the use of time expressions is necessary in the narration of events (usually at level B1) as well as in some cases when a situation is described or when a prediction is made. Based on these assumptions, Table 3 shows how often each of these types of cohesive devices is expected to appear in the candidates' linguistic performance in the tasks of Activity 2, at levels B1 and B2.

Types of cohesion	Frequency of types of cohesion	
	Level B1 (total number of tasks: 84)	Level B2 (total number of tasks: 190)
Referential (through pronouns)	52%	46%
Causal	29%	47%
Temporal	50%	22%
Additive	20%	15%
Consequential	5%	4%
Concessional	4%	10%

Table 3: Expected frequency of types of cohesion in Activity 2

Another finding relates to the test-takers' degree of familiarization with the situation presented in the task and how this influences their linguistic choices, especially in terms of 'modality', which constitutes another category in the TDM. In particular, at level B2 test takers are frequently presented with a picture, in which the situation or people depicted are unknown to them and the tasks require they say who the people are, what they are doing or how they feel. Consequently, since the candidates do not necessarily know the people in the visual stimuli and have to make inferences about what happens, they need to use specific modal verbs and adverbial phrases expressing probability (e.g. could, may, perhaps, etc.) or mental clauses (e.g. I think / believe that…), all of which constitute ways by which 'epistemic modality' is realized. On the contrary, most of the Activity 2 task rubrics designed for level B1 are phrased in such a way that presupposes that test takers participate in the situation and know the people presented on the prompts. Consequently, no inferences need to be made. In fact, test takers are mostly asked to make descriptions (of people, feelings, the situation), using present and past tenses, without any expressions of modality.

Table 4 shows how many Activity 2 tasks require test takers to express either epistemic or deontic modality or neither of them. After analysing task rubrics and their lexicogrammatical expectations, Table 4 shows that, in fact, more instances of epistemic modality are expected at level B2 than at level B1 (see Table 4). This finding also implies that mental clauses are expected to be more common at level B2 than at level B1, because they are one of the various ways in which epistemic modality can be realized. Regarding deontic modality, it seems less likely to be traced in the tasks of Activity 2, because it relates to the generic process of giving instructions or advice, which has only been encountered in a few B1 tasks. There are also many tasks which do not involve test takers into any kind of modality. These might be the tasks that presuppose that the test takers are familiar with the situation on the visual prompts and they just have to describe or explain it, by using tenses.

Concerning the tasks of Activity 3, the most prevalent type of generic process is *Instructions,* at both levels. However, similarly to Activity 2, oral mediation task rubrics seem to generate combinations of generic processes. In these cases, the most common combination is '*Instructions and Description*', both for level B1 (75%) and level B2 (59%). Moreover,

'*Description and Explanation*' is another, less frequent, combination, which mostly appears in a few B2 level oral mediation tasks.

Types of modality	B1 level tasks (total number of tasks: 84)	B2 level tasks (total number of tasks: 190)
Epistemic	25	93
Deontic	5	4
Nothing	52	92

Table 4: Number of Activity 2 tasks with expected cases of modality

Since the most common generic process in the tasks of Activity 3 is *Instructions* and test takers are mostly asked to provide instructions or advice, a frequent use of deontic modality expressions should be expected, mainly in the form of modal verbs expressing obligation (e.g. should, can, could, etc.). Interestingly, the analysis of the Greek texts that serve as input in the oral mediation tasks revealed that *Instructions* is also the most common type of generic process in the texts used at both levels. However, what differs in the Greek texts is the way advice and instructions are expressed. For example, information is mostly provided in present tenses and in bullet form. The test takers should use this information to address another interlocutor and advise him/her by using modal verbs (e.g. *should*, *could*), imperative or other expressions of deontic modality (e.g. *it is necessary that …*). Moreover, given that in most mediation tasks the test takers are asked to provide advice or instructions and give explanations, then speech cohesion is expected to be achieved by numbering (i.e. instruction steps/advice) and using causal connectives, whereby *additive* and *causal* cohesion are respectively realized.

In summary, these findings have revealed some basic differences between Activity 2 and 3, at both levels of proficiency. Regarding the type of generic process, in the tasks of Activity 2, test takers are usually involved in *Description* and *Explanation*, whereas in Activity 3 *Instructions* is the most frequent generic process. These differences in the generic processes lead to differences in the expected lexicogrammatical characteristics during oral performance in the tasks of Activities 2 and 3. More specifically, in *Description* and *Explanation* the speakers are more likely to use referential pronouns and causal links in order to perform the task. Moreover, if they are presented with an unfamiliar situation they need to resort to epistemic modality, in order to make inferences about the people and what exactly happens. Therefore, since *Description* and *Explanation* are the most frequent generic processes in the tasks of Activity 2, referential, causal cohesion and epistemic modality are expected to be frequently found in the test takers' language output. On the contrary, since *Instructions* is the most common generic process in the tasks of Activity 3, test takers are expected to use additive cohesive devices as well as expressions of deontic modality in order to give instructions or advice.

As far as the differences between levels B1 and B2 are concerned, the analysis of the task rubrics has shown that at level B2 test takers are more frequently asked to imagine that they are part of a fictitious situation. Therefore, more instances of epistemic modality are expected at level B2, whereas at level B1 the use of present tenses to provide simple descriptions is more frequent. Besides, the expression of probability or obligation (i.e. modality) is usually taught after the instruction of tenses, therefore, instances of modality are less expected at B1 level oral production.

Conclusion, implications and limitations

The description and analysis of the oral tasks used at B1 and B2 level KPG exams revealed systematic interactions between generic process and lexicogrammatical realizations in the oral tasks analyzed in terms of modality and cohesion. In particular, specific generic process combinations were repeatedly observed in the tasks of Activities 2 and 3, which are expected to lead to specific lexicogrammatical characteristics, without any particular deviating cases. This finding offers evidence of task consistency in the oral KPG tasks at levels B1 and B2, which positively adds to the validity of the task design process. Moreover, the tasks designed for Activities 2 and 3 were found to involve the test takers in different generic processes, thus leading the test-takers to different lexicogrammatical choices. Therefore, they are evaluated on different aspects of linguistic ability, something that positively adds to the fairness of their final score.

Task analysis by the means of the TDM resulted in an analytical description of some lexicogrammatical characteristics that should be found in the successful oral performance at levels B1 and B2, in the KPG exams. At the same time, the central role that generic process plays in making predictions about the characteristics of oral perfromance was highlighted. If the characteristics of expected language production are also observed in real time performance, they could be used to make the illustrative descriptors for oral production at levels B1 and B2 more accurate. Therefore, material writers would be able to base task construction on clearer guidelines. Moreover, the findings of task description could help in describing more accurately the oral assessment criteria that refer to lexicogrammatical choices at these two levels of proficiency. The improvement of the oral assessment scale combined with the oral examiners' training in relating the generic process implied by the task to specific lexicogrammatical characteristics, would hopefully ensure sounder criteria for marking oral performance in the KPG exams.

However, in order to obtain even more accurate descriptions regarding expected oral performance at levels B1 and B2, the TDM should be used to explore the interaction of even more linguistic characteristics. For example, it would be useful to explore how topic relates to lexical choices or to the use of specific tenses. This type of analysis could probably reveal more differences regarding expected oral production at levels B1 and B2. Additionally, the discourse analysis of real test performance could shed some light on how lexicogrammatical choices are actually realized at different levels of proficiency.

Future research should analyse oral task characteristics at levels A1-A2 and C1 to find out how tasks become more demanding from lower to higher levels of proficiency in the KPG exams. Ultimately, all the findings from the application of the TDM to tasks of oral production could be compared with those on written production, thus acquiring a complete picture of the linguistic characteristics of expected language production in the B level KPG exams.

Finally, this study presents some pedagogical implications for teachers, exam providers and material writers. In particular, since task design in the KPG exams is based on the assumption that context affects linguistic choices, test takers need to be taught how to analyze the information provided in the task rubrics. This does not mean that learners should necessarily be exposed to the SFL principles or get familiarized with such terms as *generic process*. Instead, they should be aware of the fact that linguistic choices are neither accidental nor can they be memorized as well as that language is a construct with social and

cultural dimensions. A final implication for language teaching and assessment, which expands the context of the KPG exams, is that language teachers should teach their students how to recognize task goal and text purpose because this ability is likely to help students use grammar and vocabulary appropriately, thus being effectively equipped for success not only in language exams, but also in real life interactions.

Author's email: elfie1903@yahoo.com

Notes

1. KPG stands for 'Kratiko Pistopiitiko Glossomathias'.
2. In May 2011 an integrated B level exam was introduced for the first time, which consists of three activities, with an equal number of B1 and B2 questions and tasks. A sample of the new integrated B level test can be found at: http://rcel.enl.uoa.gr/rcel/news.htm
3. The illustrative descriptors for language production specify what the candidates are expected to be able to do with language, depending on their level of proficiency. Information about the KPG illustrative descriptors for oral production at levels B1 and B2 can be found at: http://rcel.enl.uoa.gr/rcel2/texts/B1-Specs.pdf / http://rcel.enl.uoa.gr/rcel2/texts/B2-C1-Specs.pdf
4. The scientific head of this project is Prof. Dendrinos. The project was launched in 2007-2008 by the University of Athens within the frame of a larger research project, entitled "Exam system and language learning certification" (also known by the Greek acronym SAPiG), which was funded by EPEAK, EKT (European Community Fund) and national resources. Since January 2009, the project has evolved as autonomous scientific work, involving experts from various disciplines. More information about the project can be found at: http://rcel.enl.uoa.gr/sapig/gr_ereuna_01_glosso.htm.
5. In the one-sided talk (i.e. Activity 2), task response is based on a set of visual prompts that accompany the tasks. In Activity 3, which is also known as 'mediation', the test takers are asked to respond to the task, by relaying relevant information found in a Greek text to an interlocutor who does not understand / speak Greek.
6. According to Eggins (1994, p.179, 187), *modalization* (or else, 'epistemic modality') expresses how likely something is to happen or how frequently something happens. It can appear through modal operators (e.g. might) and mood adjuncts (e.g. possibly). Modulation (or else, 'deontic modality') expresses obligation / necessity (something should(n't) be done) or inclination (how willing I am to do something).
7. Κονδύλη & Λύκου (2009) provide a complete description of each one of all these lexicogrammatical features, by offering examples and commenting on the theoretical framework they fit into.

Acknowledgement

References

Bygate, M. (1999). 'Quality of language and purpose of task: patterns of learners' language on two oral communication tasks'. *Language Teaching Research* 3/3: 185-214.

Bygate, M. (2009). 'Teaching and testing speaking'. In M. H. Long & C. J. Doughty, (Eds.), *The handbook of language teaching*. West Sussex, Malden: Blackwell Publishing Ltd., 412-440.

Council of Europe, (2001). *Common European framework for reference for languages: learning teaching and assessment*, Cambridge: Cambridge University Press.

Dendrinos, B. (2007). 'Information booklet: the Greek state examination system to certify foreign language proficiency.' Athens: RCEL. Accessed at: http://rcel.enl.uoa.gr/rcel2/texts/B1-Specs.pdf and http://rcel.enl.uoa.gr/rcel2/texts/B2-C1-Specs.pdf on 20 Feb 2010.

Eggins, S. (1994). *An introduction to systemic functional linguistics*. London: Pinter.

Fulcher, G. & Davidson, F. (2007). *Language testing and assessment. an advanced resource book*. New York: Routledge.

Halliday, M.A.K. (2002). *On grammar*. London: Continuum.

Halliday, M.A.K. & Hassan, R. (1976). *Cohesion in English*. London: Longman.

Halliday, M.A.K. & Hassan, R. (1985). *Language, context and text: aspects of language in a social-semiotic perspective*. Oxford: Oxford University Press.

Halliday, M.A.K. & Mathiessen, C.M.I.M. (1999). *Construing experience through meaning*. London: Cassell.

Halliday, M.A.K. & Mathiessen, C.M.I.M. (2004). *An introduction to functional grammar (3rd ed.)*. London: Arnold.

Huang, J. & Mohan, B. (2009). 'A functional approach to integrated assessment of teacher support and student discourse development in an elementary Chinese programme'. *Linguistics and Education,* 20: 22-38.

Hughes, C. (2009). 'Assessment as text production: drawing on systemic functional linguistics to frame the design and analysis of assessment tasks'. *Assessment and Evaluation in Higher Education,* 34/5: 553-563.

Hunta, A. Kohonen, V. Kurki-Suonio, L. Luoma, S. (1997). *Current developments and alternatives in language assessment.* Jyvaskyla (Finland): University of Jyvaskyla.

Karavas, K. (2009). *The KPG speaking test in English: a handbook*. Athens: RCEL.

Knapp, P. & Watkins, M. (2005). *Genre, text, grammar. technologies for teaching and assessing writing*. Sydney: UNSW Press.

Κονδύλη, Μ. & Λύκου, Χ. (2009). «Γλωσσολογική περιγραφή των θεμάτων του ΚΠΓ: Η οπτική του κειμενικού είδους και της λεξικογραμματικής». *Research Periodical*, 1-17, accessed at: http://rcel.enl.uoa.gr/periodical/article_periodical_01.pdf on 12 May 2011.

Kormos, J. (1999). 'Simulating conversations in oral-proficiency assessment: a conversation analysis of role plays and non-scripted interviews in language exams'. *Language Testing,* 16/2: 163-188.

Long, M.H. and Doughty, C.J. (Eds.) (2009). *The Handbook of language teaching*. West Sussex, Malden: Blackwell Publishing Ltd.

Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.

Martin, J. (1998). 'Linguistics and the consumer: the practice of theory'. *Linguistics and Education,* 9/4: 411-448.

Milanovic, M. & Saville, N. (1996). 'Introduction'. In M. Milanovic & S. Saville (Eds.), *Performance testing, cognition and assessment: selected papers from the 15th language testing research colloquium.* Cambridge: Cambridge University Press, 1-17.

O'Sullivan, B. (2008). *Modelling performance in tests of spoken language*. Frankfurt: Peter Lang.

O'Sullivan, B., Weir, J.C. & Saville, N. (2002) 'Using observation checklists to validate speaking-test tasks'. *Language Testing,* 19/1: 33-56.

Pavlou, P. (1997). 'Do different speech interactions in an oral proficiency test yield different kinds of language?'. In A. Hunta, V. Kohonen, L. Kurki-Suonio & S. Luoma (Eds), *Current developments and alternatives in language assessment.* Jyvaskyla (Finland): University of Jyvaskyla, 185-201.

Shohamy, E. (1988). 'A proposed framework for testing the oral language of second/foreign language learners'. *Studies in Second Language Acquisition,* 10: 165-179.

Young, R. (1995). 'Conversational styles in language proficiency interviews'. *Language Learning,* 45/1: 3-42.

ΥΠΕΠΘ. (2003). *Κρατικό πιστοποιητικό γλωσσομάθειας: ενημερωτικό έντυπο*. Αθήνα: Υπουργείο Εθνικής Παιδείας και Θρησκευμάτων.

Re-examining text difficulty through automated textual analysis tools and readers' beliefs: the case of the Greek State Certificate of English Language Proficiency exam

[Επανεξετάζοντας τη δυσκολία αναγνωσιμότητας κειμένων με τη βοήθεια αυτοματοποιημένων συστημάτων ανάλυσης κειμένου και των απόψεων των αναγνωστών: η περίπτωση των εξετάσεων Αγγλικής γλώσσας του Κρατικού Πιστοποιητικού Γλωσσομάθειας]

Jenny Liontou

This article reports on an exploratory study that aimed at describing and comparing a range of linguistic features that characterize the reading texts used at the B2 and C1 level of the Greek State Certificate of English Language Proficiency exam (KPG[1]). Its ultimate purpose was to explore the contribution of such features to perceived text difficulty while at the same time examining the relationship between strategy use and test-takers' perceived level of reading comprehension difficulty reported in 7,250 questionnaires. Text analysis revealed significant differences between B2 and C1 reading texts for a specific number of text features such as word, paragraph and text length, readability indices, levels of word frequency and presence of words with rich conceptual content. A significant correlation was also found between B2 test-takers' perception of reading module difficulty and specific text features i.e. lexical diversity, abstract words, positive additive connectives and anaphoric references between adjacent sentences. With regard to C1 test-takers, data analysis showed that two specific text variables i.e. positive logical connectives and argument overlap, correlated significantly with readers' perception of reading module difficulty. Finally, problem-solving reading strategies such as rereading the text, guessing the meaning of unknown words and translating in mother tongue were found to correlate significantly with perceived text difficulty, whereas support-type reading strategies such as underlining or selectively reading parts of the text were less often employed regardless KPG test-takers' perception of text difficulty. The findings of this study could help both EFL teachers and test designers gain valuable knowledge regarding EFL learners' reading habits and also become more alert to the difficulty specific text features impose on the latter .

છ

Στο παρόν άρθρο παρουσιάζονται τα αποτελέσματα μιας μακροχρόνιας έρευνας που είχε ως στόχο τη λεπτομερή περιγραφή και σύγκριση υφομετρικών μεταβλητών των κειμένων κατανόησης γραπτού λόγου που έχουν χρησιμοποιηθεί στα επίπεδα Β2 και Γ1 των εξετάσεων Αγγλικής γλώσσας του Κρατικού Πιστοποιητικού Γλωσσομάθειας (ΚΠγ).

Απώτερος στόχος της παρούσας έρευνας υπήρξε η διερεύνηση της επίδρασης κειμενικών μεταβλητών στο βαθμό δυσκολίας αναγνωσιμότητας των συγκεκριμένων κειμένων. Παράλληλα εξετάστηκε η χρήση από τους εξεταζόμενους συγκεκριμένων στρατηγικών κατανόησης κειμένου και η σχέση των τελευταίων με το βαθμό δυσκολίας κατανόησης αγγλικών κειμένων σύμφωνα με τις απόψεις των συμμετεχόντων στις συγκεκριμένες εξετάσεις πιστοποίησης της γλωσσομάθειας όπως αυτές εκφράστηκαν σε 7.250 ερωτηματολόγια. Από την ανάλυση προέκυψαν στατιστικά σημαντικές διαφορές ανάμεσα στα κείμενα επιπέδου Β2 και Γ1 όσον αφορά σε συγκεκριμένες υφομετρικές μεταβλητές σε επίπεδο λέξης, πρότασης και παραγράφου καθώς και χαρακτηριστικών λεξιλογικής πλουσιότητας και πυκνότητας. Επιπλέον, παρατηρήθηκε η ύπαρξη στατιστικά σημαντικής συσχέτισης ανάμεσα στις απόψεις των εξεταζόμενων σχετικά με τη δυσκολία αναγνωσιμότητας κειμένων και τη χρήση συγκεκριμένων στρατηγικών ανάγνωσης όπως η ανάγνωση κειμένου περισσότερες από μια φορές, η προσπάθεια υποθετικής ερμηνείας των άγνωστων λέξεων αλλά και η μετάφραση στη μητρική γλώσσα για καλύτερη κατανόηση του κειμένου. Η χρήση «υποστηρικτικών» στρατηγικών ανάγνωσης όπως η υπογράμμιση ή η επιλεκτική ανάγνωση σημείων του κειμένου υπήρξε περιορισμένη και μη σχετιζόμενη με τις απόψεις των υποψηφίων ΚΠγ για το βαθμό δυσκολίας κατανόησης συγκεκριμένων αγγλικών κειμένων. Τα αποτελέσματα της παρούσας έρευνας θα μπορούσαν να προσφέρουν στους καθηγητές της Αγγλικής ως ξένη γλώσσα αλλά και τους συγγραφείς θεμάτων για εξετάσεις γλωσσομάθειας σημαντικές πληροφορίες σχετικά με τις στρατηγικές ανάγνωσης αγγλικών κειμένων και τα υφομετρικά χαρακτηριστικά αυτών που εντείνουν τη δυσκολία κατανόησης για συγκεκριμένους αναγνώστες.

Keywords: text difficulty, readability, lexical complexity, reading strategies, test-taking strategies

Introduction

Over the last sixty years readability or text difficulty has been an area of concern for all those who need to establish the appropriacy of a given text for any pedagogic purpose (Brabham and Villaume, 2002; Fry, 2002; Hatcher, 2000; Mesmer, 2005). It has also been recognized that providing test-takers with texts that are too difficult or too easy to process can affect their performance and contaminate test results (Bachman, 2002; Carr, 2006; Bailin and Grafstein, 2001; Kozminsky and Kozminsky, 2001). A range of reader factors that affect the reading process have also been recognized including motivation, background knowledge and previous reading experience (Brantmeier, 2005; Drucker, 2003; Keshavarz *et al*., 2007; Krekeler, 2006; Rupp *et al*., 2006; Salataci and Akyel, 2002). However, as these variables are essentially beyond the control of the researcher, it is facets of the text that have received the most attention (Leong *et al*., 2002; Parker *et al*., 2001). Particularly in relation to reading strategies[2] a number of studies have used introspective methods (e.g. Anderson *et al*., 1991; Cohen & Upton, 2007; Nikolov, 2006; Storey, 1997) or questionnaires (e.g. Brand-Gruwel *et al*.,1998; Nevo, 1989; Phakiti: 2003a, 2003b; Purpura, 1997; Zimmerman & Pons, 1986) to examine the processes employed by test-takers and provide evidence for the construct validity of an examination. However, given the lack of sufficient research evidence regarding the relationship between text difficulty and employed strategies, more validation studies in this area are needed for both test designers and test-takers to become aware of the nature of such processes and their contribution to exam performance.

Literature review

Although a lot of research has been conducted in the field of second language acquisition with specific reference to ways of reading and text processing strategies, Alderson (2000: 104) stresses language testers' lack of success "to clearly define what sort of text a learner of a given level of language ability might be expected to be able to read or define text difficulty in terms of what level of language ability a reader must have in order to understand a particular text". Fulcher (1997) also draws on the importance of text difficulty or text accessibility as a crucial but much neglected area in language testing. For him test developers need to be aware of the range of factors that make texts more or less accessible in order to be able to select reading texts at appropriate levels for inclusion into the reading sub-tests of their examinations. He further points out that research in this area is particularly pertinent because text difficulty is re-emerging as an area of great concern not only in language teaching and materials writing but also in the testing community.

Despite the considerable advances that have been made in exploring and understanding the various aspects of foreign language acquisition and reading performance, the available research has been rather unsuccessful in clearly defining those text features that have a direct impact on text complexity (Davies and Irvine, 1996; Dale and Chall, 1995). Consequently, for reasons of practicality, many researchers are still resorting to readability formulae or their own experience for assigning reading levels to texts (Juan, 2006; Ko, 2005; Kobayashi, 2002; Oakland and Lane, 2004; Trites and McGroarty, 2005). However, many researchers have pointed to the serious limitations of readability formulae and stressed the need for a more in-depth analysis of text features (Lehner, 1993; Shokrpour, 2004; Spadorcia, 2005). Weir (2005) further acknowledges that although the Common European Framework of Reference for Languages (CEFR) attempts to describe language proficiency through a group of scales composed of ascending level descriptors, it fails to provide specific guidance as to the topics that might be more or less suitable at any level of language ability or define text difficulty in terms of text length, content, lexical and syntactic complexity. Alderson *et al.* (2004: 13) also stress that many of the terms in the CEFR remain undefined. They further argue that "difficulties arise in interpreting it [i.e. the CEFR] because it does not contain any guidance, even at a general level, of what might be simple in terms of structures, lexis or any other linguistic level".

Considering all the above, in the present research it is assumed that a better measure of text complexity can be achieved by using systemic functional grammar as a basis for the text analysis, since such a model of language can help analyze prose in a manner that classifies ideas according to their role in conveying the total meaning of the passage and further show how relationships among ideas account for the overall coherence of a text (Freebody and Anderson, 1983). To this end, the framework for the analysis of text cohesion in this paper has been largely based on Halliday & Hasan's work (1976) and further supplemented with features proposed by Kintsch and van Dijk (1978). At this point, it is worth mentioning that Halliday & Hasan's model of text cohesion has been applied by a number of researchers in general discourse analysis, but its applicability and usability in examining foreign language reading comprehension text difficulty remains greatly unexplored.

Aim and methodology

The aim of the research presented in this paper has been twofold: a) to delineate and compare a range of linguistic features that characterize the reading texts used at the B2 (Independent User) and C1 (Proficient User) level of the Greek State Certificate of English Language Proficiency in order to explore their contribution to text difficulty and b) to examine whether specific reading and test-taking strategies are related to text complexity or

test-takers' perceptions of text difficulty. The stimulus for such an investigation has been the need for empirical evidence to supplement the to date mainly intuitive selection by item writers[3] of reading texts to be used at the B2 and C1 reading sub-tests of the specific exam. At this point it should be emphasized that the lack of research based evidence on the way levels of difficulty are assigned to test texts does not exclusively characterize KPG exams, but is a rather common feature among various well-established and long-administered exam systems that fail to provide sound evidence of their text selection processes (Bachman *et al.*, 1988; Chalhoub-Deville and Turner, 2000; Fulcher, 2000). These two levels were also chosen for reasons of practicality since, when the research began, they were the only ones available and had attracted a great number of test-takers.

Although it is beyond the scope of the present paper to provide a detailed description of the KPG English Language exam it is worth mentioning that each test consists of four sub-tests designed to assess the following competencies: a. reading comprehension and language awareness, b. writing and mediation, c. listening comprehension and d. speaking. According to the KPG specifications, at the B2 level of the reading comprehension and language awareness subtest candidates are required to skim through, scan or read closely longer or shorter texts of average difficulty and respond to a total of 75 items (reduced to 50 since November 2007) of various types designed to assess their overall reading skills, their knowledge of discourse and text grammar as well as their ability to make appropriate lexicogrammatical choices (KPG B2 specifications[4], 2003). At the C1 level candidates are asked to skim through or read carefully longer and more linguistically demanding texts of varied discourse, register and style, which they are likely to encounter in their social, professional or academic environment, and respond to a total of 75 items of various types (reduced to 60 since November 2007) designed to assess their ability to understand the overall meaning or partial meanings of these texts, to make reasoned inferences and draw conclusions as well as understand the relationships between different parts of a single text or among various texts (KPG C1 specifications[4], 2005). The level of the reading texts has been broadly defined in the Common Framework of the KPG examinations, applicable to all languages[5], according to which "the B2 Level exams are designed to test at an Independent User level the candidates' abilities to use English in order to understand the main ideas of texts of *average difficulty* on various topics, including abstract ideas or specialized information that requires some technical knowledge" whereas "the C1 Level exams are designed to test at a Proficient User level the candidates' abilities to understand texts relatively long and of a *high level of difficulty*" (Common Framework of the KPG examinations, 2003, p.6). However, it has not yet been possible to define, based on empirical evidence, the readability level of texts and the specific lexicogrammatical features that could be more appropriate to the intended audience i.e. prospective B2 or C1 test-takers. The current research has, thus, been designed to fill this void and further add to our present state of knowledge on EFL text difficulty in general. In order to explore these issues, the following research questions have been formed:

1. Are there any statistically significant differences between the B2 and C1 test texts with regard to specific lexicogrammatical features?
2. Is there a relationship between test-takers' impressions of text difficulty and the specific lexicogrammatical features estimated for each set (B2, C1) of KPG texts?
3. What reading and test-taking strategies do B2 and C1 KPG test-takers report using when responding to a multiple-choice reading comprehension test?
4. Is there a relationship between reported strategies and test-takers' perceived level of text difficulty?

In relation to text analysis, a range of lexicogrammatical features was automatically measured through computer programs such as Coh-Metrix 2.0, Web VocabProfile 3.0, AceReader Pro Deluxe, TextAnalyzer 2.0 and CLAN. To be more specific, Coh-Metrix 2.0 was

used to measure syntactic complexity and the frequency of particular syntactic classes along with text abstractness and conceptual similarity across sentences and paragraphs of the same text (see variables V1-V10, V24-V28, V31-V69 in Table 1). In addition, more surface text features that have been reported to contribute to text difficulty at a word and sentence level were estimated using Web VocabProfile 3.0 (see variables V17-V22, V29-V30 in Table 1), AceReader Pro Deluxe (see variables V11-V13, V16 in Table 1), TextAnalyzer 2.0 (see variables V14-V15 in Table 1) and CLAN (see variable V23 in Table 1). All in all, 24 B2 reading texts used between November 2003 to November 2010 examination periods and 24 C1 texts used between April 2005 to November 2010 examination periods were analyzed with regard to sixty-nine text variables. For the texts to be appropriate for comparisons a specific set of criteria was followed during the selection process i.e. only those reading comprehension texts that contained ten multiple choice questions with three options (A, B or C) per item were considered appropriate for further analysis.

In order to collect information regarding KPG test-takers' profile and reading strategies as well as their perceptions of text and task difficulty, a paper-and-pencil survey was conducted by the Research Centre for English Language Teaching Learning and Assessment (RCEL) at the University of Athens in cooperation with the Greek Ministry of Education. Taking into consideration test-takers' feedback -and in a way treating them as "judges" of reading texts- is a crucial part of the research, since they were the actual readers who had to interact with the texts in order to perform a set of tasks and demonstrate a successful performance in the specific exams. So, their feelings and opinions about the texts are believed to be of great importance since they can give us more in-depth information about text difficulty from the reader's point of view. To this end a questionnaire was administered on the day of the exam to all KPG test-takers sitting for the B2 and C1 English language exams in the May and November 2006 and 2007 and May 2008 examination periods[6]. To date, 7,250 questionnaires from five examination periods have been analyzed of which 4,750 referred to the B2 level and 2,500 to the C1 level. For the survey sample to be appropriate for statistical analysis and as representative as possible of the target population, a decision was made for its size to be at least 10% of the total number of test-takers taking each exam, with a minimum of 500 participants per examination period. In order to ensure a balanced geographical distribution and avoid any variation in response rates due to urbanization, 50% of the participants was randomly drawn from the five most densely populated[7] cities in Greece i.e. Athens, Thessaloniki, Patras, Larisa and Herakleion, Crete, 25% from rural areas and the remaining 25% from the Greek islands. Once test-takers finished their exam and before leaving the examination room, they were kindly requested to fill in a questionnaire and anonymously express on a five-point Likert scale their agreement or disagreement with a variety of statements regarding text and task difficulty, topic and genre familiarity, topic preference and text lexical complexity. Apart from the Likert scaling system, respondents also answered dichotomous questions in order to provide more general information regarding their reading and test-taking strategies.

Results and discussion

Once the analysis of text characteristics per level was completed, independent sample t-tests[8] were carried out in order to explore and further determine the significance of existing differences between the B2 and C1 level texts and thus answer the first research question (*Are there any statistically significant differences between the B2 and C1 test texts with regard to specific lexicogrammatical features?).* As can be seen in Table 1, among the range of analyzed text features, significant differences were found for a specific number of text variables including word (V2-V5), paragraph (V8) and text length (V1), readability indices (V11-V16), levels of word frequency (V17, V25) and Latent Semantic Analysis[9] cosines for adjacent sentences (V67) and paragraph-to-paragraph units (V69). In other words, texts used

at the C1 level included a significantly (p<0.05) higher number of words in text (tokens) than their B2 counterparts (V1).They also contained a significantly higher number of unique words (V2), which could have made comprehension comparatively more difficult since more content words would need to be decoded and integrated within the same discourse context. Moreover, C1 texts were characterized by significantly longer words in terms of syllables per word (V3), average number of characters per word (V4) and average number of syllables per 100 words (V5) as well as a significantly higher number of words per sentence (V6) and sentences per paragraph (V8), all of which could have contributed to overall text difficulty. It is notable that despite the serious limitations of readability formulae, in the present research a significant difference was revealed between B2 and C1 texts in relation to the six employed readability indices i.e. the Flesch Reading Ease Index (V11) and the Flesch-Kincaid Grade Level Index (V12), the Dale-Chall Grade Level Index (V13), the Spache Grade Level (V14), the Gunning's Fog Index (V15) and the Fry Readability Graph (V16), all of which rated B2 texts as less difficult than those used at the C1 level. This finding supports the view that despite their apparent simplicity, readability formulae do seem to come in some agreement with KPG test-designers' perception of text difficulty and could be of practical usefulness to them during the text selection and validation process. Finally, C1 texts were found to consist of a significantly lower percentage of the one thousand most frequent English words of the BNC frequency list (V17), a feature that could lead to increased text difficulty. Most importantly, the fact that C1 texts were characterized by significantly less conceptually similar sentences (V67) and paragraphs (V69) could have made the comprehension process more demanding since readers had to process and decode a higher number of ideas. Against our expectations no significant differences were found regarding additional text features especially in relation to syntactic and semantic complexity and text abstractness (V23-V66). This may be taken to suggest that more explicit text differences across levels could be drawn should test designers become more alert to such features as verb and noun hypernym levels (V33-V34), word concreteness(V26-V27), anaphoric reference (V54-V55) and content word overlap (V66) and take them into account during the text selection process.

	B2 N=24		C1 N=24			
	Mean	SD	Mean	SD	t (48)	Adj. sig. (2-tailed)
V1. No. of words in text	417.75	111.273	590.42	148.949	-4.550	**.000**
V 2. No. of different words in text (types)	223.96	49.074	298.58	55.013	-4.959	**.000**
V3. Syllables per word	1.539	.093	1.611	.084	-2.804	**.007**
V 4. Average number of characters per word	4.700	.2309	4.913	.2346	-3.091	**.003**
V 5 Average number of syllables per 100 words	150.986	8.106	158.917	9.106	-3.187	**.003**
V 6. Words per sentence	18.102	3.652	20.607	6.300	-1.685	.099
V7. No. of sentences	23.92	9.325	29.08	11.225	-1.735	.090
V8. Sentences per paragraph	3.346	1.530	5.704	3.404	-3.095	**.004**
V9. Average number of sentences per 100 words	5.988	1.666	5.071	1.105	2.245	.030
V10. No. of paragraphs	8.17	3.559	7.83	3.130	.345	.732
V 11. Flesch Reading Ease	58.205	8.969	48.530	9.133	3.703	**.001**
V 12. Flesch-Kincaid Grade Level	9.574	1.750	11.143	1.458	-3.374	**.002**
V 13. Dale-Chall Grade Level	9.521	1.913	11.750	2.658	-3.334	**.002**
V 14. Spache Grade Level	4.829	.700	5.488	.896	-2.835	**.007**
V 15. Gunning's Fog Index	10.241	2.314	13.390	2.743	-4.297	**.000**
V 16. Fry Readability Graph	9.17	1.949	11.38	2.163	-3.716	**.001**
V 17. K1 Words (1-1000)	79.781	5.217	74.573	4.857	3.579	**.001**
V 18. K2 Words (1001-2000)	8.024	2.257	8.817	1.880	-1.323	.192
V 19. K3 Words (2001-3000)	2.902	1.576	3.930	1.764	-2.127	.039
V 20. K4 Words (3001-4000)	1.986	1.423	2.437	.907	-1.307	.198

V 21. K5 Words (4001-5000)	.960	.701	1.137	.675	-.891	.378
V 22. Academic Words	4.228	2.488	4.894	1.688	-1.084	.284
V 23. Lexical Diversity	107.401	22.468	103.511	20.243	.630	.532
V 24. Log freq. content words	2.253	.154	2.143	.145	2.537	**.015**
V 25. Log min. freq. content words	1.162	.349	1.109	.298	.571	.571
V 26. Concreteness content words	376.891	23.404	386.434	16.312	-1.639	.108
V 27. Min. concreteness content words	177.63	19.763	174.71	15.909	.563	.576
V 28. Lexical Density	.552	.042	.554	.030	-.233	.817
V 29. Anglo-Sax Index:	73.410	4.948	72.812	5.389	.401	.691
V 30. Greco-Lat/ Fr-Cognates	26.589	4.948	27.187	5.389	-.401	.691
V 31. Negations	6.081	4.534	6.479	4.963	-.290	.773
V 32. Passive sentences	.118	.085	.173	.080	-2.315	.025
V 33. Noun hypernym	4.905	.455	5.284	1.979	-.914	.365
V 34. Verb hypernym	1.559	.148	1.741	.650	-1.338	.188
V 35. Causal cohesion	.834	.467	.881	.385	-.375	.709
V 36. Temporal cohesion	.827	.102	.816	.082	.389	.699
V 37. Spatial cohesion	.488	.062	.493	.085	-.241	.811
V 38. Causal content	50.594	13.273	47.180	9.732	1.016	.315
V 39. Conditional operators	1.617	2.084	1.080	1.362	1.055	.297
V 40. Intentional content	13.371	4.709	13.754	7.035	-.222	.825
V 41. All connectives	73.432	16.441	71.115	11.411	.567	.573
V 42. Pos. additive connectives	37.086	14.287	31.761	11.496	1.423	.162
V 43. Pos. temporal connectives	6.821	5.055	9.076	3.951	-1.721	.092
V 44. Pos. causal connectives	21.521	5.954	20.610	6.369	.512	.611
V 45. Pos. logical connectives	19.278	7.065	19.283	5.351	-.003	.998
V 46. Neg. additive connectives	12.132	7.281	9.501	4.826	1.475	.147
V 47. Neg. temporal connectives	.390	1.098	.424	.922	-.119	.906
V 48. Neg. causal connectives	1.111	1.906	.412	.762	1.667	.106
V 49. Neg. logical connectives	12.390	6.738	10.117	4.679	1.357	.181
V 50. Logic operators	43.114	13.120	37.504	10.736	1.621	.112
V 51. Syntactic structure similarity adjacent	.080	.020	.084	.019	-.682	.499
V 52. Syntactic structure similarity all_01	.078	.014	.078	.018	-.124	.902
V 53. Syntactic structure similarity all_02	.088	.015	.085	.024	.452	.654
V 54. Anaphor reference	.145	.097	.117	.076	1.125	.266
V 55. Adjacent anaphor reference	.285	.170	.288	.140	-.067	.947
V 56. Adjacent argument overlap	.445	.118	.487	.136	-1.139	.261
V 57. Adjacent stem overlap	.408	.146	.467	.173	-1.272	.210
V 58. Argument overlap	.368	.153	.407	.118	-.995	.325
V 59. Stem overlap	.324	.144	.379	.140	-1.339	.187
V 60. Noun Phrase incidence	273.873	19.515	281.133	23.049	-1.178	.245
V 61. Pronoun ratio	.210	.109	.170	.071	1.476	.147
V 62. Personal pronouns	57.375	30.424	48.342	21.944	1.180	.244
V 63. Modifiers per Noun Phrase	.879	.176	.959	.152	-1.679	.100
V 64. Higher level constituents	.726	.042	.708	.032	1.683	.099
V 65. Words before main verb	4.239	1.407	5.233	2.358	-1.773	.083
V 66. Content word overlap	.071	.022	.078	.018	-1.172	.247
V 67. LSA sentence adjacent	.149	057	.208	.078	-2.946	**.005**
V 68. LSA sentence all	.155	.059	.182	.083	-1.322	.193
V 69. LSA paragraph	.235	.141	.321	.149	-2.034	**.048**

Table 1. Results of Independent samples t-tests between B2 and C1 KPG reading texts

In order to answer the second research question (*Is there a relationship between test-takers' impressions of text difficulty and the specific lexicogrammatical features estimated for each set (B2, C1) of KPG texts?)* Pearson correlation coefficients were estimated for all B2 and C1 KPG texts included in the questionnaires and data analysis showed a significant correlation between B2 test-takers' perception of reading module difficulty and lexical diversity (r .862,

p<0.05). This implies that the wider the range of vocabulary displayed in a text the more difficult its processing became for the B2 candidature. Apart from lexical diversity, it was found that positive additive connectives (r.-646, p<0.05) and anaphoric references between adjacent sentences (r. -530, p<0.05) negatively correlated with perceived module difficulty. This could be due to the fact that both of these variables relate to the semantic and syntactic complexity of a text and their absence could lead to information presented in a less straightforward and thus more difficult way to decode. Moreover, perceived text vocabulary difficulty was found to correlate significantly with the mean hypernym value of nouns (r. -835, p<0.05). This means that B2 test-takers found it difficult to decode abstract words with fewer hypernym levels possibly due to their lack of language competence or limited exposure to abstract words. Finally, the analysis showed that the higher the proportion of logical operators (r .709, p<0.05) the more lexically demanding the text became for B2 test-takers. This finding is in agreement with the fact that texts with a high density of logical operators[10] are considered difficult for most readers to process since such operators are often used to link rather long sequences of events. With regard to C1 test-takers, data analysis showed that two specific text variables i.e. positive logical connectives (r. -789, p<0.05) and argument overlap (r. -794, p<0.05) correlated significantly with their perception of reading module difficulty. In other words, the lower the proportion of positive logical connectives or the proportion of sentences in a paragraph that shared one or more arguments, the more syntactically and semantically complex the text became for the majority of C1 test-takers. Finally, at the C1 level, reported text lexical complexity was found to correlate significantly with word frequency i.e. K5 words (r .666, p<0.05), word abstractness[11] (r. 706, p<0.05) and concreteness content words[12] (r .752, p<0.05), all of which relate to lexicogrammatical complexity and may be more difficult for EFL learners to master.

Regarding the third research question i.e. *what reading and test-taking strategies do B2 and C1 KPG test-takers report using when responding to a multiple-choice reading comprehension test*, statistical analysis showed that a limited number of reading and test-taking strategies were more frequently employed by KPG test-takers when sitting for the B2 exam (Table 2). To be more specific, the most frequent strategies employed by B2 test-takers were "*trying to guess the meaning of unknown words*" which ranked first in the frequency list with a mean of almost 71% followed by "*reading the text more than once*" (68%) and "*translating words in L1 to better understand their meaning*" (65%). The next strategy that emerged was a test-taking one i.e. "*reading first the questions and then the text*" with almost 53%, followed by "*combining information from different parts of a text*" with a mean of 48%. "*Underlining parts of a text*" was chosen by less test-takers (almost 32%) whereas "*selectively reading parts of the text*" and "*keeping notes while reading*" ranked rather low in the list with mean values of 17% and 16% respectively. It is interesting that the strategy running first in the frequency list for the C1 level is similar to the one reported by most B2 test-takers i.e. "*trying to guess the meaning of unknown words*". However, as we move down the list, a noticeable difference can be traced in comparison with the B2 level since "*combining information from different parts of the text*" ranked third with a mean value of 58% followed by "*translating words in L1 to better understand their meaning*" with 53%. Moreover, "*underlining parts of the text*" appeared to be more often employed by C1 test-takers with a mean of 44% compared with the 31% of the B2 level. Interestingly, a test-taking strategy such as "*reading first the questions and then the text*" was less often employed by C1 test-takers with a mean of almost 40% in comparison with their B2 counterparts who reported more frequent use of it with a mean of 53%. Similarly to the B2 level, "*selectively reading parts of the text*" and "*keeping notes while reading*" ranked low in the list with a mean value of 11% and 10% respectively.

When checking for significant differences across the employed strategies between the two levels it became apparent that although the same nine strategies were frequently employed by both B2 and C1 test-takers, their levels of frequency in certain cases varied significantly

(Table 2). The independent samples t-test showed that three specific strategies i.e. "*reading items first*", "*translating words in L1 to better understand their meaning*" and "*selectively reading parts of the text*" were significantly more often employed by B2 level test-takers who seemed more prone to problem-solving strategies. On the other hand, C1 test-takers appeared to employ a mixture of problem-solving and support reading strategies with "*reading the text more than once*" being the strategy used at a significantly higher proportion. This was possibly due to the fact that C1 texts were significantly longer than their B2 counterparts and test takers would need to read them more than once to understand the presented information. It is worth mentioning at this point that the frequent use of similar strategies at both levels maybe due to the fact that, since C1 test-takers have progressed from the B2 level, they make frequent use of strategies they have already mastered. Moreover, the use of more reading (S2-S7, S9) than test-taking strategies (S1, S8) seems to support the view that KPG test-takers were to some extent actively involved in the reading process and tried to understand the text rather than simply answer items by taking shortcuts and resorting to test-taking strategies. However, more cognitively demanding strategies such as "*combining information from different parts of the text*" or "*selectively reading parts of the text*" ranked rather low in the list, which raises the question of whether and to what extent test-takers have been exposed to such strategies and would be able to make use of them, if needed.

	B2 Level		C1 Level			
	Mean	SD	Mean	SD	t (48)	Adj. sig. (2-tailed)
S1. Reading first the questions and then the text.	.527	.010	.395	.005	22.051	**.000**
S2. Underlining parts of the text	.316	.044	.445	.034	-6.525	**.000**
S3. Trying to guess the meaning of unknown words.	.709	.074	.776	.050	-1.488	.187
S4. Translating words in L1 to better understand their meaning	.651	.050	.534	.026	4.119	**.006**
S5. Selectively reading parts of the text	.177	.015	.112	.030	3.743	**.016**
S6. Reading the text more than once	.683	.107	.772	.039	-2.419	**.032**
S7. Combining information from different parts of the text.	.488	.082	.584	.039	-2.113	.079
S8. Answering items without reading the text	.109	.010	.120	.033	-.631	.552
S9. Keeping notes while reading	.163	.021	.107	.009	3.322	.080

Table 2. Results of Frequencies and Independent samples t-tests comparing B2 and C1 KPG test-takers' reading strategies

Finally, in order to answer the fourth research question i.e. *if there is a relationship between reported strategies and test-takers' perceived level of text difficulty*, Pearson correlations were calculated between the valid percent of test-takers' responses to questions regarding text and item difficulty and the mean values of various reading strategies (Table 3)[13]. The data analysis revealed that test-takers were using different strategies given their perceptions of text and item difficulty. Regardless of their level of language competence they seem to have processed text in the following way: the closer to their expectations the text was the less they resorted to guessing and combining information from different parts of the text. If the text seemed easier than expected they refrained from translating, rereading, guessing the meaning of unknown words or combining information and opted for more test-taking strategies such as reading the items first or even selectively reading parts of the text to answer the questions. However, when the text seemed more difficult than expected, test-takers focused more on the text instead of reading the items first and started making use of specific strategies such as guessing, combining information, underling parts of the text or reading the text more than once. A similar approach was observed in relation to perceived item difficulty since the more demanding a question became the more times test-takers read

the text, tried to guess the meaning of unknown words and combined information from different parts of the text. These findings could help both EFL teachers and test designers gain valuable knowledge regarding learners' ways of processing a reading text. They could also make them alert to the fact that test-takers seem to have been more familiar with and have at hand a limited set of strategies which they repeatedly employ when dealing with a more linguistically complex text, possibly because they had been more exposed to them through their coursebooks.

	Text as difficult as expected by test-takers	Text less difficult than expected by test-takers	Text more difficult than expected by test-takers	Items as difficult as expected by test-takers	Items less difficult than expected by test-takers	Items more difficult than expected by test-takers
Translating words in L1		-.845				
Underlining parts of the text			.641		-.657	
Rereading the text		-.738	.545	-.515	-.677	.636
Guessing unknown words	-.819	-.909	.956	-.899		.950
Combining information	-.872	-.903	.881	-.875	-.935	.870

Table 3. Pearson correlations between text/item difficulty & employed strategies

Conclusion

The present study aimed at providing a detailed description of text features characterizing the reading texts used at the B2 and C1 level of the KPG English language exams while at the same time exploring the textual differences between them. The comparison identified significant differences between the two exams for a specific number of text variables including paragraph and text length, readability indices, levels of word frequency and presence of words with rich conceptual content as well as estimates of LSA cosines for paragraph-to-paragraph units. The absence of differences on many other of the employed syntactic, referential and semantic measures could be of practical usefulness to KPG test providers who may wish to add such a list of text features to their text selection guidelines for texts used at different levels of the KPG exams to be more clearly distinguishable.

This study also provided useful insights into the reading comprehension process of Greek learners of English while shedding light on the effect specific text and reader variables have on the reading outcome and the extent to which they interfere with text difficulty. Until now, these aspects remained unexplored in the context of the KPG English language exams. However, since KPG test-takers made their own interpretations to questionnaire items and reported strategy use to the best of their conscious knowledge, we must be wary of the limitations of the obtained data. As in other studies, at best, these data indicated trends in strategy use. It should also be noted that other strategies that were not included in the questionnaire might have been employed or even that the reported ones might have been used more or less often than test-takers indicated.

On the other hand, the fact that a large number of responses was repeatedly and consistently collected over a long period of time could add to the reliability of the findings. In other words, despite the inherent limitations of the research instrument, the collected data were to a great extent reflective of the type of strategies KPG test-takers employed when answering the multiple-choice questions of the B2 or C1 KPG English language exams. Based on these findings, it seems useful to draw EFL teachers' attention to the fact that more effort should be made to guide and train students into using a wider range of strategies, especially when processing texts that appear more difficult than expected. In addition, it might be

useful to make EFL learners more aware of some test-taking strategies such as reading all instructions attentively, answering questions based on their reading of the text and rechecking questions before submitting their test papers. This way they may achieve the best of their performance and even feel less anxious when taking an exam.

Finally, the present study attempted to make a methodological contribution in that instead of resorting to readability formulae, it made use of a range of computational tools and proposed a mixed model of estimating text difficulty. At the same time it investigated test-takers' perspectives on various aspects of text comprehensibility but, instead of studying specific variables separately, it explored and cross-related the effect of various variables for their interaction to be better defined and predicted within the context of the KPG exams. Nevertheless, given the complexity of the reading comprehension process and the limitations of the present study, more research is needed to better define text difficulty in terms of actual reader performance and further explore the effect of complex linguistic features on task-based performance across a variety of text types.

Author's email: tliontou@enl.uoa.gr

Notes
1. The initials KPG correspond to the acronym ΚΠΓ which in Greek stands for Kratiko Pistopiitiko Glossomathias, translated in English as State Certificate of Language Proficiency. Exams in six languages i.e. English, French, German, Italian, Spanish and Turkish, are administered by the Greek Ministry of Education and Religious Affairs in co-ordination with the National and Kapodistrian University of Athens and the Aristotle University of Thessaloniki. The KPG English language exams are developed by a team of test designers under the guidance of Professor B. Dendrinos at the Faculty of English Studies at the National and Kapodistrian University of Athens. Despite being in its infancy, KPG is rapidly gaining acceptance as a high-stakes exam in Greece and as such it can influence one's future prospects for employment and education. Exams are administered twice a year and since their introduction in November 2003 more than 500.000 test-takers have taken part in the English language exams.
2. Reading strategies are defined as those "conscious procedures" that readers deliberately employ to increase or enhance their comprehension of a text. As opposed to processes which are unconscious and more automatic, strategies are thus believed to be controlled by the readers. Test-taking strategies, on the other hand, are defined as those procedures which respondents consciously decide to use when taking an exam in order to perform a specific language task. Thus, under specific testing conditions, respondents may show test-wiseness and employ certain strategies to arrive at answers when their language knowledge is of limited help to them (Nevo, 1989).
3. Test-designers seemed to assign levels to texts through a holistic interpretation of exam specifications and their subjective judgment.
4. The full text in Greek is available online at: www.ypepth.gr/docs/kpg_english_examples.doc (Last access: 13/12/11).
5. The full text in Greek is available online at: www.ypepth.gr/docs/kpg_plaisio_eniaio.doc (Last access: 13/12/11). The KPG exams are currently offered in six languages i.e. English, French, German, Italian, Spanish and Turkish. The first four exams certify relevant language proficiency at

levels ranging from A1 to C1, whereas the Spanish exams are currently offered at B1-C1 levels and the Turkish ones at the B1-B2 levels.

6. The survey was commissioned and funded by the Research Centre for English Language Teaching Learning and Assessment at the University of Athens. The author was responsible for organizing and coordinating the surveys conducted between 2006, 2007 and 2008 examination periods.

7. According to the most recent official records of the Hellenic Statistical Authority (available online at: http://www.statistics.gr/portal/page/portal/ESYE).

8. The alpha level of 0.05 was corrected for multiple tests using the Holm-Bonferroni adjustment.

9. Latent Semantic Analysis (LSA) is a statistical technique for representing world knowledge, based on a large corpus of texts. LSA uses singular value decomposition, a general form of principle component analysis, to condense a very large corpus of texts to 100-500 dimensions (Graesser *et al.*, 2004).

10. Logical operator incidence score (and + if + or + cond + neg). Logical operators are prevalent in syllogisms and texts that express logical reasoning. They include the Boolean operators (and, or, not, if, then) and a small number of other similar cognate terms (Graesser *et al.*, 2004).

11. Adjusted mean for content words (0-6) (Graesser *et al.*, 2004).

12. Concreteness measures how concrete a word is, based on human ratings of the of the MRC Psycholinguistics Database. High numbers lean toward concrete and low numbers to abstract. The more abstract the words in a text the more difficult the comprehension process (Graesser *et al.*, 2004).

13. Only statistically significant Pearson correlations are presented in Table 3.

References

Alderson, C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

Alderson, J. C., Figueras, N., Kuijper, H., Nold, G., Takala, S., & Tardieu, C. (2004). 'The development of specifications for item development and classification within the Common European Framework of Reference for Languages: Learning, teaching, assessment. Reading and Listening. (Final Report of the Dutch CEFR Construct Project)'. Lancaster: Lancaster University. Retrieved 04/09/2011 from http://eprints.lancs.ac.uk/44/01/final_report.pdf.

Bachman, L. (2002). 'Some reflections on task-based language performance assessment.' *Language Testing*, 19/4: 453-476.

Bachman, L., Kunnan, A., Vanniaraian, S. & Lynch, B. (1988). 'Task and ability analysis as a basis for examining content and construct comparability in two EFL proficiency test batteries.' *Language Testing*, 5/2: 128-159.

Bailin, A. & Grafstein, A. (2001). 'The linguistic assumptions underlying readability formulae: a critique.' *Language & Communication*, 21: 285-301.

Brabham, E. & Villaume, S. (2002). 'Levelled text: The good news and the bad news.' *The Reading Teacher*, 55/5: 438-441.

Brantmeier, C. (2005). 'Effects of Reader's Knowledge, Text Type and Test Type on L1 and L2 Reading Comprehension in Spanish.' *Modern Language Journal*, 89/1: 37-53.

Carr, N. (2006). 'The factor structure of test task characteristics and examinee performance.' *Language Testing*, 23/3: 269-289.

Chalhoub-Deville, M. & Turner, C. (2000). 'What to look for in ESL admission tests: Cambridge certificate exams, IELTS and TOEFL.' *System*, 28/4: 523-539.

Cohen, A. & Upton, T. (2007). 'I want to go back to the text: Response strategies on the reading subtest of the new TOEFL*.' Language Testing*, 24/2: 209-250.

Dale, E. & Chall, J. (1995). '*Readability revisited: the New Dale-Chall readability formula*.' Cambridge, Massachusetts: Brookline Books Inc.

Davies, A. & Irvine, A. (1996). 'Comparing test difficulty and text readability in the evaluation of an extensive reading programme.' In M. Milanovic & N. Saville (Eds.), *Performance Testing, Cognition and Assessment: Selected Papers from the 15th Language Testing Research Colloquium, Cambridge and Arnhem*. Cambridge: Cambridge University Press, 165-183.

Drucker, M. (2003). 'What reading teachers should know about ESL learners.' *International Reading Association,* 22-29.

Freebody, P. & Anderson, R. (1983). 'Effects of vocabulary difficulty, text cohesion and schema availability on reading comprehension.' *Reading Research Quarterly*, 18/3: 277-294.

Fry, E. (2002). 'Readability versus levelling.' *The Reading Teacher*, 56/3: 286-291.

Fulcher, G. (1997). 'Text difficulty and accessibility: reading formulae and expert judgment.' *System*, 25/4: 497-513.

Fulcher, G. (2000). 'The `cummunicative' legacy in language testing.' *System*, 28, 4, 483-497.

Graesser, A., McNamara, D., Louwerse, M. & Cai, Z. (2004). 'Coh-Metrix: analysis of text on cohesion and language.' *Behavioral Research Methods, Instruments, and Computers*, 36: 193-202.

Halliday, M. & Hasan, R. (1976). *Cohesion in English*. New York: Longman.

Hatcher, P. (2000). 'Predictors of Reading Recovery book levels.' *Journal of Research in Reading*, 23/1: 67-77.

Juan, E. (2006). 'The Compensatory Nature of Discipline-Related Knowledge and English-Language Proficiency in Reading English for Academic Purposes.' *The Modern Language Journal*, 90: 210-227.

Keshavarz, M., Atai, M. & Ahmadi, H. (2007). 'Content schemata, linguistic simplification and EFL readers' comprehension and recall.' *Reading in a Foreign Language*, 19/1: 19-33.

Kintsch, W. & van Dijk, T. (1978). 'Toward a model of text comprehension and production.' *Psychological Review*, 85/5: 363-394.

Ko, M. (2005). 'Glosses, comprehension and strategy use.' *Reading in a Foreign Language*, 17/2: 125-143.

Kobayashi, M. (2002). 'Method effects on reading comprehension test performance: text organization and response format.' *Language Testing*, 19/2: 193-220.

Kozminsky, E. & Kozminsky, L. (2001). 'How do general knowledge and reading strategies ability relate to reading comprehension of high school students at different educational levels?' *Journal of Research in Reading*, 24/2: 187-204.

Krekeler, C. (2006). 'Language for special academic purposes (LSAP) testing: the effect of background knowledge revisited.' *Language Testing*, 23/1: 99-130.

Lehner, F. (1993). 'Quality control in software documentation based on measurement of text comprehension and text comprehensibility.' *Information Processing & Management*, 29/5: 551-568.

Leong, E., Ewing, M. & Pitt, L. (2002). 'E-comprehension: Evaluating B2B websites using readability formulae.' *Industrial Marketing Management*, 31: 125-131.

Mesmer, H. (2005). 'Introduction: Text Accessibility and the Struggling Reader.' *Reading & Writing Quarterly*, 21/1: 1-5.

Nevo, N. (1989). 'Test-taking strategies on a multiple-choice test of reading comprehension.' *Language Testing*, 6/2: 199-215.

Nikolov, M. (2006): 'Test-taking strategies of 12- and 13-year-old Hungarian Learners of EFL: Why Whales have Migraines.' *Language Learning*, 56/1: 1-51.

Oakland, T. & Lane, H. (2004). 'Language, Reading, and Readability Formulas: Implications for Developing and Adapting Tests.' *International Journal of Testing*, 4/3: 239-252.

Parker, R., Hasbrouck, J. & Weaver, L. (2001). 'Spanish Readability Formulas for Elementary-Level Texts: A Validation Study.' *Reading & Writing Quarterly*, 17: 307-322.

Phakiti, A. (2003a). 'A Closer Look at Gender and Strategy Use in L2 Reading.' *Language Learning*, 53/4: 649-702.

Phakiti, A. (2003b). 'A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance.' *Language Testing*, 20/1: 26-56.

Purpura, J. (1997). 'An analysis of the relationships between test-takers' cognitive and metacognitive strategy use and second language test performance.' *Language Learning*, 47/2: 289-325.

Rupp, A., Ferne, T. & Choi H. (2006). 'How assessing reading comprehension with multiple-choice questions shapes the construct: a cognitive processing perspective.' *Language Testing,* 23/4: 441-474.

Salataci, R. & Akyel, A. (2002). 'Possible effects of strategy instruction on L1 and L2 reading.' *Reading in a Foreign Language*, 14/1: 1-17.

Shokrpour, N. (2004). 'Systemic functional grammar as a basis for assessing text difficulty.' *Indian Journal of Applied Linguistics*, 30/2: 5-26.

Spadorcia, S.A. (2005). 'Examining the text demands of high-interest, low-level books.' *Reading & Writing Quarterly*, 21/1: 33-59.

Sudman, S., Bradburn, N. & Schwarz, N. (1995). *Thinking about answers: the application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass Publishers.

Trites, L. & McGroarty, M. (2005). 'Reading to learn and reading to integrate: new tasks for reading comprehension tests.' *Language Testing*, 22/2: 174-210.

Weir, C. (2005). 'Limitations of the Common European framework for developing comparable examinations and tests.' *Language Testing*, 22/3: 281-300.

'Co-construction' in the B2 and C1 KPG oral exams:
a comparison of examiners as a factor involved in candidates' performance

['Συν-οικοδόμηση' στις προφορικές εξετάσεις ΚΠΓ, Β2 και Γ1 επιπέδου: Σύγκριση των εξεταστών ως παράγοντα που εμπλέκεται στην απόδοση του υποψηφίου]

Xenia Delieza

Researchers who investigate oral testing invariably allude to the complexity of the procedure residing in the multitude of factors which influence its final outcome. One of these factors is the examiner who has been found to affect test takers' performance through his/her role both as interlocutor and rater. There has also been a long discussion on the characteristics of this role in the so-called oral paired-exam in comparison to the oral proficiency interview. The present paper looks into two oral tests of the same examination battery, both of which are paired in that two candidates go into an examination room where there are two examiners. However, only in one of the two tests do the candidates engage in a paired activity. This article aspires to describe the differences between the two tests in terms of the ways in which the examiner is involved in the candidates' language performance and discuss the implications of the findings for the two types of oral examination.

൦ช

Οι επιστήμονες που διερευνούν την προφορική δοκιμασιολογία αναφέρονται μονίμως στη πολυπλοκότητα αυτής της διαδικασίας η οποία και οφείλεται στο πλήθος των παραγόντων οι οποίοι επηρεάζουν το τελικό της αποτέλεσμα. Ένας από αυτούς τους παράγοντες είναι ο εξεταστής, ο οποίος έχει αποδειχτεί ότι επηρεάζει την απόδοση του υποψηφίου μέσω του ρόλου του και ως συνομιλητή και ως βαθμολογητή. Γίνεται επίσης μεγάλη συζήτηση για τα χαρακτηριστικά αυτού του ρόλου στην ονομαζόμενη εξέταση σε ζευγάρια σε αντιδιαστολή με τη συνέντευξη γλωσσομάθειας προφορικού λόγου. Το παρόν άρθρο εξετάζει δύο εξετάσεις προφορικού λόγου του ιδίου πιστοποιητικού γλωσσομάθειας, και οι δύο από τις οποίες γίνονται σε ζευγάρια καθώς δύο υποψήφιοι εισέρχονται σε ένα χώρο εξέτασης όπου βρίσκονται δύο εξεταστές. Παρόλ'αυτά, μόνο σε μία από τις δύο εξετάσεις εμπλέκονται οι υποψήφιοι σε δραστηριότητα για ζευγάρι υποψηφίων. Αυτό το άρθρο αποβλέπει στο να περιγράψει τις διαφορές μεταξύ των δύο εξετάσεων όσον αφορά

την εμπλοκή του εξεταστή στην γλωσσική απόδοση του υποψηφίου και να συζητήσει τη σημασία των ευρημάτων για τα δύο είδη προφορικής εξέτασης.

Introduction

High-stakes oral proficiency examinations have been synonymous with the oral proficiency interview (OPI), where candidates are usually individually examined by an examiner in a specially arranged room. One of the most important issues arising because of the OPI nature is that many factors or facets are involved in and interact during the assessment process (Ross, 1992; McNamara, 1995, 1997; Lazaraton, 1996; Milanovic & Saville, 1996; McNamara & Lumley, 1997; Brown, 2003, 2005; among others) such as tasks, examiners as 'interlocutors' and 'raters', candidates and assessment criteria.[1] For this reason, researchers have frequently used the term *co-construction* of the candidates' language output (Jacoby & Ochs, 1995; McNamara, 1995; He & Young, 1998; Fulcher, 2003; Brown, 2003, 2005; May, 2010, among others), first introduced by Kramsch (1986) proposing the so-called *Interactional Competence Theory* as an applied linguistics approach. The term co-construction captures what McNamara (1997, p. 459) calls 'the social dimension of interaction'. It refers to the fact that candidates do not speak alone, but their performance is collaboratively constructed, i.e. co-constructed by all the co-participants' contributions; this claim raises concerns for construct validity and reliability in oral testing.

Since the examiner is one of the co-participants, their role has been investigated in many empirical studies. Some researchers have emphasised the asymmetric nature of the interaction taking place in oral interviews; since the examiner is the one who controls the speech event, (that is the candidate cannot introduce topics or change the direction of the conversation), the validity of the interview as a representation of real life interactions is threatened (van Lier, 1989; Perret, 1990; Young and Milanovic 1992; Johnson, 2001; Csepes, 2002). Additionally, studies have looked into examiners' variation, their distinct or personal styles and the ways in which these might affect the candidates' performance (Ross, 1992; Ross & Berwick, 1992; Lazaraton, 1996; Brown & Lumley, 1997; Brown, 2003, 2005). Research on the issue of inter- and intra-examiner-as-interlocutor variation concludes that it might threaten the validity and reliability of any examination, and therefore, should be dealt with continuous monitoring of the examiners and examiner training.

The paired exam

Growing awareness of the aforementioned issues has oriented language testers towards the introduction of the paired testing pattern, during which there is one or more activities of interaction between the two candidates. The pair-work approach has its origins in classroom language learning throughout the world.

An increasing amount of literature provides theoretical and empirical data about the paired approach testing promoting its advantages. Ikeda (1998, p. 71) proposes the paired learner interview as 'an effective means to reduce communicative stress [...] and elicit authentic learner participation'. Iwashita (1996) suggests that peer-peer interaction creates a non-threatening environment and

generates similar scores to those obtained by the traditional interview (also in Norton, 2005). Együd and Glover (2001) argue that candidates can use language not as 'inferiors' (candidates) addressing 'superiors' (examiners), but as they normally use it in everyday speech situations. Saville and Hargreaves (1999) and Galazci (2003) argue that the paired format allows for more varied patterns of interaction, which is also advocated by Taylor (2000, 2001), who talks about a greater range of functions in the paired speaking test, and Brooks (2009) who found higher complexity in the test takers' interaction. Other researchers have looked into the paired format from the raters' perspective. May (2009) and Ducasse and Brown (2009) conclude that raters do recognise and assess the way(s) candidates contribute to successful interaction. And although discourse in paired activities is collaboratively produced by definition since there are two interlocutors-candidates who must be assessed separately, such co-construction is not regarded as a negative feature of the paired activity, but should be embraced as being more reflective of real world communication (Brookes, 2009, p.361). As long as interlocutor variability in peer-peer interaction is regarded as 'part of the ability construct we are interested in measuring' (Taylor & Wigglesworth, 2009, p.332) and appears to 'directly inform the assessment scale' (Nakatsuhara, 2003, p 22), assessment is achieved in a reliable and valid way.

This paper – being part of a more extensive PhD research – looks into the paired-type activity from the point of view of the examiner and the potential value of this testing pattern in the elimination of examiner variation. The study compares existing variation in examiners' performance in oral exams of two different levels by presenting a comparative study of the role of the examiner-as-interlocutor in the two tests. It also draws some conclusions concerning the examiners' role as this depends on the level and type of activity (paired or not paired). Finally, it reports on data from three different sources: observation of actual oral examinations, oral examiners' feedback forms and simulated oral examinations.

The context of the study

The present research was conducted by the Research Centre for Language Teaching, Learning and Assessment (RCEL) of the Faculty of English Studies, University of Athens, Greece[2], within the context of the Greek State Certificate of English Language Proficiency exams, known as KPG exams, and more specifically, the oral tests at B2[3] and C1 levels. These are the two levels the KPG battery started with, and also, in which the role of examiner as interlocutor has been defined quite differently by construct (see below). In both levels, two candidates enter the examination room and are examined by one examiner, while another examiner is also present. This second examiner is an observer who only assigns marks and does not participate in the speech event. At the end of the test, the examiner who conducts the interview also marks the two candidates. Both examiners use the same set of criteria, i.e. rating scale, to assign marks for separate skills or competences which are then added to produce two total marks (one from each examiner). The average of the two marks provides each candidate's final oral test mark. (See Appendix for the content and structure of the KPG oral tests at the B2 and C1 levels).

The role of the examiner as interlocutor differs in the two levels. At B2 level, the examiner reads out the questions (Activity 1) and tasks (Activities 2 and 3) to each of the two candidates, thus interacting with each one of them in turn. The candidates go into the examination room together, for reasons of organisation and time economy, but they do not speak to each other at all; thus the speaking test is not a paired test in the form traditionally discussed in the literature. On the other hand, at C1 level, the examiner reads out one question (for each test-taker) in Activity 1 for the purpose of which s/he

interacts with each of the two candidates; then in Activity 2 s/he becomes a listener while the candidates engage into interaction in order to reach a decision or solve a set problem on the basis of input presented in reading texts in Greek.

Regarding the rating scale, candidates are assessed on the basis of different types of criteria in the two levels, (see Table 1). Only at B2 level, they are evaluated for task achievement. Only at C1 level, test-takers are assessed for interactional competence because there is a task-type requiring interaction between the two candidates.[4] There are some more differences but they fall outside the scope of the present article.

B2	C1
<u>**Overall performance on task**</u> (i.e. the degree to which the candidate has responded to the requirements of the task)	
■ Dialogue (0-2)	
■ One sided talk (0-2)	
■ Mediation (0-2)	
<u>**Overall language performance**</u> (i.e. the quality and level of candidates' output in relation to certain criteria)	
■ Phonological competence (0-2)	■ Phonological competence (1-2)
■ Linguistic competence (0-4)	■ Lexical range and control (0-3) ■ Grammatical accuracy (0-3)
■ Sociolinguistic competence (0-4)	■ Appropriateness of language choices (0-3)
■ Pragmatic competence (0-4)	■ Cohesion/coherence/fluency (0-3) ■ Conversational competence (0-3)
	■ Mediation (0-3)

Table 1: The B2 and C1 Oral Assessment Criteria

The study

The present study is based on the many-relevant-research findings that inter- and intra-examiner variation might threaten the validity and reliability of the examination and therefore should be controlled. More specifically, it examines whether variation, also called *intervention* or *involvement*, differs according to the level of the exam as well as the design of it.

Research questions

This paper seeks to answer the following questions.
- In what ways does examiner-as-interlocutor involvement differ between B2 and C1 KPG oral tests?
- Can the differences be attributed to the level, the type of activities involved, or both?

This study draws data from three different sources to shed light into the ways examiners-as-interlocutors affect the candidates' language output in two very different tests of oral proficiency. These three sources are separately described below along with their results, which are further discussed in the final section of this article.

The KPG Oral Examiners Observation project

The KPG observation project commenced as a pilot study in November 2005. Since then, it has been conducted in five more phases providing invaluable information about the efficiency of oral examiner conduct as well as other issues related to the test procedure and administration. As such, the observation project constitutes an on-going effort of the RCEL to control and monitor examiners' performance in the KPG oral tests. It is carried out in the biggest examination centres all around Greece, thus producing a representative amount of data in terms of both quality and quantity (see also Delieza, 2008 and Karavas & Delieza, 2009)

Data for this paper are drawn from the third and fourth phases of the observation project, carried out in May and November 2007. These phases aimed to collect information in relation to the type and frequency of interlocutor intervention in each of the test activities. Through previous piloting observation phases, the assigned observers (the writer herself among them) had detected and listed various types of interlocutor interventions, which are presented in Table 2 below; these types were classified into two general categories, namely *change or interference with the questions/tasks rubrics* and *interruption of candidate or interference with their language output*.[5]

Changes to or interference with the rubrics	Interruption of the candidate or interference with his/her language output in order to
• Use of an introductory question • Change of one-two words from the rubric • Supplying a synonym for a word • Expansion of the original exam question • Explanation of the rubric (through the use of examples) • Repetition of the rubric (more slowly)	• redirect the candidate because s/he misunderstood something by repetition of the rubric or part of it • make some kind of correction • supply one or more words the candidate was unable to find • add something

Table 2: Types of examiner intervention identified by observers.

Since this paper looks into the differences between the two levels in terms of the role of the examiner-interlocutor, I only present results directly relevant to this issue. Thus, in May 2007, 32 observers observed 156 examiners who examined 588 candidates for the B2 level oral test and 105 examiners examining 342 candidates for the C1 level oral test. In November 2007, 42 observers observed 133 examiners examining 514 B2 candidates and 66 examiners examining 232 C1 candidates.

Table 3 presents results (percentages and numbers) for the two general types of intervention for each Activity in the two levels for the two observation phases.[6] Percentages of interventions have been calculated on the basis of the total of candidates observed, since instances of intervention have been counted per candidate.

Comparing the results from the two phases it is obvious that C1 examiners generally tend to intervene less than B2 examiners. It becomes clear that C1 examiners make changes to the question/task rubrics much less frequently than B2 examiners; this is seen in both sets of findings (for instance, May 2007, Activity 1 – B2=57% as opposed to C1=22.5%). Examiners, therefore, appear to feel the need to provide support to and/or facilitate candidates of B2 more than those of C1 level. Furthermore, percentages in November 2007 are generally higher in all B2 Activities and in C1

Activity 1, in which the examiner is the candidate's interlocutor, than in C1 Activity 2, in which the two candidates interact, the examiner being a mere listener.

May 2007 Intervention per activity	B2 (588 in total)			C1 (342 in total)	
	Activity 1	Activity 2	Activity 3	Activity 1	Activity 2
Changes to the rubrics	57.5% (338)	31% (184)	33% (194)	22.5% (77)	23.5% (80)
Interruptions and/or interferences	20.5% (121)	33.5% (198)	28.5% (168)	*36%* (123)	*55.5%* (188)

November 2007 Intervention per activity	B2 (514 in total)			C1 (232 in total)	
	Activity 1	Activity 2	Activity 3	Activity 1	Activity 2
Changes to the rubrics	30% (155)	23% (117)	22% (112)	21.5% (53)	13.5% (31)
Interruptions and/or interferences	25% (128)	40% (204)	30% (152)	*34.5%* (80)	21% (49)

Table 3: Results of examiner intervention per activity in the B2 and C1 level oral tests

There are three percentages (depicted in bold letters in the table) which appear to contradict the assumed preference of examiners to intervene in B2. Firstly, *36%* of the C1 examiners in May 2007 interrupted the candidates or interfered with their language output in some way in Activity 1 and *55.5%* in Activity 2. Karavas and Delieza (2009) attribute these percentages to the tendency of examiners to a) expand on the given opinion question in Activity 1, where the candidates sometimes do not produce enough assessable language; and, b) intervene in Activity 2 in order to remind the candidates of test procedures and requirements which they sometimes forget, since the two of them are in a process of information exchange and interaction, which the examiner is supposed to listen and only *monitor* if necessary. In Activity 1, where examiners are interlocutors, they appear to be *facilitators* of the talk, while in Activity 2, where they are listeners, they act as *instructors*, ensuring the procedure is conducted according to regulations.

The third somewhat odd percentage is *34.5%* in C1 Activity 1, in November 2007, since examiners are not expected to intervene to such an extent at this level. Again, it can be explained by the fact that examiners as interlocutors in Activity 1 often expand on the candidate's answer in order to help them produce longer or more complete answers.

Feedback from Examiners

All KPG oral examiners participating in the oral examination are asked to complete *anonymously* the Oral Examiner Feedback Forms at the end of each examination day. Some of the content of these forms has varied from some examinations to others, but their core questions always refer to the efficiency of the questions and tasks. With the introduction of an Interlocutor Script[7] in the English KPG oral test which was introduced in November 2007, some questions relevant to the examiner conduct were added in the feedback forms. These questions concerned the use and efficiency of the newly introduced Interlocutor script; they also asked the examiners to state how often they changed or interfered with the questions or tasks rubrics and how often they interrupted the candidates while they were talking. Table 4 presents the questions (4a, 4b and 6a, 6b) from the feedback form which relate to this study, along with the results for the two levels. For B2 level, 258 Feedback forms

and for C1 level 180 Feedback forms were collected and analysed.[8] Table 4 shows the frequencies of what examiners themselves stated they did in terms of activity rubrics and interruptions. The last column shows the percentage of forms in which the questions were answered, as in a few cases they were left blank.

B2 level (total of feedback forms=258)				
4. Did you change or interfere with the rubrics in any of the following ways?	**VERY OFTEN**	**SOMETIMES**	**NEVER**	**% of forms**
a. Change one-two words and/ or supplying a synonym for a word.	2.71% (7)	66.67% (172)	29.46% (76)	98.84% (3 not answered)
b. Expand the question and/ or use examples to explain.	2.71% (7)	50.78% (131)	43.02% (111)	96.51% (9 not answered)
6. Did you generally interrupt the candidates or intervene while they were talking in order to:	**VERY OFTEN**	**SOMETIMES**	**NEVER**	**% of forms**
c. correct or add information?	0% (0)	20.54% (53)	72.48% (187)	93.02% (18 not answered)
d. help the candidate by repeating the whole or part of the question?	3.88% (10)	71.32% (184)	21.71% (56)	96.91% (8 not answered)
C1 level (total of feedback forms=180)				
4. Did you change or interfere with the rubrics in any of the following ways?	**VERY OFTEN**	**SOMETIMES**	**NEVER**	**% of forms**
a. Change one-two words and/ or supplying a synonym for a word.	2,22% (4)	38,33% (69)	57,78% (104)	98.33% (3 not answered)
b. Expand the question and/ or use examples to explain.	32,22% (58)	0,56% (1)	62,22% (112)	95% (9 not answered)
6. Did you generally interrupt the candidates or intervene while they were talking in order to:	**VERY OFTEN**	**SOMETIMES**	**NEVER**	**% of forms**
c. correct or add information?	0,56% (1)	14,44% (26)	77,78% (140)	92.78% (13 not answered)
d. help the candidate by repeating the whole or part of the question?	46,67% (84)	3,33% (6)	45,56% (82)	95.56% (8 not answered)

Table 4: Results for questions 4 and 6 of the November 2007, B2 and C1 Oral Examiner Feedback Forms

Almost 67% of the B2 examiners who answered question 4 sometimes make some kind of change to the question or task rubric given in the Examiner Pack, and almost 51% expand the question or use examples to explain it. Because the questions are verbalised in a generalised way, it is not clear how the examiners exactly interfere with the given rubrics (henceforth the need for discourse analysis). However, more than two thirds of the examiners admit that they sometimes 'tamper' with the questions or tasks, which is a major threat to the reliability of the examination.

In question 6, findings are more encouraging. Almost 21% of the examiners, who answered this question, state that they sometimes corrected the candidates or added information while they were talking, while 72% did not do so at all. 21% is not negligible, this being a practice which examiners have been advised to avoid. Although the use of correction or elaboration can be explained by the

fact that KPG examiners are also EFL teachers, it cannot but be eliminated as a source of examiner variability which may threaten the validity and reliability of the test. Additionally, around 71% of the examiners stated that they sometimes used repetition of the whole or part of the questions or tasks. Repetition has been promoted by KPG oral exam designers and examiner trainers as an efficient way of helping the candidate out of a trouble situation (e.g. being 'stuck' or showing lack of understanding etc.) without providing the linguistic means to do so; i.e. without supplying the candidate with language which s/he is expected to produce.

Findings for C1 are different from B2. First of all, almost 58% and 62% of the examiners, respectively, stated that they never changed the rubrics or expanded them etc. (questions 4a and 4b). This could be attributed to two factors, the first being the level of the candidates' linguistic competence. The second is the amount of questions and tasks in the C1 test. The examiner is supposed to ask *one* opinion question in Activity 1 (as opposed to *two* to *four* personal questions in Activity 1 in B2). Moreover, the examiner assigns one task in Activity 2 (as opposed to two tasks, one for Activity 2 and one for Activity 3 in B2) in which s/he is a mere listener, allowing the candidates to engage in a long conversation (as opposed to his/her being each candidate's interlocutor). However, percentages are quite high for 'sometimes' (38%) in 4a and for 'very often' (32%) in 4b. Observation of actual exams has shown that these types of intervention may be connected with the Activity 1 opinion question which is sometimes a source of 'trouble' in two ways. Either a word or phrase is incomprehensible for the candidate or the candidate fails to answer fully, both of which cause explanation and/or expansion of the given question.

In relation to question 6c, the answers are similar with B2 level: almost 78% of the examiners stated that they never correct candidates or add information to their language production. Concerning question 6d, however, almost half of the examiners state that they *very often* use repetition to help the candidates or that they *never* do so. This can be explained as follows. Firstly, repetition is one of the strategies examiners have been advised (by the KPG exam developers) to utilise with candidates in their effort to assist language production. Secondly, drawing from experience, C1 candidates sometimes ask for repetition of the opinion question in Activity 1 and also need to be reminded of part of their task while they have been interacting for some time in Activity 2. When repetition is not used, it is probably because higher proficiency candidates may not ask or need to be reminded of the question or task as often as lower proficiency ones.

To conclude, it is evident that examiners at B2 level tend to intervene more often than in C1, by changing rubrics, or expanding the question/task or through correction and addition of information and also repeating the whole or part of the question/task. It appears that examiners at B2 level seem to try to facilitate candidates more than at C1, either because of the candidates' level or because of the role they are supposed to play in the communicative event, or for both of these reasons.

Simulated Oral Tests

From October to December 2006, 14 simulations of the actual KPG oral exams were conducted (7 for each level) with learners preparing to take the KPG exams in May 2007. These simulations were conducted by highly experienced KPG examiners but without the Interlocutor Script – since this was only introduced in November 2007. These simulations were video- and audio-recorded and then transcribed. The transcribed data analysis produced valuable findings both in quantity and quality. The coding of the types of intervention was done through careful study of the transcripts; different categories arose inductively through analysis of the raw data and also deductively through use of the

findings from observation and relevant studies (Ross, 1992; Ross & Berwick, 1992). Within the limits of the present article, I will present some quantitative data which can be compared to the data presented in the two sections above. Table 5 presents the most frequently used types for each activity in the two levels.

The categories of intervention outlined in Table 5 are not directly comparable to the ones produced though observation and feedback forms. This comes as a result of the discourse analysis and coding of the transcribed data. Close examination of the language used by examiners revealed similarities among some types, as these were defined in observation and feedback collection, but rendered their 'boundaries' unclear. Therefore, types of involvement were reconsidered and re-defined. On the basis of the new analysis, *expansion* includes any addition to the candidate's language production, i.e. providing words or phrases and asking further questions. *Repetition* was defined as a general type including repetition of a question, a task or part of them and repetition of the candidates' word(s). *Explanation* is a type which includes mainly explanation of a word/words in the given question or task, and this differentiates it from expansion. Finally, *comment/evaluation* is a type which came to light through this study and had scarcely been detected by observers.

B2	No of interventions per activity	Total no of interventions = 279			
		expansion	repetition	explanation	comment /evaluation
Act 1	133	8.6% (24)	*12.2% (34)*	4.3% (12)	3.9% (11)
Act 2	79	9.3% (26)	10.8% (30)	0.7% (2)	0.7% (2)
Act 3	67	*12.5%* (35)	4.7% (13)	0.7% (2)	0.4% (1)
C1	No of intervention per activity	Total no of interventions = 86			
		expansion	repetition	explanation	comment /evaluation
Act 1	50	*15.1%* (13)	7% (6)	10.5% (9)	5.8% (5)
Act 2	36	*18.6%* (16)	5.8% (5)	2.3% (2)	0.0% (0)

Table 5: Most frequent types of intervention in simulated oral tests in the B2 and C1 levels

As shown in Table 5, examiner intervention is much more frequent in B2 than in C1, which was also indicated by the observation and feedback forms results. Additionally, it is in B2-Activity 1 that examiners mostly prefer to involve themselves in the candidates' language output. It could be argued that examiners tend to facilitate candidates because Activity 1 is the introductory activity to the whole examination and consists of a series of (two to four) personal questions asked by the examiner-interlocutor. Repetition is the most frequently used type of intervention (12.2%) in this activity, while expansion also lies among the most preferable choices of examiners (8.6%).

Explanation and comments/evaluation are also frequent. It appears that examiners involve themselves in more ways than they do in Activities 2 and 3 in B2 level, in which intervention is much less frequent in the first place (79 and 67 as opposed to 133). In Activity 2, examiners most frequently choose to make repetitions (10.8%) but also expansions (9.3%), the frequency of these two types being very close, while the percentages for explanation and comment/evaluation are almost negligible. In Activity 3, expansion (12.5%) is much more frequent than the rest, some of which are scarce. Activities 2 and 3 are more cognitively challenging (according to their construct definition) requiring the candidates to carry out a task on the basis of one or more pictures and

Greek input text respectively. Given that these tasks are sometimes long and consisting of sub-questions (see example in Appendix), examiners tend to facilitate the candidates either by repeating part(s) of them or expanding them in order to prompt the candidates to speak.

C1 examiners get remarkably less frequently involved than B2 examiners (86 interventions as opposed to 279 respectively), which can be attributed to their expectations of the candidates' proficiency level as defined by the specifications as well as the nature of the role they play in the process. C1 examiners ask each candidate one opinion question (Activity 1) and then assign a task to both candidates to be conducted on the basis of Greek texts (Activity 2), thus restricting the role of the examiner to that of a listener. As can be seen in Table 5, examiners get more involved in Activity 1 than in Activity 2 (50 as opposed to 36 interventions). In Activity 1, they most frequently expand on or explain the question and much less frequently repeat it or make comments/evaluations. This can be attributed to two reasons: a) the opinion question itself causes trouble because of a word or phrase which creates difficulty[9] and b) according to the examiner, the candidate does not fully answer the question. In Activity 2, although the percentage appears to be the highest (18.6%), in fact, it is only 16 times that expansion was used in all 7 simulated tests. These 16 interventions were actually against test-conduct instructions and can be attributed to the lack of Interlocutor script and the personality and choices of the specific examiner, who opted for a more 'interventionist' role in this Activity. Such tactics are not uncommon in actual exams, thus threatening the validity and reliability of the procedure and its outcome.

Discussion and conclusions

This article presents data from three studies investigating the differences in interlocutor conduct and the extent to which they depend on the level, the paired activity format or both. Although the three studies are not directly comparable due to differences in methodology, they all offer insights into interlocutor variation.

First of all, examiners tend to use different types of intervention, and therefore vary in the way they conduct the test, more frequently in the B2 exam than the C1. It appears that when they are given the role of the sole interlocutor of the candidate, they have the tendency to use more facilitative techniques –even if they have been advised to be mere deliverers of questions and tasks. On the other hand, they prefer to intervene mainly in a more instructive role when they are listeners of a paired-type task. This is especially evident in the C1 exam, where examiners get more involved in Activity 1. Examiners themselves ask each candidate an opinion question while in Activity 2, they remain silent, listening to the candidates' interaction and interfere mainly for instructive reasons.

Additionally, comparing all activities in both levels, examiner intervention is more frequent in Activity 1 of the B2 exam since it consists of personal questions which seem to be more prone to changes or expansions by the examiner-as-interlocutor. Although this activity is called 'Dialogue' and it is not supposed to be conducted as any dialogue in a real-time situation, examiners seem to be undertaking a 'freer' and/or more accommodating kind of interlocutor role. It is, finally, possible that, because it is the very first activity, examiners tend to use facilitation strategies in order to encourage the candidates. In contrast, examiners generally refrain from getting involved in C1 Activity 2, which is the only peer-peer interaction activity. Examiners mainly intervene in an instructive role, to repeat the task or remind the candidates of procedure details.

The analyses also showed that repetition is a strategy frequently used by examiners. This is a very positive fact for the way KPG exams are conducted, because repetition can be a type of involvement which will not provide the candidates with the linguistic means to continue thus affecting their language output.

The results presented in this paper contribute to the literature on examiner variation in oral tests by shedding light into differences between levels of oral proficiency (KPG B2 and C1) as well as between two activity patterns (paired and non-paired). It also provides evidence that the involved candidates' performance seems to be co-constructed by examiners in cases where the latter intervene. Evidently, further study into the effect of types of variation on candidate language output as well as on the final score could further support this idea of co-construction. Nevertheless, it could be stated that examiners tend to be factors of co-construction more frequently in the B2 exam than in the C1 exam and more frequently in the non-paired activities than in the paired ones. This has implications for oral examiner training in the assessment of proficiency at different levels and also supports the value of the paired activity itself.

In conclusion, examiner variability in oral proficiency tests has always been a major concern for high-stakes proficiency testers. Continuous research into and monitoring of oral tests provides information about the high complexity of the oral test procedure. This article shows in what ways examiners vary in the practices they use when conducting two tests whose level and interactional pattern differ. It has offered some insight into the ways examiners collaboratively construct the candidates' performance and highlighted a positive aspect of paired oral activities.

However, more research into oral tests is required, since many variables affect such procedures – whatever the types of activities involved in them. Additionally, the co-constructed nature of oral performance in both paired and non-paired activities is an issue awaiting more empirical research, because safe conclusions should be drawn on the ways co-construction is or can be included in assessment scales or criteria and then internalised, interpreted and applied by raters.

Evidently, such research results provide material for continuous and vigorous training of examiners and raters aiming at dealing with the complexity and elusiveness of oral communication in tests of oral proficiency.

Author's email: xdelieza@enl.uoa.gr

Notes

1. Milanovic and Saville (1996, p. 6) propose a most comprehensive diagram of variables interacting with each other and affecting assessment.
2. The writer – in her capacity as a research assistant at the RCEL – coordinated (and *participated as real-time observer* in) the observation project with KPG candidates from November 2005 to November 2008 examination periods. This project is part of a larger ongoing research project, directed by Prof. V. Dendrinos and Assist. Prof. E. Karavas, investigating the KPG oral exam.
3. Since May 2011, B1 and B2 have become an integrated exam.
4. B2 – Activity 1 is a Dialogue only to the extent that examiners deliver 2-4 personal questions and test-takers answer them – they do not engage in some kind of conversation and, for this reason, are not assessed for interactional competence (for details see Karavas, 2009).
5. Karavas and Delieza (2009) present the results from the May 2007 phase with special reference to a) their impact on the improvement of the oral test by means of the use of an interlocutor script – which was introduced in all English KPG oral exams in November 2007; and b) on the training of the oral examiners – who were re-trained and supplied with a list of 'acceptable and non-acceptable types of intervention' and were further evaluated on the applicability of these instructions through observation again in November 2007.
6. For detailed results of different types of intervention per activity for May 2007, see Karavas and Delieza (2009).
7. An Interlocutor Script or Frame is a set of written instructions which examiners read out to candidates guiding them through the examination; it is similar to the scripts actors use when rehearsing their lines.
8. C1 feedback forms were fewer than the B2 ones because the C1 level usually has at least 50% fewer candidates than the B2 level.
9. They sometimes contain sayings or difficult words – see example in Appendix.

References

Brooks, L. (2009). 'Interacting in pairs in a test of oral proficiency: Co-constructing a better performance.' *Language Testing,* 26/3: 341-366.

Brown, A. (2003). 'Interviewer variation and the co-construction of speaking proficiency.' *Language Testing,* 20/1: 1-25.

Brown, A. (2005). *Interviewer variability in oral proficiency interviews*. Frankfurt am Mein: Peter Lang.

Brown, A., & Lumley, T. (1997). 'Interviewer variability in specific-purpose language performance tests.' In A. Huhta, V. Kohonen, L. Kurki-Suonio & S. Luoma (Eds.), *Current developments and alternatives in language assessment: proceedings of LTRC 96*. Jÿvasskÿla: University of Jÿvasskÿla and University of Tampere, 173-191.

Csepes, I. (2002). 'Measuring oral proficiency through paired performance.' Unpublished PhD Dissertation, Eötvös Loránd University.

Delieza, X (forthcoming). 'Monitoring KPG examiner conduct.' *Directions,* 1. RCEL publications, University of Athens.

Ducasse, A.M. & Brown, A (2009). 'Assessing paired orals: raters' orientation to interaction.' *Language Testing*, 26/3: 423-443.

Együd, G. & Glover, P. (2001). 'Oral testing in pairs: a secondary school perspective.' *ELT Journal*, 55/1: 70-76.

Fulcher, G. (2003). *Testing second language speaking*. London: Pearson Longman.

Galazci, E.D. (2003). 'Interaction in a paired speaking test: the case of the First Certificate in English.' *Cambridge ESOL Research Notes*, 14: 19-23.

He, A. W. & Young, R. (1998). 'Language proficiency interviews: a discourse approach.' In R. Young, & A. W. He, (Eds) *Talking and testing. discourse approaches to the assessment of oral proficiency.* Amsterdam: John Benjamins, 1-23.

Ikeda, K. (1998). 'The paired learner interview: a preliminary investigation applying Vygotskian insights.' *Culture and Curriculum,* 11/1: 71-96.

Iwashita, N. (1996). 'The validity of the paired interview in oral performance assessment.' *Melbourne Papers in Applied Linguistics*, 5/2: 51-65.

Jacoby, S. & Ochs, E. (1995). 'Co-construction: an introduction.' *Research on Language and Social Interaction*, 28/3, 171-183.

Johnson, M. (2001). *The art of non-conversation: a re-examination of the validity of the oral proficiency interview*. New Haven, CT: Yale University Press.

Karavas, E. (Ed.) (2009). *The KPG speaking test in English: a handbook*. University of Athens: University of Athens RCEL Publications.

Karavas, E. & Delieza X. (2009). 'On site observation of KPG oral examiners: implications for oral examiner training and evaluation.' *Apples – Journal of Applied Language Studies*, 3/1: 51-77.

Kramsch, C. (1986). 'From language proficiency to interactional competence.' *Modern Language Journal*, 70, 366-372.

Lazaraton, A. (1996). 'Interlocutor support in oral proficiency interviews: the case of CASE.' *Language Testing* 13: 151-172.

May, L. (2009). 'Co-constructed interaction in a paired-speaking test: the rater's perspective.' *Language Testing*, 26/3: 397-421.

May, L. (2010). 'Developing speaking assessment tasks to reflect the 'social turn' in language testing.' *University of Sydney Papers in TESOL,* 5: 1-30.

McNamara, T. F. (1995). 'Modelling performance: opening Pandora's box.' *Applied Linguistics* 16/2: 159-179.

McNamara, T. F. (1997). 'Performance testing'. In C. Clapham & D. Corson (Eds.), *Language testing and assessment.* Dordrecht: Kluwer Academic Publishers, 131-139.

McNamara, T. F., & Lumley, T. (1997). 'The effect of interlocutor and assessment variables in overseas assessment of speaking skills in occupational settings.' *Language Testing,* 14: 140-156.

Milanovic, M., and Saville, N. (1996). *Performance testing, cognition and assessment: selected papers form the 15th language testing research colloquium*. Cambridge: Cambridge University Press.

Nakatsuhara, F. (2006). 'The impact of proficiency-level on conversational styles in paired speaking test.' *Cambridge ESOL Research Notes*, 25: 15-20.

Norton, J. (2005). 'The paired format in the Cambridge speaking tests.' *ELT Journal*, 59/4, 287-297.

Perret, G. (1990). 'The language testing interview: a reappraisal.' In J. de Jong & D. K. Stenenson (Eds.) *Individualising the assessment of language abilities*. Philadelphia. PA: Multilingual Matters, 225-238.

Ross, S. (1992). 'Accommodative questions in oral proficiency interviews.' *Language Testing* 9: 173-185.

Ross, S., and Berwick, R. (1992). 'The discourse of accommodation in oral proficiency interviews.' *Studies in Second Language Acquisition,* 14: 159-179.

Saville, N & Hargreaves, P. (1999). 'Assessing speaking in the revised FCE.' *ELT Journal*, 53: 42-51.

Taylor, L. (2000). 'Investigating the paired speaking test format.' *Cambridge ESOL Research Notes*, 2: 14-15.

Taylor, L. (2001). 'The paired speaking test format: Recent Studies.' *Cambridge ESOL Research Notes*, 6: 15-17.

Taylor, L. & Wigglesworth, G. (2000). 'Are two heads better than one? pair work in L2 assessment contexts.' *Language Testing*, 26/3: 325-339.

van Lier, L. (1989). 'Reeling, writhing, drawling, stretching, and fainting in coils: oral proficiency interviews as conversations.' *TESOL Quarterly,* 23/3: 489-508.

Young, R. & Milanovic, M. (1992). 'Discourse variation in oral proficiency interviews.' *Studies in Second Language Acquisition,* 14: 403-424.

Appendix

The B2 and C1 levels exam structure and content accompanied by *examples* for all activities.

B2	C1
Duration of test	
15-20 minutes	20 minutes
Pattern of participation	
Candidates are tested in pairs but do not converse with each other.	Candidates are tested in pairs and converse with each other in Activity 2.
Content of oral test	
a) **Dialogue** (3-4 minutes) between examiner and each candidate who answers questions about him/herself and his/her environment posed by the examiner. For instance, the candidate is asked questions such as '*Do you have a lot of friends or just a few close ones? Tell us about them',* and/ or, '*Do you prefer listening to music at home or going to live performances? Why?*' and/ or, '*What would be the ideal school environment/ working environment for you? Why?*' (November 2007 – English KPG – B2 – Module 4, Examiner Pack, page 2). b) **One-sided talk** (5-6 minutes) by each candidate who develops a topic on the basis of a visual prompt. For instance, the candidate is shown a page with pictures depicting 'People's emotions' (November 2007 – English KPG – B2 – Module 4, Candidate Booklet, page 6) and the relevant task is '*Look at photos 1 & 2. Tell us how you think the people are feeling, what has happened and what is going to happen next'* (Examiner Pack, page 3). c) **Mediation** by each candidate who develops a topic based on input from a Greek text. (6 minutes for both) For instance, the candidate is given text in Greek about 'how to take care of contact lenses'. (November 2007 – English KPG – B2 – Module 4, Candidate Booklet, page 12) and the relevant task is '*Imagine I am going to wear contact lenses for the first time. Using information from Text 1, give me some advice on how to take care of them'* (Examiner Pack, page 4).	a) **Warm-up** (not assessed – 1 minute) Examiner asks each candidate a few ice-breaking questions (age, studies/work, hobbies) b) **Open-ended response** (4 minutes): The candidate responds to a single question posed by the examiner expressing and justifying his/her opinion about a particular issue/topic. For instance, the candidate is asked questions such as, **'***Do you think that some professions are more appropriate for men and some for women?'* or '*Do you think that the saying "Hard work never did anyone harm" is always true? Explain why or why not'* (November 2007 – English KPG – C1 – Module 4, Candidate Booklet, page 2. c) **Mediation and open-ended conversation** (15 minutes): Candidates carry out a conversation in order to complete a task using input from a Greek text. For instance, the candidates are given two different but related texts in Greek about 'ways of saving energy'. (November 2007 – English KPG – C1 – Module 4, Candidate Booklet, pages 6 & 11) and the relevant task is '*Imagine that you have been asked to design campaign leaflets on how to save energy at home. Read your texts, and together decide on the two most useful pieces of advice you would include'* (Examiner Pack, page 3).

Assessment practices in the English language classroom of Greek Junior High School

[Πρακτικές αξιολόγησης της Αγγλικής Γλώσσας στο Ελληνικό Γυμνάσιο]

Stavroula Vlanti

The present study aimed to compare teacher and student perceptions concerning assessment in the English language classroom of Greek Junior High Schools. It focuses on exploring teacher assessment practices and students' view of these. The degree of agreement between teacher and student views can affect the results of instruction and learning. The central finding of the study was that English language teachers follow an approach that keeps a balance between the requirements of the Cross-Thematic Curriculum (DEPPS) for performance assessment and the official specifications which define final achievement tests. Students understand the purpose of assessment and the importance of attitudes towards learning and have a clear picture of methods and tasks used for their assessment. On the basis of these research findings, a series of suggestions about teachers and teacher trainers are put forward, focusing on teacher professional development, which, in turn, will promote student involvement and responsibility for learning.

႘

Η παρούσα μελέτη επιχειρεί να συγκρίνει τις αντιλήψεις καθηγητών και μαθητών σχετικά με την αξιολόγηση της Αγγλικής γλώσσας στο Γυμνάσιο. Σκοπός της έρευνας είναι να διερευνηθούν οι πρακτικές που χρησιμοποιούν οι εκπαιδευτικοί για την αξιολόγηση και πώς οι μαθητές αντιλαμβάνονται αυτές τις πρακτικές. Ο βαθμός συμφωνίας μεταξύ καθηγητών και μαθητών στα υπό εξέταση ζητήματα μπορεί να είναι καθοριστικός για τα αποτελέσματα της διδασκαλίας και της μάθησης. Βασικό εύρημα της παρούσας μελέτης ήταν ότι οι εκπαιδευτικοί ακολουθούν μια προσέγγιση η οποία ισορροπεί ανάμεσα στις απαιτήσεις του Διαθεματικού Ενιαίου Πλαισίου Προγράμματος Σπουδών, το οποίο συστήνει την αξιολόγηση των μαθητών ως προς τη δυνατότητα χρήσης της γλώσσας, και τις επίσημες οδηγίες οι οποίες καθορίζουν τη μορφή των τελικών εξετάσεων. Οι μαθητές κατανοούν το σκοπό της αξιολόγησης και της σημασίας των στάσεων προς τη μάθηση και έχουν σαφή εικόνα των μεθόδων και των δραστηριοτήτων που χρησιμοποιούνται στην αξιολόγησή τους. Βάσει αυτών των δεδομένων, αναπτύσσεται μια σειρά από προτάσεις για τους εκπαιδευτικούς και τους επιμορφωτές με στόχο την επαγγελματική ανάπτυξη των

καθηγητών, η οποία ακολούθως θα οδηγήσει στη διευκόλυνση της συμμετοχής των μαθητών στη διαδικασία ης μάθησης και της αξιολόγησης.

Key words: Junior High School, language assessment, assessment practices, tests, alternative assessment.

Introduction

The teaching of English as a foreign language in the English language classroom of Greek Junior High Schools is dictated by the Cross-Thematic Curriculum Framework for Compulsory Education (2003), also known as 'Diathematikon Programma' (DEPPS). The section devoted to modern languages in this document has greatly been based on the Common European Framework of Reference for Modern Languages (CEFR, Council of Europe, 2001).

Nonetheless, the final achievement test is controlled by official specifications, which constitute a synthesis of various Presidential Decrees. According to these specifications, formal testing consists of the assessment of reading, through a particular pre-taught text by means of comprehension questions, and the testing of grammar and syntax. The assessment of other skills, namely writing and speaking, is not included in the prescribed form of the formal achievement test. Listening is tested in classes B´ and C´ by means of dictation.

However, the new curriculum for modern languages stresses the need to supplement traditional testing methods with alternative assessment techniques and defines students as partners in the process (Pedagogical Institute, 2003, p. 381). It refers to assessment as a continuous, transparent and valid procedure (ibid) and focuses on performance assessment and evaluation of student progress. The ultimate aim is to locate sources of student weaknesses in order to provide feedback and remedial teaching. The curriculum encourages the use of a range of assessment methods such as dynamic interaction between the participants in the teaching/learning process, teacher observation, student portfolio, self-assessment and peer-assessment. Moreover, it makes special reference to projects, which are considered to be the most creative approach to foreign language learning and defines students as partners in the process (ibid, p. 379).

In this context, the aim of the present study was to investigate how Greek Junior High School teachers balance assessment procedures in order to meet the demands of both the curriculum and official test specifications and how students experience assessment in their English language classrooms.

Literature Review

Cimbricz (2002) argues that teachers devote time to test preparation instead of instruction avoiding creative work, like projects and cooperative activities. However, other works (Chan, 2008) reveal that teachers believe that alternative assessment is more effective than traditional testing and prefer classroom-based assessment because of "the limitations of the current formal procedures" (Troudi, Coombe and Al-Hamly, 2009).

Furthermore, research into the learners' views shows that students believe that alternative assessment is fair, stimulating, creative and engaging but time-consuming (Struyven, Dochy

and Janssens, 2002). Additionally, Gijbels and Dochy (2006) claim that students need consistency between learning and assessment tasks.

Research into students' and teachers' views (Tarnanen and Huhta, 2011) reveals that they both believe that self- and peer-assessment do not really determine students' grades. On the contrary, they believe that effort, participation and attitude influence grades. Furthermore, teachers report that it is the ability to communicate that determines students' grades; yet, students claim that grammar is the most important determiner of assessment (ibid).

Eventually, the literature review reveals that although teachers and students appear to be positively disposed towards alternative assessment, traditional assessment is believed to determine student grades. Concerns exist mainly about issues of practicality and reliability of alternative assessment.

Research Methodology

The aim of the present study is to compare teacher and student perceptions about English language assessment in the Greek Junior High School. In particular, the present study explored – from two different viewpoints, namely the teachers' and the students' – how integration of instruction and assessment is materialised in Greek EFL classrooms, the degree to which authentic assessment tasks are employed and the degree of student involvement in assessment procedures. The following research questions were addressed:
1. What is the purpose of assessment in the English language classrooms of Greek Junior High Schools?
2. What are the assessment methods and tasks employed in the Greek EFL classroom?
3. What are the procedures followed in the assessment of EFL students?

For the purposes of the study, two questionnaires were constructed (see Appendix I: Teacher and Student Questionnaires) based on Tsagari (2011) and Cheng, Rogers and Hu (2007). Yet, personal interests and experience, the review of relevant literature and the mandates of the Ministry to Education also guided the creation of the questionnaires. The questionnaires were divided into four sections namely, background information, purposes of testing and assessment, language skills and assessment techniques and, finally, assessment procedures.

Data Analysis

SPSS (Wiersma, 2000) was used to conduct statistical analysis to test whether student perceptions concerning assessment practices in the English language classroom of Greek Junior High School coincide with teacher perceptions. Before aggregating the samples of the two groups (teachers and students), factor analysis (Oppenheim, 1999) was applied to identify a small number of factors that could explain most of the variance in a much larger number of variables regarding "Purposes of assessment", "Aspects of learning" and "Types of assessment". To check the strength of the association between factors, Pearson's correlation coefficient (r) was used (Oppenheim, 1999; Wiersma, 2000).

Cronbach alpha was applied to present the reliability of the data (Oppenheim, 1999; Wiersma, 2000). For the comparison of teacher and student perceptions regarding purposes and types of assessment, t-tests and analysis of variance (ANOVA) (Oppenheim, 1999; Wiersma, 2000) were applied. To determine whether the observed frequencies according to

references of teachers differ significantly from the frequencies referred by students, chi-square (I) test was applied (Oppenheim, 1999; Wiersma, 2000). Chi-square test was used to check differences or similarities between teacher and student perceptions regarding assessment methods, language skills, assessment methods and assessment procedures.

Sample

A total of 80 English language teachers, recruited from different regions of Greece, participated in the study. The student questionnaire was distributed to 75 Junior High school students in Athens (capital of Greece). Tables 1, 2 and 3 provide background information of the participants.

	Teachers		Students	
Gender	n	%	n	%
Male	8	10	34	45,3
Female	72	90	41	54,7
	80	100	75	100

Table 1. Gender of participants

Age	n	%	Age	n	%
22-30	20	25	13	15	20
31-40	47	58.8	14	37	49.3
41-50	10	12.5	15	22	29.3
51+	3	3.8	16	1	1.3
	80	100		75	100

Table 2. Teachers according to age   *Table 3. Students according to age*

Purposes and Types of Assessment

During your academic and professional career, how have you been informed about assessment practices?	f	%
1. I attended a course/module during my undergraduate studies.	31	38.8
2. I attended a course/module during my postgraduate studies.	46	57.5
3. I attended pre-service seminars on assessment.	25	31.3
4. I have attended in-service seminars on assessment.	28	35
5. I have studied recent theories on assessment.	48	58.8
6. I have studied the guidelines in CEF.	52	65
7. I have studied the guidelines in DEPPS.	51	63.8
8. I discuss with colleagues	49	61.3
9. I rely on my teaching experience.	43	53.8
10. I have studied the suggestions in the teacher's book.	38	47.5
11. Other	1	1.3

Table 4. Teachers' sources of information about assessment

Purposes of student assessment (Question B1): Factor analysis was applied to identify a small number of factors which could explain most of the variance in a much larger number of variables. This process was considered important as it would facilitate the comparison of teacher and student answers for this part.

According to factor analysis, there are four dimensions of assessment purposes, which can be categorized under the following headings: (F1) formative assessment, (F2) motivation, (F3) summative assessment and (F4) administrative assessment. The question *"To place my students at appropriate levels"* is considered as a distinct dimension, which is not related to other purposes of assessment (Table 5).

	Questions	*F1*	*F2*	*F3*	*F4*
	14. To determine whether remedial instruction is necessary.	.803			
	13. To identify learning difficulties.	.744			
	5. To plan further instruction.	.744			
F1	15. To encourage my students to develop problem-solving skills.	.701			
	16. To encourage my students to develop a sense of responsibility.	.608			
	11. To identify my students' strengths and weaknesses.	.546			
	7. To make my students work harder.		.815		
	17. To enhance my students' self-confidence and self-esteem.		.694		
F2	6. To motivate my students.		.694		
	12. To maintain discipline in class.		.603		
	2. To measure my students' progress.			.772	
	4. To evaluate whether teaching objectives have been achieved.			.663	
F3	10. To provide students with information about their progress.			.624	
	8. To determine student term and final grades.				.834
F4	9. To provide information to the central administration and parents.				.691
	3. To assess my students' performance.				,466

Table 5. Factor analysis of "Purposes of student assessment"

According to descriptive statistics, teachers basically use assessment for summative, formative and administrative purposes. Although teachers and students consider the same factors as important, teachers indicate more than students that formative, administrative and summative assessment are the main purposes of assessment (Table 6).

Aspects of learning (Question B2): Factor analysis indicated two dimensions of aspects of learning: (F1) attitudes towards learning and (F2) performance. The first dimension refers to the extent to which cooperation, initiative, creativity, participation, effort and interest are considered as important aspects of assessment, while the second dimension measures the importance teachers attribute to performance (Table 7).

Assessment purposes	Teachers		Students		t-value	p-value
	M	S.D.	M	S.D.		
F1 – Formative assessment	3.86	0.74	3.18	0.77	5.11	<0.001
F2 - Motivation	3.05	0.83	2.88	0.88	1.25	>0.05
F3 - Administrative assessment	4.38	0.52	3.26	0.70	11.16	<0.001
F4 - Summative assessment	3.85	0.74	3.43	0.92	3.32	<0.01

Table 6. Mean and standard deviation of "assessment purposes"

	Questions	F1	F2
F1	9. Cooperation	.803	
	7. Initiative	.744	
	8. Creativity	.744	
	6. Interest	.701	
	4. Participation	.608	
	2. Effort	.546	
F2	5. Performance in tests		.815
	1. Performance of language skills.		.694
	3. Preparation		.694
	10. Respect to school regulations.		.384

Table 7. Factor analysis of "Aspects of learning"

Table 8 shows that teachers believe more than students that attitudes toward learning are important aspects of learning that contribute to assessment. By contrast, students refer more than teachers to performance, which is thought to be of importance to grading than attitudes towards learning.

Learning aspects	Teachers		Students		t-value	p-value
	M	S.D.	M	S.D.		
F1 – Attitudes toward learning	4.21	0.59	3.94	0.67	2.43	<0.05
F2 – Performance	3.72	0.59	3.96	0.61	-2.36	<0.05

Table 8. Mean and standard deviation of "learning aspects"

Types of assessment (Question B3): Factor analysis yielded five dimensions indicating different types of assessment: (F1) informal assessment, (F2) student portfolio and student diary, (F3) assessment of progress, (F4) teacher-oriented assessment and teacher observation and (F5) initial assessment (Table 9).

As Table 10 shows, teachers and students agree that progress and teacher-oriented assessment are the most usual types of assessment. Additionally, although teachers refer to informal assessment, students state that it is rarely used. Finally, both teachers and students share the view that student diary is never used.

Importance of assessment types for grades (Question B4): Teachers believe that progress assessment is very important while they refer to informal and teacher-oriented assessment as important. Finally, they consider student portfolio/diary and initial important as aspects

of little importance. On the other hand, students believe that progress and teacher-oriented assessment are important for their grades whereas informal assessment and student portfolio/diary are not very important (Table 11).

	Questions	F1	F2	F3	F4	F5
F1	10. peer-assessment	.853				
	9. self-assessment	.818				
	1. mini quizzes	.576				
F2	8. student portfolio		.905			
	6. student diary/journal		.673			
F3	2. progress tests			.684		
	11. projects			-.681		
F4	3. achievement tests				-.792	
	7. teacher observation				.675	
F5	5. placement tests					.812
	4. diagnostic tests					.746

Table 9. Factor analysis of "Types of assessment"

Assessment types	Teachers		Students		t-value	p-value
	M	S.D.	M	S.D.		
F1 – Informal assessment	3.11	0.88	2.08	0.73	7.5	<0.001
F2 – Student portfolio/diary	2.27	0.97	1.71	0.74	4.21	<0.001
F3 – Progress assessment	3.39	0.80	3.23	0.71	1.37	>0.05
F4 – Teacher-oriented assessment	2.71	0.84	1.99	0.67	5.64	<0.001
F5 – Initial assessment	3.35	0.94	2.53	0.76	5.70	<0.001

Table 10. Mean and standard deviation of use of assessment types

Assessment types	Teachers		Students		t-value	p-value
	M	S.D.	M	S.D.		
F1 – Informal assessment	2.98	0.88	2.30	0.69	5.02	<0.001
F2 – Student portfolio/diary	2.56	1.04	1.78	0.88	5.11	<0.001
F3 – Progress assessment	3.93	0.68	3.30	0.65	5.42	<0.001
F4 – Teacher-oriented assessment	2.09	1.00	2.23	0.96	-0.79	>0.05
F5 – Initial assessment	3.24	0.90	2.56	0.84	4.82	<0.001

Table 11. Mean and standard deviation of importance of assessment types for grades

Language Skills and Assessment Methods

Language Skills (Question C1): Table 12 indicates that almost all teachers assess vocabulary, writing, grammar and reading. A large percentage of teachers assess speaking (80%) and listening (71.3%). However, most students believe that teachers assess mainly writing, grammar and speaking.

Language skills/aspects	Teachers		Students		x^2	p-value
	f	%	f	%		
1. Reading	78	97.5	67	89.3	4.28	<0.05
2. Writing	79	98.9	71	94.7	2.07	>0.05
3. Listening	57	71.3	53	70.7	0.06	>0.05
4. Speaking	64	80	69	92	4.58	<0.05
5. Grammar	79	98.8	70	93.3	3.05	>0.05
6. Vocabulary	80	100	67	89.3	9	<0.01

Table 12. Assessment of language skills/aspects

• *Reading (Question C2):* According to Table 13, teachers believe that the most common reading tasks are multiple-choice questions (63 teachers), open-ended questions (59 teachers), information transfer (56 teachers) and multiple-matching activities (45 teachers). Students believe that the most usual reading tasks are open-ended questions (67 students), cloze texts (62 students), projects (61 students) and multiple-matching activities (59 students). However, only 6 teachers and 2 students state that student diary/journal is used to assess reading. It is worth mentioning that although information-transfer activities are mentioned by 56 teachers, only 15 students admit that such activities are used.

Tasks	Teachers		Students	
	f	%	f	%
1. Open-ended questions	59	73.8	67	89.3
2. Dual-choice questions	25	31.3	16	61.3
3. Multiple-choice questions	63	78.8	52	69.3
4. Information-transfer questions	56	70	15	26
5. Multiple-matching activities	45	56.3	59	78.7
6. Cloze	26	32.5	62	82.7
7. Jumbled sentences/paragraphs	43	53.8	36	48
8. Note-taking	21	26.3	36	48
9. Translation	16	20	40	53.3
10. Summary	24	30	17	22.7
11. Student diary/journal	6	7.5	2	2.7
12. Student portfolio	15	18.8	19	25.3
13. Projects	27	33.8	51	68

Table 13. Reading

• *Writing (Question C3):* Table 14 shows that both groups state that paragraph writing (66 teachers/57 students), projects (44 teachers/57 students) and controlled writing (56 teachers/44 students) are the most common writing tasks. Teachers include guided writing (64 teachers) as well, whereas 42 students claim that sentence joining is used. Teachers claim more than students that self-assessment (28 teachers/16 students) and peer-assessment (25 teachers/5 students) are used to assess writing. Student diary/journal is scarcely used.

• *Listening (Question C4):* Most teachers indicate that the activities used to assess listening are: true/false activities (51 teachers), multiple-choice questions (45 teachers) and matching activities (41 teachers). Generally, as Table 15 shows, there are differences in teacher and student answers as to the type of listening tasks used.

Tasks	Teachers		Students	
	f	%	f	%
1. Sentence joining	34	42.5	42	56
2. Paragraph writing	66	82.5	57	76
3. Free writing tasks	37	46.3	32	43
4. Controlled writing tasks	56	70	44	58.7
5. Guided writing tasks	64	80	36	48
6. Summary writing	22	27.5	16	21.3
7. Editing a sentence/a paragraph	26	32.5	34	45.3
8. Mediation activities	24	30	27	36
9. Student diary/journal	5	62.5	1	1.3
10. Student portfolio	25	31.3	27	36
11. Projects	44	55	57	76
12. Self-assessment	28	35	16	21.3
13. Peer-assessment	25	31.3	5	6.7

Table 14. Writing

Tasks	Teachers		Students	
	f	%	f	%
14. Open-ended questions	25	31.3	42	56
15. Dual-choice questions	15	18.8	39	52
16. True/false questions	51	63.8	58	77.3
17. Multiple-choice questions	45	56.3	48	64
18. Listening cloze	30	37.5	51	68
19. Matching activities	41	51.3	35	46.7
20. Jumbled pictures	25	31.3	42	56
21. Following directions	27	33.8	21	28
22. Summarising/note-taking	19	23.8	6	8
23. Dictation	23	28.8	36	48
24. Student portfolio	3	3.8	20	26.7

Table 15. Listening

• *Speaking (Question C5):* According to Table 12, teachers and students have different views concerning the types of tasks used to assess speaking. Nonetheless, both groups believe that a variety of tasks are used.

Tasks	Teachers		Students	
	f	%	f	%
1. Reading aloud	23	28.8	34	45.3
2. Responding to a cue	35	43.8	45	60
3. Translation	5	62.5	35	46.7
4. Guided conversation	54	67.5	38	50.7
5. Mediation	21	26.3	12	16
6. Problem solving	49	61.3	28	37.3
7. Commenting on a text/picture	51	63.8	57	76
8. Group discussion	50	62.5	49	65.3
9. Roleplay	52	65	27	36
10. Presentation	27	33.8	55	73.3
11. Interview	21	26.3	18	24
12. Discussion/interview based on picture stimulus	41	51.3	40	53.3
13. Summarising	14	17.5	8	10.7

Table 16. Speaking

• *Grammar (Question C6):* According to Table 17, most teachers (79) and students (63) state that gap-filling activities are used for the assessment of grammar. Multiple-choice activities are also widely used.

Tasks	Teachers		Students	
	f	%	f	%
1. Multiple choice	66	82.5	58	77.3
2. True/false	38	47.5	53	70.7
3. Cloze	46	57.5	55	73.3
4. Transformations	57	71.3	52	69.3
5. Scrambled activities	28	35	45	60
6. Matching activities	56	70	61	81.3
7. Gap-filling activities	79	98.8	63	84
8. Short-answer questions	42	52.5	36	48
9. Essay writing	25	31.3	28	37.3
10. Summary	8	10	11	14.7
11. Roleplay	7	8.8	6	8
12. Other	1	1.3	18	24

Table 17. Grammar

• *Vocabulary (Question C7):* As Table 18 shows, most teachers (74) claim that they use gap-filling activities to assess vocabulary. Multiple-choice activities are also mentioned by 69 teachers. Teachers also state that they greatly use word-building (64 teachers), synonyms/antonyms (63 teachers) and cloze (51 teachers). Students also refer to these activities but to a lesser degree.

Tasks	Teachers		Students	
	f	%	f	%
1. Multiple choice	69	86.3	62	82.7
2. Cloze	51	63.8	56	74.7
3. Gap-filling activities	74	92.5	56	74.7
4. Sentence-writing tasks	46	57.5	52	69.3
5. Translation	21	26.3	37	49.3
6. Summary writing	12	15	16	21.3
7. Synonyms/antonyms	63	78.8	38	50.7
8. Word building	64	80	39	52

Table 18. Vocabulary

Assessment Procedures

• *Test design (Question D1):* Table 19 shows that the great majority of teachers design the tests they use by themselves, using the teacher's book or other material. Additionally, 37.5 % use material from the Internet and international tests. Only a low percentage (23.8%) cooperates with colleagues in the design of tests and only 5% involve students in the process of test preparation. At the same time, Table 20 shows that about 65% of students believe that their teachers themselves prepare their tests, 61.3% refer that they do not participate in the preparation of the tests and almost all students support that their teachers do not use international tests.

Do you use tests and/or other assessment methods which are …?	
	%
… designed by you?	93.8
… designed by you together with a colleague?	23.8
… designed by you together with your students?	5
… ready-made from the teacher's book?	65
… from other published material?	70
… from existing international tests?	37.5
… found on the internet?	43.8

Table 19. Percentage of teacher practices regarding test design

Test design	Yes		No		I don't know	
	f	%	f	%	f	%
1. designed by your teacher	49	65.3	7	9.3	19	25.4
2. designed by your teacher with other teachers	7	9.3	19	25.3	49	65.4
3. designed by your teacher and you	5	7.6	46	61.3	24	32.1
4. ready-made from teacher's book	15	20	11	14.7	49	65.3
5. from other books	15	20	21	28	39	52
6. from international tests	1	1.3	71	94.7	3	4
7. found on the Internet	28	37.3	21	28	26	34.7

Table 20. Teacher practices regarding test design according to students

• *Test activities (Questions D2-D3):* The two groups share the same opinion: 53.8% of the teachers and 42.7% of the students agree that the test activities are similar to the textbook activities (Table 21). There are significant differences between teachers and students regarding similarity of test activities to the activities done in the classroom. More teachers than students claim that test activities are similar to the work done in class (Table 22).

Are the activities in your tests similar to the activities of the textbook?		
Answers	Teachers	Students
	%	%
Yes	53.8	42.7
No	2.8	2.7
Sometimes	42.5	54.7

Table 21. Similarity between test activities and textbook activities (%)

Are the activities in your tests similar to the activities done in class?		
Answers	Teachers	Students
	%	%
Yes	88.8	50.7
No	2.8	2.7
Sometimes	11.2	49.3

Table 22. Similarity between test activities and activities done in class (%)

• *Informing students about test content (Questions D4-D5):* According to Table 23, teachers state that they inform their students about the content of the test, at least sometimes. Yet, 12% of the students claim that they have no information about test content. The majority of teachers and students state that revision is the most frequent method of

informing students about test content. Only few teachers and students claim that they jointly decide about the content of the test (Table 24).

Are the students informed about the content of the test?		
Answers	Teachers	Students
	%	%
Yes	85	70.7
No	0	12
Sometimes	15	17.3

Table 23. Student information about test content (%))

Methods of student information about test content	Teachers		Students	
	f	%	f	%
1. Revision before the test	71	88.8	77	93.3
2. Oral instructions before the test	54	67.5	53	70.7
3. Emphasis on important points during teaching	57	71.3	46	61.3
4. Teacher-student joint decision	9	11.3	15	20

Table 24. Methods of student information about test content

• *Informing students about test format (Questions D6-D7):* According to Table 25, most teachers and students indicate that students are informed about test format. Additionally, Table 26 shows that teachers and students believe that oral instructions and revision are the most common ways of informing students about test format. Information about the official format and experience from previous tests are also referred as applicable methods of information. Only one teacher and two students claim that they jointly decide about test format.

Are the students informed about the format of the test?		
Answers	Teachers	Students
	%	%
Yes	75	76
No	1.3	2.7
Sometimes	23.8	21.3

Table 25. Student information about test format (%))

Methods	Teachers		Students	
	f	%	f	%
1. Official format	50	62.5	46	61.3
2. Oral instructions	62	77.5	64	85.3
3. Revision	68	85	62	82.7
4. Teacher-student joint decision	1	1.3	2	2.7
5. Experience from previous tests	37	46.3	30	40

Table 26. Methods of student information about test format

• *Informing students about grading criteria (Questions D8-D9):* Most teachers and students claim that students are informed about the grading criteria (Table 27). Moreover, Table 28 shows that the majority of teachers (85%) and students (85.3%) state that indication on the test is the most common method regarding student information about the grading criteria which will be applied during the correction of the test.

Are the students informed about the the grading criteria?		
Answers	Teachers %	Students %
Yes	73.8	62.7
No	2.5	8
Sometimes	23.8	29.3

Table 27. Student information about the grading criteria (%))

Methods of informing about grading criteria	Teachers		Students	
	f	%	f	%
1. Indication on test	68	85	64	85.3
2. Revision	25	31.3	47	62.5
3. Experience from previous tests	23	28.8	19	25.3

Table 28. Methods of student information about grading criteria

• *Feedback (Question D10):* Table 29 shows that students believe less than teachers that verbal feedback, written comments, checklists, total test scores and letter grades are used. Furthermore, they claim more than teachers that conferences with individual students are used. Both teachers and students believe that class conference is "often" used.

Feedback to students	Teachers		Students		F-value	p-value
	M	S.D.	M	S.D.		
1. Verbal feedback	4.11	1.03	3.25	1.28	21.21	<0.001
2. Written comments	3.96	1.11	2.20	1.37	77.67	<0.001
3. Checklist	2.02	1.10	1.52	1.03	8.66	<0.01
4. Conference with student	2.92	1.14	3.33	1.40	3.98	<0.05
5. Class conference	3.81	1.14	3.70	1.34	0.28	>0.05
6. Total test score	4.31	1.16	3.49	1.59	13.52	<0.001
7. A letter grade	1.84	1.28	1.29	0.90	9.31	<0.01

Table 29. Feedback types

• *Progress test and achievement test (Questions D11, D12, D13 & D14):* Progress and achievement tests involve the assessment of grammar, reading, vocabulary and writing; yet, there are differences in the frequency of their use. Listening and speaking are hardly ever tested according to students. More teachers than students claim that listening and vocabulary are parts of the progress test. Almost all teachers (98.8%) and students (96.6%) mention grammar, 92.5% of teachers and 79.7% of students mention reading, while 70% of teachers and 55.9% of students mention writing as parts of the achievement test (Tables 30 & 31).

• *Views regarding student participation in the design of the test (Question D15):* More than 40% of teachers and students agree that student participation in the design of the test would be useful. However, most teachers are not sure about the participation of students in the design of the test (Figure 17).

• Views regarding self- and peer-assessment (Questions D16-D17): Teachers appear to be positively inclined towards self- and peer- assessment. Most students favour self-assessment; however, an important percentage of students state that they are uncertain

about it. As far as peer-assessment is concerned, most students do not favour it (Tables 33 & 34).

Parts of a progress test	*Teachers*		*Students*		x^2	*p-value*
	f	%	f	%		
1. Reading	68	85	51	81	0.52	>0.05
2. Writing	48	60	28	44.4	3.43	>0.05
3. Listening	10	12.5	0	0	8.47	<0.01
4. Speaking	2	2.5	2	3.2	0.81	>0.05
5. Grammar	78	97.5	61	96.8	0.81	>0.05
6. Vocabulary	67	83.8	26	41.3	27.97	<0.001
7. Dictation	5	6.3	1	1.6	1.91	>0.05

Table 30. Parts of a progress test

Parts of an achievement test	*Teachers*		*Students*		x^2	*p-value*
	f	%	f	%		
1. Reading	74	92.5	47	79.7	4.97	<0.05
2. Writing	56	70	33	55.9	2.92	>0.05
3. Listening	13	16.3	1	1.7	7.94	<0.01
4. Speaking	3	3.8	1	1.7	0.47	>0.05
5. Grammar	79	98.8	57	96.6	0.74	>0.05
6. Vocabulary	66	82.5	27	45.8	20.70	<0.001
7. Dictation	8	10	21	35.6	13.47	<0.001

Table 31. Parts of the achievement test

Do you think that student participation in the design of the tests would be useful?		
Answers	Teachers %	Students %
Yes	42.5	44
No	13.8	37.3
I am not sure	43.8	18.7

Table 32. Student participation in the design of the tests (%)

Do you think that students should participate in their own assessment?		
Answers	Teachers %	Students %
Yes	80	48
No	6.3	18.7
I am not sure	13.8	33.3

Table 33. Self-assessment

Do you think that students should participate in the assessment of their peers?		
Answers	Teachers %	Students %
Yes	62.5	17.3
No	11.3	60
I am not sure	26.3	22.7

Table 34. Peer-assessment
according to teachers and students (%)

Discussion

The present study attempted to explore Greek Junior High school teacher and student views concerning English language assessment. In particular, the study focused on:
- the purpose, forms and frequency of assessment;
- the roles of teachers and students in the assessment process;
- their attitudes towards different modes of assessment and the type of feedback provided to students.

According to the findings, grading has a central role in Greek EFL classes and is most frequently informed by testing. Formative, summative and administrative assessment are really important for both teachers and students; yet, both groups agree that student motivation (Bachman, 1990) is also an important purpose of assessment. Although teachers claim they assign more importance to student attitudes towards learning (Rea-Dickins, 2000), students believe that performance is more significant in their classes. Additionally, teachers use performance assessment mainly in the form of projects. However, there is disagreement between teachers and students as to whether and to what extent self- and peer-assessment are taken into consideration for grading.

Placement is a distinct purpose which does not inform instructional decisions. However, the assessment carried out at the early stages of instruction can productively guide lesson planning and assessment. Teachers can use this information to develop instructional and assessment material that caters for student needs and characteristics.

Teachers mainly use paper-and-pencil language tests (see also McNamara, 2000) to assess separate components of language knowledge, such as grammar and vocabulary, in line with previous research on the subject (see Tsagari, 2008). They also use tests to assess language skills, mainly reading and writing. Nevertheless, progress and achievement tests do not focus on all skills (also in Tsagari & Pavlou, 2009), e.g. they do not include the assessment of listening and speaking skills (Rea-Dickins, 2007). This, of course is in line with the official specifications which define the content of the achievement test. The official form of the final achievement test constitutes a barrier which restricts teachers to employing tasks which cannot be perceived as authentic and meaningful. Yet, the use of authentic tasks in language assessment seems to gain ground.

Furthermore, various methods of giving feedback to students are employed. Feedback is not restricted in a managerial aspect of assigning grades. Though teachers and students do not completely agree, it is obvious that there is an open channel of information between the two groups. This interaction between teachers and students is a factor which promotes positive washback (Bachman & Palmer, 1996), ensures the formative role of assessment (Bachman, 1990) and determines the classroom atmosphere (Clark, 2006).

Teachers ensure student familiarity with test content by emphasizing important points during teaching and reviewing material prior to tests. Familiarity with question formats is also ensured (Airasian, 2005). Students are also informed about the grading criteria, providing assessment with increased transparency (Cheng & Wang, 2007), which constitutes an effort for reliability. Yet, a variety of multiple sources of assessment is also necessary to provide assessment with reliability (Taylor & Nolen, 1996). Thus, assessment procedures in Greek Junior High Schools present a certain degree of content validity, since the content and the format of the test correspond to material known and reflect classroom objectives (Airasian, 2005), as these are represented in the classroom and the textbook activities.

Furthermore, this fact supports the face validity of the test. It can also be advocated that for this reason, test and assessment have positive washback on teaching and learning: the assessment tasks are related to the objectives of the course (Weir, 1988).

Teachers are found to be more positively inclined towards self- and peer-assessment than students (Noonan & Duncan, 2005). It is worth investigating, however, why students do not think self- and peer-assessment would be useful. Falchikov (1996) argues that students feel uneasy during peer-assessment. Nunan (1988) suggests that students have to be assisted to use such techniques. There is, therefore, a need to sensitise learners to become autonomous, to involve in the assessment of their performance and their progress, the materials and the activities used in teaching and testing. Yet, to achieve this, learners must be informed about course objectives and most aspects of the curriculum (ibid). The teacher can undertake the responsibility and become the agent of this development (Nunan, 1988; Tudor, 1996) through student training (McDonough & Shaw, 2003; Ross, 2006).

Teachers do not collaborate for the design and implementation of their assessment (Tsagari & Pavlou, 2009). Teachers develop their own tests (Sanders & Horn, 1995), consulting different sources (Coniam, 2009) but not their colleagues or their students. This could be attributed to the fact that a number of teachers have not had any – or they had little – formal training in assessment, in agreement with results from previous studies (Al-Saadat, 2004; Cumming, 2009; Taylor 2009; Xu & Liu, 2009). Moreover, teachers of the English language in the Greek Junior High School have to base their assessment practices on sources that are at hand: studying CEFR and DEPPS, discussing with colleagues and studying assessment theories; yet, all of these do not constitute a uniform and consistent source of information, as they depend on personal choices and interpretation.

In conclusion, it can be claimed that teachers conform to official documents in order to carry out assessment and they realise the importance of alternative assessment and the necessity to involve learners. Furthermore, students understand the purpose of assessment and the importance of different aspects of learning and they have a clear picture of assessment methods and tasks. However, they are ambivalent as to their involvement in assessment.

Pedagogical Implications

These results highlight that the specifications for the design of tests result in inconsistencies between class work and assessment procedures. First of all, it is obvious that adaptation is necessary to achieve alignment with DEPPS. However, the findings also reveal the need for teacher support and accessible professional development (Brindley, 1992; Clark, 2006; Hamp-Lyons, 2009) which will include technical and practical knowledge along with theoretical knowledge (Taylor, 2009).

A necessary step to be taken is the introduction of relevant undergraduate and graduate courses, where they do not exist (Cumming, 2009; Tsagari & Pavlou, 2009), followed by continuing support in the form of in-service training (Coniam, 2009; Cumming, 2009; Stiggins & Conklin, 1992) and the organisation of teacher workshops (Richards, 2001).

White (1988) puts forward a normative-re-educative model of innovation on the grounds that people are self-activating and non-passive. This model advocates normative change, that is, changes in attitudes, values and skills, through alteration in teachers' personal theories of teaching. He bases his suggestion on the belief that the innovations decided and

implemented by teachers themselves will produce more effective and more lastingly established results than top-down innovations.

Perhaps the most significant pedagogical implication to be drawn is that students need to be involved in the whole process of learning, including the process of assessment. The most secure way to do this is classroom interaction (Clark, 2006). Yet, student training is necessary (Black & Wiliam, 1998; Sadler, 1989), for only when learners become aware of the purposes of learning and assessment, can they be actively involved in the process.

Suggestions for Future Research

This study of teacher and student perceptions about assessment in the English language classroom of Greek Junior High Schools attempted to investigate a much studied issue, yet, taking into consideration student voices as well. Future research is hoped to override the limitations and weaknesses of the present study. Specifically, a more extensive sample of teachers and students could enhance the results of the present study. Also, the limited focus on aspects like grading and teacher training could also be expanded. Further research into the students' point of view of assessment practices is also necessary (Rea-Dickins, 2007).

It is also hoped that further research on teacher training needs on assessment (Davison & Leung, 2009) will be supported, which will eventually lead to the enhancement of awareness and sensitivity of administration to teachers' claims for support and training (Brindley, 1997) and, ultimately, to the enhancement of teacher effectiveness. Ultimately, the area of assessment constitutes a significant context for investigation of special interest to policy makers as well.

Conclusion

The present study concludes that EFL teachers in Greek Junior High Schools conform to official documents and carry out assessment procedures as prescribed. At the same time, they appear to understand the importance of alternative assessment methods and the necessity to engage learners in the process of assessment. However, this is not materialised, probably due to lack of teacher training and support on assessment. Students, also, are aware of the assessment procedures used in their class. Students understand the purposes of assessment, the importance of different aspects of learning, like motivation and effort, and they have a clear picture of methods and tasks used for their assessment, which indicates a certain degree of interaction with teachers. Nonetheless, learners appear to be ambivalent as to their involvement in assessment, which probably reflects uncertainty. A need for student training is obvious here, too, which will eventually lead to student empowerment.

Author's email: svlanti@gmail.com

References

Airasian, P. W. (2005). *Classroom assessment: concepts and applications.* Boston: McGraw-Hill.
Al-Saadat, A. I. (2004). 'A close look at Saudi EFL teachers preparation in language testing.' *Scientific Journal of King Faisal University*, 5/2: 227-254.

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Bachman, L. F. & Palmer, A. S. (eds). (1996). *Language testing in practice*. Oxford: Oxford University Press.

Black, P. & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment.* London: School of Education King's College.

Brindley, G. (1997). 'Assessment and the language teacher: trends and transitions.' *The Language Teacher Online*, accessed at http://jalt-publications.org/tlt/files/97/sep/brindley.html on 21 Aug 2010.

Chan, Y. (2008). 'Elementary school EFL teachers' beliefs and practices of multiple assessments.' *Reflections on English Language Teaching,* 7/1: 37-62.

Cheng, L., Rogers, T. & Hu, H. (2004). 'ESL/EFL instructors' classroom assessment practices: purposes, methods and procedures.' *Language Testing,* 21: 360.

Cheng, L. & Wang, X. (2007). 'Grading, feedback, and reporting in ESL/EFL classrooms.' *Language Assessment Quarterly*, 4/1: 85-107.

Clark, I. (2006). 'Assessment for learning: assessment in interaction.' *Essays in Education*, 18: 1-16.

Coniam, D. (2009). 'Investigating the quality of teacher-produced tests for EFL students and the effects of training in test development principles and practices on improving test quality.' *System*, 37: 226-242.

Davison, C. & Leung, C. (2009). 'Current issues in English language teacher-based assessment.' *TESOL Quarterly*, 43/3: 393-415.

Falchikov, N. (1996). 'Improving learning through critical peer feedback and reflection.' Different Approaches: Theory and Practice in Higher Education. *Proceedings HERDSA Conference 1996*. Perth, Western Australia.

Hamp-Lyons, L. (2009). 'Principles for large-scale classroom-based teacher assessment of English learners' language : an initial framework from school-based assessment in Hong Kong.' *TESOL Quarterly*, 43/3: 524-529.

McDonough, J. & Shaw, C. (2003). *Materials and methods in ELT*. Oxford: Blackwell Publishing.

McNamara, T. (2000). *Language testing.* Oxford: Oxford University Press.

Noonan, B. & C. R. Duncan. (2005). 'Peer and self-assessment in High Schools.' *Practical Assessment Research & Evaluation*, 10/17: 1-8.

Nunan, D. (1988). *The learner-centred curriculum.* Cambridge: Cambridge University Press.

Oppenheim, A. N. (1999). *Questionnaire design, interviewing and attitude measurement*. London: Pinter.

Pedagogical Institute (2003). 'Cross-Thematic Curriculum Framework for Modern Foreign Languages.' Accessed at http://www.pi-schools.gr/download/programs/depps/english/14th.pdf on 21 Aug 2010.

Rea-Dickins, P. (2000). 'Classroom assessment.' In T. Hedge (Ed.), *Teaching and learning in the language classroom.* Oxford: Blackwell Publishing, 375-401.

Rea-Dickins, P. (2007). 'Classroom-based assessment: possibilities and pitfalls.' In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching*, 505-520.

Richards, J. C. (2001). *Curriculum development in language teaching*. New York: Cambridge University Press.

Ross, J. A. (2006). 'The reliability, validity, and utility of self-assessment.' *Practical Assessment Research & Evaluation*, 11/10: 1-13.

Sadler, D. R. (1989). 'Formative assessment and the design of instructional systems.' *Instructional Science*, 18/2: 119-144.

Sanders, W. L. & Horn, S. P. (1995). 'Educational assessment reassessed: the usefulness of standardized and alternative measures of student achievement as indicators for the assessment of educational outcomes.' *Educational Policy Analysis Archives*, 3/6: 1-15.

Stiggins, R. J. & Conklin, N. F. (1992). *In teachers' hands: investigating the practices of classroom assessment.* New York: State University of New York.

Tarnanen, M. & A. Huhta. (2011). 'Foreign language assessment and feedback practices in Finland.' In D. Tsagari & I. Csépes (Eds.), *Classroom-based language assessment*. Frankfurt: Peter Lang, 129-146.

Taylor, C. S. & Nolen, S. B. (1996). 'What does the psychometrician's classroom look like? Reframing assessment concepts in the context of learning.' *Education Policy Analysis Archives*, 4/17: 1-46.

Taylor, L. (2009). 'Developing assessment literacy.' *Annual Review of Applied Linguistics*, 29: 21-36.

Troudi, S., Coombe, C. & Al-Hamly, M. (2009). 'EFL teachers' views of English language assessment in Higher Education in the United Arab Emirates and Kuwait.' *TESOL Quarterly*, 43/3: 546-555.

Tsagari, D. (2008). 'Assessment literacy of EFL teachers in Greece: current trends and future prospects.' [Powerpoint presentation] Accessed at http://www.ealta.eu.org/conference/2008/programme.htm on accessed 30 Oct 2010.

Tsagari, D. (2011). 'Practices, policies and tensions in classroom-based assessment.' In D. Tsagari & I. Csépes (Eds.), *Classroom-based language assessment*. Frankfurt: Peter Lang, 169-190.

Tsagari, D. & Pavlou, P. (2009). 'Practices, policies and tensions in classroom-based assessment.' [Powerpoint presentation] Accessed at http://www.ealta.eu.org/conference/2009/programme.htm on 30 Oct 2010.

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

Weir, C. J. (1988). *Communicative language testing*. New York: Prentice Hall.

Wiersma, W. (2000). *Research methods in education: an introduction*. Boston: Allyn and Bacon.

White, R. V. (1988). *The ELT curriculum: design, innovation and management*. Oxford: Blackwell Publishers Inc.

Xu, Y. & Liu, Y. (2009). 'Teacher assessment knowledge and practice : a narrative inquiry of a Chinese college EFL teacher's experience.' *TESOL Quarterly*, 43/3: 493-513.

Appendix I. The Questionnaires

A. Teacher Questionnaire[1]

Dear Colleague,

My name is Stavroula Vlanti and I am a postgraduate student at the Hellenic Open University working on my dissertation. I would like to kindly invite you to participate in my research project. The purpose of the study is to compare teacher and student perceptions concerning assessment practices in the English language classroom of Greek Junior High Schools. The survey will take approximately 15 minutes to complete and it is anonymous. If it is possible, please complete it by December 20th.

If you have any questions about this project, please contact me at the following mails: svlanti@sch.gr and svlanti@gmail.com .

Thank you very much for your help.

Stavroula Vlanti
Teacher of English
9th Gymnasium of Aigaleo

A. Background information

1. Age
☐ 22-30   ☐ 31-40   ☐ 41-50   ☐ 51+

2. Gender
☐ Male   ☐ Female

3. Years of teaching experience
☐ 0-5   ☐ 6-10   ☐ 11+

4. Years of teaching experience in a Junior High School
☐ 0-5   ☐ 6-10   ☐ 11+

5. Apart from your BA in English Language and Literature, do you hold any other degrees?
☐ Yes.   ☐ No.

6. If <u>yes</u>, please specify:
☐ MA in ELT Methodology
☐ Other. Please, state: _____

7. During your academic and professional career, how have you been informed about assessment practices? Please, tick all that apply.
☐ I attended a course/module on assessment during my undergraduate studies.
☐ I attended a course/module on assessment during my postgraduate studies.
☐ I attended an induction/pre-service seminar on assessment.
☐ I have attended in-service seminars/workshops on assessment.
☐ I have studied recent theories on assessment.
☐ I have studied the guidelines in the Common European Framework of Reference (CEF).
☐ I have studied the guidelines in the Cross-Thematic Curriculum (DEPPS) and the syllabus for the English language.
☐ I discuss with colleagues.
☐ I rely on my teaching experience.
☐ I have studied the suggestions concerning assessment in the teacher's book I use.
☐ Other. Please, explain: _____

---

[1] https://spreadsheets.google.com/viewform?formkey=dFZPYlhCSFFOVDJJZHA2RHZBQnlMWUE6MQ

B. Purposes of testing and assessment

1. Why do you assess students? Please, tick all that apply.

1: Not important - 5: Most important

	1	2	3	4	5
1. To place my students at appropriate levels.	☐	☐	☐	☐	☐
2. To measure my students' progress.	☐	☐	☐	☐	☐
3. To assess my students' performance.	☐	☐	☐	☐	☐
4. To evaluate whether teaching objectives have been achieved.	☐	☐	☐	☐	☐
5. To plan further instruction.	☐	☐	☐	☐	☐
6. To motivate my students.	☐	☐	☐	☐	☐
7. To make my students work harder.	☐	☐	☐	☐	☐
8. To determine student term and final grades.	☐	☐	☐	☐	☐
9. To provide information to the central administration (for example, school) and parents.	☐	☐	☐	☐	☐
10. To provide students with information about their progress.	☐	☐	☐	☐	☐
11. To identify my students' strengths and weaknesses.	☐	☐	☐	☐	☐
12. To maintain discipline in class.	☐	☐	☐	☐	☐
13. To identify learning difficulties.	☐	☐	☐	☐	☐
14. To determine whether remedial instruction is necessary.	☐	☐	☐	☐	☐
15. To encourage my students to develop problem-solving skills.	☐	☐	☐	☐	☐
16. To encourage my students to develop a sense of responsibility.	☐	☐	☐	☐	☐
17. To enhance my students' self-confidence and self-esteem.	☐	☐	☐	☐	☐

2. What aspects of learning are important when you assign grades, for example, at the end of the term? Please, tick all that apply.

1: Not important - 5: Most important

	1	2	3	4	5
1. Performance of language skills	☐	☐	☐	☐	☐
2. Effort	☐	☐	☐	☐	☐
3. Preparation (for example, homework)	☐	☐	☐	☐	☐
4. Participation	☐	☐	☐	☐	☐
5. Performance in tests	☐	☐	☐	☐	☐
6. Interest	☐	☐	☐	☐	☐
7. Initiative	☐	☐	☐	☐	☐
8. Creativity	☐	☐	☐	☐	☐
9. Cooperation	☐	☐	☐	☐	☐
10. Respect to school regulations	☐	☐	☐	☐	☐

3. What types of assessment do you use? Please, tick all that apply.

	Never	Rarely	Sometimes	Often	Most usually
1. Mini quizzes	☐	☐	☐	☐	☐
2. Progress tests (for example, at the end of the unit)	☐	☐	☐	☐	☐
3. Achievement tests (final tests)	☐	☐	☐	☐	☐
4. Diagnostic tests	☐	☐	☐	☐	☐
5. Placement tests	☐	☐	☐	☐	☐
6. Student diary/journal	☐	☐	☐	☐	☐
7. Teacher observation	☐	☐	☐	☐	☐
8. Student portfolio	☐	☐	☐	☐	☐
9. Self-assessment	☐	☐	☐	☐	☐
10. Peer-assessment	☐	☐	☐	☐	☐
11. Projects	☐	☐	☐	☐	☐

4. Which of these are most important when you assign grades (for example, at the end of the term)? Please, tick all that apply.

1: Not important - 5: Most important

		1	2	3	4	5
1.	Mini quizzes	☐	☐	☐	☐	☐
2.	Progress tests	☐	☐	☐	☐	☐
3.	Achievement tests	☐	☐	☐	☐	☐
4.	Diagnostic tests	☐	☐	☐	☐	☐
5.	Placement tests	☐	☐	☐	☐	☐
6.	Student diary/journal	☐	☐	☐	☐	☐
7.	Teacher observation	☐	☐	☐	☐	☐
8.	Student portfolio	☐	☐	☐	☐	☐
9.	Self-assessment	☐	☐	☐	☐	☐
10.	Peer-assessment	☐	☐	☐	☐	☐
11.	Projects	☐	☐	☐	☐	☐

C. Assessment methods

1. Do you assess your students on … ? Please, tick all that apply.
☐ Reading
☐ Writing
☐ Listening
☐ Speaking
☐ Grammar
☐ Vocabulary

2. If you assess your students on <u>reading</u>, what types of tasks do you use? Please, tick all that apply.
☐ Open-ended questions
☐ Dual-choice questions
☐ Multiple-choice questions
☐ Information-transfer questions
☐ Multiple-matching activities
☐ Cloze
☐ Jumbled sentences/paragraphs
☐ Note-taking
☐ Translation
☐ Summary
☐ Student diary/journal
☐ Student portfolio
☐ Projects
☐ Other. Please, state: _____

3. If you assess your students on <u>writing</u>, what types of tasks do you use? Please, tick all that apply.
☐ Sentence joining
☐ Paragraph writing
☐ Free writing tasks
☐ Controlled writing tasks
☐ Guided writing tasks
☐ Summary writing
☐ Editing a sentence/a paragraph
☐ Mediation activities
☐ Student diary/journal
☐ Student portfolio
☐ Projects
☐ Self-assessment
☐ Peer-assessment
☐ Other. Please, state: _____

4. If you assess your students on <u>listening</u>, what types of tasks do you use? Please, tick all that apply.
- ☐ Open-ended questions
- ☐ Dual-choice questions
- ☐ True/false questions
- ☐ Multiple-choice questions
- ☐ Listening cloze
- ☐ Matching activities
- ☐ Jumbled pictures
- ☐ Following directions
- ☐ Summarising/note-taking
- ☐ Dictation
- ☐ Student portfolio
- ☐ Other. Please, state: _____

5. If you assess your students on <u>speaking</u>, what types of tasks do you use? Please, tick all that apply.
- ☐ Reading aloud
- ☐ Responding to a cue
- ☐ Translation
- ☐ Guided conversation
- ☐ Mediation
- ☐ Problem solving
- ☐ Commenting on a text/picture
- ☐ Group discussion
- ☐ Roleplay
- ☐ Presentation
- ☐ Interview
- ☐ Discussion/interview based on picture stimulus
- ☐ Summarising
- ☐ Other. Please, state: _____

6. If you assess your students on <u>grammar</u>, what types of tasks do you use? Please, tick all that apply.
- ☐ Multiple-choice
- ☐ True/false
- ☐ Cloze
- ☐ Transformations
- ☐ Scrambled activities
- ☐ Matching activities
- ☐ Gap-filling activities
- ☐ Short-answer questions
- ☐ Essay writing
- ☐ Summary
- ☐ Roleplay
- ☐ Other. Please, state: _____

7. If you assess your students on <u>vocabulary</u>, what types of tasks do you use? Please, tick all that apply.
- ☐ Multiple-choice
- ☐ Cloze
- ☐ Gap-filling activities
- ☐ Sentence-writing tasks
- ☐ Translation
- ☐ Summary writing
- ☐ Synonyms/antonyms
- ☐ Word building
- ☐ Other. Please, state: _____

D. Assessment procedures

1. Do you use tests and/or other assessment methods which are …? Please, tick all that apply.
☐ … designed by you?
☐ … designed by you together with a colleague?
☐ … designed by you together with your students?
☐ … ready-made from the teacher's book?
☐ … from other published material?
☐ … from existing international tests?
☐ … found on the internet?
☐ Other. Please, state: _____

2. Are the activities in your tests similar to the activities of the textbook? Please, choose only one answer.
☐ Yes.   ☐ No.   ☐ Sometimes.

3. Are the activities in your tests similar to the activities done in class? Please, choose only one answer.
☐ Yes.   ☐ No.   ☐ Sometimes.

4. Are the students informed about the content of the test? Please, choose only one answer.
☐ Yes.   ☐ No.   ☐ Sometimes.

5. If the students are informed (even sometimes) about the content of the test, how is this done? Please, tick all that apply.
☐ We have a revision session before the test.
☐ I give oral instructions before the test.
☐ I emphasise the important points during teaching.
☐ We decide together about the content of the test.
☐ Other. Please, explain. _____
_____

6. Are the students informed about the format of the test? Please, choose only one answer.
☐ Yes.   ☐ No.   ☐ Sometimes.

7. If the students are informed (even sometimes) about the format of the test, how is this done? Please, tick all that apply.
☐ I inform students about the official format (defined by the Ministry) of the test.
☐ I give students necessary information before the test.
☐ We do similar activities during revision.
☐ We decide together about the format of the test.
☐ They have experience from previous tests.
☐ Other. Please, explain. _____
_____

8. Are the students informed about the grading criteria? Please, choose only one answer.
☐ Yes.   ☐ No.   ☐ Sometimes.

9. If the students are informed (even sometimes) about the grading criteria, how is this done? Please, tick all that apply.
☐ Every section of the test has appropriate indication.
☐ Students are informed about grading during revision.
☐ They have experience from previous tests.
☐ Other. Please, explain. _____
_____

10. How do you give feedback to your students after a test or other method of assessment? Please, tick all that apply.

	Never	Rarely	Sometimes	Often	Most usually
Verbal feedback (oral comments)	☐	☐	☐	☐	☐
Written comments	☐	☐	☐	☐	☐
Checklist	☐	☐	☐	☐	☐
Conference with individual student	☐	☐	☐	☐	☐
Class conference	☐	☐	☐	☐	☐
Total test score	☐	☐	☐	☐	☐
A letter grade	☐	☐	☐	☐	☐

11. When you design a progress test, how many parts does it usually include? Please, choose only one answer.
☐ One   ☐ Two   ☐ Three   ☐ More than three

12. Can you please name these parts (for example, reading, writing, grammar etc.)?
_____
_____

13. When you design an achievement test, how many parts does it usually include? Please, choose only one answer.
☐ One   ☐ Two   ☐ Three   ☐ More than three

14. Can you please name these parts (for example, reading, writing, grammar etc.)?
_____
_____

15. Do you think that student participation in the design of the tests would be useful? Please, choose only one answer.
☐ Yes.   ☐ No.   ☐ I am not sure.

16. Do you think that students should participate in their own assessment? Please, choose only one answer.
☐ Yes.   ☐ No.   ☐ I am not sure.

17. Do you think that students should participate in the assessment of their peers? Please, choose only one answer.
☐ Yes.   ☐ No.   ☐ I am not sure.

Thank you very much for your help!

B. Student Questionnaire[2]

Dear Student,

My name is Stavroula Vlanti and I am a postgraduate student at the Hellenic Open University working on my dissertation. I would like to kindly invite you to participate in my research project. The purpose of the study is to compare teacher and student perceptions concerning assessment practices in the English language classroom of Greek Junior High Schools. The survey will take approximately 15 minutes to complete and it is anonymous.

Thank you very much for your help.

Stavroula Vlanti
Teacher of English
9th Gymnasium of Aigaleo

A. Background information

1. Age
☐ 12   ☐ 13   ☐ 14   ☐ 15   ☐ 15+

2. Gender
☐ Boy   ☐ Girl

3. Grade in Junior High School
☐ B' class - Advanced   ☐ B' class - Beginner
☐ C' class - Advanced   ☐ C' class - Beginner

4. As a Junior High School student, how many English language teachers have you had up to now?
☐ 1   ☐ 2   ☐ 3   ☐ More than 3

B. Purposes of testing and assessment

1. Why does your teacher assess you? Please, tick all that are true.

	1: Not important	**-**	**5: Most important**		
	1	2	3	4	5
1. To place you at appropriate levels (Advanced/Beginners).	☐	☐	☐	☐	☐
2. To see whether teaching has been successful.	☐	☐	☐	☐	☐
3. To measure the progress you have made.	☐	☐	☐	☐	☐
4. To measure your ability to understand and use the English language.	☐	☐	☐	☐	☐
5. To plan further teaching.	☐	☐	☐	☐	☐
6. To motivate you.	☐	☐	☐	☐	☐
7. To make you work harder.	☐	☐	☐	☐	☐
8. To decide on term and final grades.	☐	☐	☐	☐	☐
9. To provide information to the central administration (for example, school) and parents.	☐	☐	☐	☐	☐
10. To provide you with information about your progress.	☐	☐	☐	☐	☐
11. To identify your strengths and weaknesses.	☐	☐	☐	☐	☐
12. To keep you disciplined in class.	☐	☐	☐	☐	☐
13. To identify learning difficulties.	☐	☐	☐	☐	☐
14. To decide whether a unit/a structure needs revision.	☐	☐	☐	☐	☐
15. To encourage you to develop the ability to deal with problems.	☐	☐	☐	☐	☐
16. To encourage you to be responsible.	☐	☐	☐	☐	☐
17. To make you feel self-confident.	☐	☐	☐	☐	☐

---

[2] https://spreadsheets.google.com/viewform?formkey=dEpIU2g1SjNKOUxQNERxdjcGhValE6MQ

2. What aspects of learning do you think are most important for your grades? Please, tick all that are true.

1: Not important - 5: Most important

		1	2	3	4	5
1.	Your ability to understand and use the English language	☐	☐	☐	☐	☐
2.	The effort you make	☐	☐	☐	☐	☐
3.	Your preparation (for example, homework)	☐	☐	☐	☐	☐
4.	Your participation in class	☐	☐	☐	☐	☐
5.	Your test results	☐	☐	☐	☐	☐
6.	The interest you show during the lesson	☐	☐	☐	☐	☐
7.	Your willingness to participate and work	☐	☐	☐	☐	☐
8.	Your creativity	☐	☐	☐	☐	☐
9.	Your cooperation	☐	☐	☐	☐	☐
10.	Your respect to school regulations	☐	☐	☐	☐	☐

3. What types of assessment does your teacher use? Please, tick all that are true.

		Never	Rarely	Sometimes	Often	Most usually
1.	Mini quizzes (for example, ten-minute tests)	☐	☐	☐	☐	☐
2.	Progress tests (for example, at the end of the unit)	☐	☐	☐	☐	☐
3.	Achievement tests (final tests)	☐	☐	☐	☐	☐
4.	Diagnostic tests	☐	☐	☐	☐	☐
5.	Placement tests	☐	☐	☐	☐	☐
6.	Student diary/journal	☐	☐	☐	☐	☐
7.	Teacher observation	☐	☐	☐	☐	☐
8.	Student portfolio	☐	☐	☐	☐	☐
9.	Self-assessment	☐	☐	☐	☐	☐
10.	Peer-assessment	☐	☐	☐	☐	☐
11.	Projects	☐	☐	☐	☐	☐

4. Which of the following do you think are most important when your teacher assigns grades (for example, at the end of the term)? Please, tick all that are true.

1: Not important - 5: Most important

		1	2	3	4	5
1.	Mini quizzes	☐	☐	☐	☐	☐
2.	Progress tests	☐	☐	☐	☐	☐
3.	Achievement tests	☐	☐	☐	☐	☐
4.	Diagnostic tests	☐	☐	☐	☐	☐
5.	Placement tests	☐	☐	☐	☐	☐
6.	Student diary/journal	☐	☐	☐	☐	☐
7.	Teacher observation	☐	☐	☐	☐	☐
8.	Student portfolio	☐	☐	☐	☐	☐
9.	Self-assessment	☐	☐	☐	☐	☐
10.	Peer-assessment	☐	☐	☐	☐	☐
11.	Projects	☐	☐	☐	☐	☐

C. Assessment methods

1. Are you assessed on … ? Please, tick all that are true.
- ☐ Reading
- ☐ Writing
- ☐ Listening
- ☐ Speaking
- ☐ Grammar
- ☐ Vocabulary

2. If you are assessed on <u>reading</u>, what types of tasks do you usually have to do? Please, tick all that are true.
- ☐ Open-ended questions
- ☐ Dual-choice questions
- ☐ Multiple-choice questions
- ☐ Information-transfer questions
- ☐ Multiple-matching activities
- ☐ Cloze
- ☐ Jumbled sentences/paragraphs
- ☐ Note-taking
- ☐Translation
- ☐ Summary
- ☐ Student diary/journal
- ☐ Student portfolio
- ☐Projects
- ☐ Other. Please, state: _____

3. If you are assessed on <u>writing</u>, what types of tasks do you usually have to do? Please, tick all that are true.
- ☐ Sentence joining
- ☐ Paragraph writing
- ☐ Free writing tasks
- ☐ Controlled writing tasks
- ☐ Guided writing tasks
- ☐ Summary writing
- ☐ Editing a sentence/a paragraph
- ☐ Mediation activities
- ☐ Student diary/journal
- ☐ Student portfolio
- ☐ Projects
- ☐ Self-assessment
- ☐ Peer-assessment
- ☐ Other. Please, state: _____

4. If you are assessed on <u>listening</u>, what types of tasks do you usually have to do? Please, tick all that are true.
- ☐ Open-ended questions
- ☐ Dual-choice questions
- ☐ True/false questions
- ☐ Multiple-choice questions
- ☐ Listening cloze
- ☐ Matching activities
- ☐ Jumbled pictures
- ☐ Following directions
- ☐ Summarising/note-taking
- ☐ Dictation
- ☐ Student portfolio
- ☐ Other. Please, state: _____

119

5. If you are assessed on <u>speaking</u>, what types of tasks do you usually have to do? Please, tick all that are true.
- ☐ Reading aloud
- ☐ Responding to a cue
- ☐ Translation
- ☐ Guided conversation
- ☐ Mediation
- ☐ Problem solving
- ☐ Commenting on a text/picture
- ☐ Group discussion
- ☐ Roleplay
- ☐ Presentation
- ☐ Interview
- ☐ Discussion/interview based on picture stimulus
- ☐ Summarising
- ☐ Other. Please, state: _____

6. If you are assessed on <u>grammar</u>, what types of tasks do you usually have to do? Please, tick all that are true.
- ☐ Multiple-choice
- ☐ True/false
- ☐ Cloze
- ☐ Transformations
- ☐ Scrambled activities
- ☐ Matching activities
- ☐ Gap-filling activities
- ☐ Short-answer questions
- ☐ Essay writing
- ☐ Summary
- ☐ Roleplay
- ☐ Other. Please, state: _____

7. If you are assessed on <u>vocabulary</u>, what types of tasks do you usually have to do? Please, tick all that are true.
- ☐ Multiple-choice
- ☐ Cloze
- ☐ Gap-filling activities
- ☐ Sentence-writing tasks
- ☐ Translation
- ☐ Summary writing
- ☐ Synonyms/antonyms
- ☐ Word building
- ☐ Other. Please, state: _____

D. Assessment procedures

1. Are the tests and/or other assessment methods used in your class …? Please, tick all that are true.

	Yes.	No.	I don't know.
… designed by your teacher?	☐	☐	☐
… designed by your teacher together with other teachers?	☐	☐	☐
… designed by your teacher and you?	☐	☐	☐
… ready-made from the teacher's book?	☐	☐	☐
…from other books?	☐	☐	☐
… from existing international tests (for example, KET, PET, FCE, ECCE, KPG)?	☐	☐	☐
… found on the internet?	☐	☐	☐

2. Are the activities in the tests similar to the activities of the textbook? Please, choose only one answer.
☐ Yes.   ☐ No.   ☐ Sometimes.

3. Are the activities in the tests similar to the activities done in class? Please, choose only one answer.
☐ Yes.   ☐ No.   ☐ Sometimes.

4. Do you know what the test will include? Please, choose only one answer.
☐ Yes.   ☐ No.   ☐ Sometimes.

5. If you know (even sometimes) what the test will include, how have you been informed? Please, tick all that are true.
☐ We have a revision lesson before the test.
☐ The teacher gives us oral instructions before the test.
☐ The teacher emphasises the important points during lessons.
☐ We decide together with the teacher about what to study.
☐ Other. Please, explain. _____
_____

6. Are you informed about the format of the test (for example, the number and the types of the tasks)? Please, choose only one answer.
☐ Yes.   ☐ No.   ☐ Sometimes.

7. If you are informed (even sometimes) about the format of the test, how is this done? Please, tick all that are true.
☐ The teacher informs us about the official format (defined by the Ministry) of the test.
☐ The teacher gives us necessary information before the test.
☐ We do similar activities during revision.
☐ We decide together with the teacher about the number and types of tasks.
☐ We have experience from previous tests.
☐ Other. Please, explain. _____
_____

8. Do you know how your teacher scores the tests (for example, how many points each task carries, what kinds of mistakes are considered important)? Please, choose only one answer.
☐ Yes.   ☐ No.   ☐ Sometimes.

9. If you know how your teacher scores the tests (even sometimes), how have you been informed? Please, tick all that are true.
☐ Every section of the test has indication about the marks it carries.
☐ The teacher informs us during revision.
☐ We have experience from previous tests.
☐ Other. Please, explain. _____

10. What kind of information (feedback) do you get about your performance in a test or other method of assessment? Please, tick all that are true.

	Never	Rarely	Sometimes	Often	Most usually
Oral comments	☐	☐	☐	☐	☐
Written comments	☐	☐	☐	☐	☐
Checklist	☐	☐	☐	☐	☐
Discussion with teacher	☐	☐	☐	☐	☐
Whole class discussion	☐	☐	☐	☐	☐
Total test score	☐	☐	☐	☐	☐
A letter grade	☐	☐	☐	☐	☐

11. Think of the most recent progress test you had. How many parts did it consist of? Please, choose only one answer.
☐ One   ☐ Two   ☐ Three   ☐ More than three

12. Can you please name the parts the progress test consisted of (for example, reading, grammar etc.)?
_____
_____

13. Think of the achievement test you had last June. How many parts did it consist of? Please, choose only one answer.
☐ One   ☐ Two   ☐ Three   ☐ More than three

14. Can you please name the parts the achievement test consisted of (for example, reading, grammar etc.)?
_____
_____

15. Do you think that student participation in the design of the tests would be useful? Please, choose only one answer.
☐ Yes.   ☐ No.   ☐ I am not sure.

16. Do you think that students should participate in their own assessment? Please, choose only one answer.
☐ Yes.   ☐ No.   ☐ I am not sure.

17. Do you think that students should participate in the assessment of their peers? Please, choose only one answer.
☐ Yes.   ☐ No.   ☐ I am not sure.

☺   **Thank you very much for your help!**

The Writing Portfolio:
an alternative assessment tool with young learners of English

[Ο Φάκελος Γραπτού Λόγου:
ένα εργαλείο εναλλακτικής αξιολόγησης μικρών μαθητών στην Αγγλική Γλώσσα]

Melpomeni Barabouti

This paper presents the implementation of a portfolio for assessment purposes, with a group of primary school learners of English, aged 11-12. The focus of the portfolio is specific, concentrating on a purposeful and systematic collection of samples of students' written language. The paper explores theories on portfolios as an alternative method of assessment and the benefits deriving from their use in the classroom. Moreover, it analyses all the stages of the organization of the learner portfolios from beginning to end. The findings shed light not only on positive outcomes but also possible problems. The paper discusses the implications of the study for classroom practice and provides suggestions for further experimentation.

☙

Η παρούσα εργασία παρουσιάζει την εφαρμογή ενός φακέλου εργασιών σε μια ομάδα μαθητών Δημοτικού Σχολείου, ηλικίας 11-12 που μαθαίνουν Αγγλικά. Ο φάκελος εστιάζει συγκεκριμένα στη σκόπιμη και συστηματική συλλογή γραπτών εργασιών των μαθητών. Συγκεκριμένα, η εργασία εξελίσσεται σε τρία στάδια: αρχικά, γίνεται αναφορά σε θεωρίες σχετικά με τη χρήση του φακέλου εργασιών ως εναλλακτική μέθοδο αξιολόγησης και τα οφέλη που προκύπτουν από αυτή. Κατόπιν παρουσιάζονται αναλυτικά όλα τα βήματα οργάνωσης του φακέλου που ακολουθήθηκαν από της αρχή ως το τέλος της έρευνας. Τέλος γίνεται αξιολόγηση των ευρημάτων της έρευνας, συζητούνται όχι μόνο τα θετικά αποτελέσματα αλλά και τα πιθανά προβλήματα και προτείνονται τομείς περαιτέρω έρευνας.

Key words: assessment, portfolio, alternative, writing, young learners

Introduction

In an attempt to meet the demands of a changing society, which values meta-cognitive skills and individual thought and expression, educators have suggested a number of alternative methods of assessment (i.e. portfolios, interactive diaries, student conferences, etc). These methods have been characterized by Genesse (1994) as classroom-referenced assessment because they are useful for making decisions not only about students, but also about the effectiveness of classroom instruction and about instructional planning

to promote language learning.

One method of alternative assessment is 'portfolios'. Zessoules and Gardner (1991), very successfully name them 'process-folios', as students can collect in them not only works of best quality or highest achievement, but also evidence of the process of learning, reflections on their own progress, their strengths, as well as weaknesses in understanding.

As far as the writing skill is concerned, which is of interest to the present study, portfolios can be an effective means of measuring it because they can provide a more complete and accurate portrait of students' writing growth on various occasions (Elbow & Belanoff, 1991; Jongsma, 1989). As Hamp-Lyons and Condon (2000:61) suggest, students have significant amounts of time to revise their work in portfolio assessment. They are able to find and correct their own errors, without being overwhelmed by time pressure. Particularly when teaching young learners, who are in a process of developing both cognitively and psychologically, the assessment of writing through portfolios can systematically record vital elements of the art of writing such as generating ideas, drafting and editing. It can also show the learners' improvement and encourage good writing habits.

Portfolio Assessment

Portfolios have been associated with the arts, where architects, photographers and other artists collect their pieces of work in order to display them to future employers or prospective buyers. Since the 90's, however, the use of portfolios has been expanded in various educational settings such as teacher education, assessing advanced learners and, more recently, young learners. Since the appearance of the portfolio on the educational assessment scene, various definitions have been proposed. The following two definitions depict more clearly what has been the use of portfolios in the present study.

Genesee & Upshur, (1996: 99) defined the portfolio as *'a purposeful collection of students' work that demonstrates to students and others their efforts, progress and achievements in given areas'.* Trim (1997:3) talked of the language portfolio as '*a document […] in which individual learners […] can assemble over a period of time, and display in a systematic way, a record of their qualifications, achievements and experiences in language learning, together with samples of work they have themselves produced*'.

The focus of language portfolios can be very specific, containing samples of students' written work, as is the case in the present study, or it can be broad, containing samples of oral work such as interviews and narrations through recordings, reports on reading work, art work or anything else that can provide evidence of language development and personal growth. However, the collection should be done selectively so as to avoid the danger of the portfolios becoming mere junk drawers of students' work, and more importantly, to simplify the teacher's task to assess work that draws on different skills.

On the contrary, portfolios as an alternative form of assessment are on-going or formative in nature and are intended to be primarily diagnostic. They allow learners plenty of time to generate rather than choose a response. According to Huerta-Macias (2002: 339), alternative forms of assessment, portfolios included, do not intrude on regular classroom activities. They reflect the curriculum that is actually being implemented, provide information on both the strengths and the weaknesses of each individual student, provide various sources of student development rather than one single method and, therefore, students' progress and growth can be assessed more reliably than in traditional testing.

A considerable number of advantages of portfolios are widely discussed in the EFL literature (Genesee & Upshur, 1996:100; Hamp-Lyons & Condon, 2000; Smith, 1996; Στεφανάκη, 1997; Φωτιάδου, 2001) and are briefly reported below. For example, portfolios enable students to:
• Enhance their self-image as they participate in the decisions about content

- Assume responsibility for self- assessment
- Interact with other students, teachers and parents about learning
- Use metalanguage to talk about language
- Exhibit their creativity and originality
- Think critically about school work

They also provide:
- An answer to today's need for a measurement system that can have a generative, rather than a reductive effect on education (Hamp-Lyons & Condon, 2000:xv)
- A chance to integrate teaching and assessment in a continuous process
- A rich source of information for teachers who can improve their own teaching methods and plan further instruction

Prompted by the theoretical considerations on the use of portfolios as an alternative form of assessment in the language classroom, this study aims at investigating the following research questions:
- How far can the purposeful and systematic collection of writing samples in a portfolio develop students' writing skills?
- Are there any negative aspects deriving from the use of portfolios in the classroom?
- Can the writing portfolios promote positive attitudes, such as increased motivation and development of learning-to-learn strategies?

The Study

Participants

The participants in this study were eleven male and nine female primary school EFL students, aged 10-11. These young learners attended the same Greek state school in Agia Paraskevi, a suburb in the north-east of Athens, Greece and had been studying English for one and a half years with the same teacher.

Although these particular students constituted in their majority a well-motivated class, willing to enthusiastically embrace any classroom innovation, they also had individual differences, concerning their interests, abilities and intelligences. The portfolio project managed to depict these personal characteristics and differences and helped the teacher appreciate each student's accomplishments according to their potential.

Procedure

The project began in February 2003 and extended over a period of three and a half months. The study followed several procedural steps, which, according to Pierce (1998:6), can help relate assessment activities directly to instructional activities and, therefore, make the collected information more manageable and useful both for the teacher and the students. For example:

Step1: Specify the purpose of the portfolio. The portfolios in the present study were basically used for the assessment of students' writing skills, with the emphasis on the process of writing rather than the final product.

Step 2: Specify use of assessment information. The information collected from the portfolios was used to:
- monitor students' growth in writing, focusing both on the content (organization of thoughts, variety of vocabulary, richness of ideas) and the form (correct grammar, syntax, style).
- diagnose students' strengths so as with positive comments to make them aware of what they can do.
- diagnose weaknesses, so as to adapt the teaching materials accordingly.

Step 3: Match entries to purpose. As far as written language production was concerned, the required entries in the portfolio were intended to match the instructional goals and objectives set by the curriculum and dictated by the syllabus and the course material (Funway 2 book, ΥΠΕΠΘ, 1999), used with the particular group of students. The working units from the book during the period of the study were units six to nine, where the language focus was the present simple, the present continuous, the simple past and the formation of wh- questions. The language functions practised in the relevant units were: talking and asking about routine and general facts, about things people were doing during present and past events respectively. As a result, the required entries in the portfolios aimed at depicting students' knowledge of the grammar forms and functions mentioned above, while at the same time gave them the opportunity to practise writing descriptive and narrative discourse types (Table 1).

No of entry	Activity title	Language function
1	Look at the pictures and write a paragraph about Nick's daily activities	Writing about people's routines
2	Write a postcard to a friend from a place you are now on holiday	Writing about present (every day or now)
3	Write in your diary how you spent your weekend	Describing a place and narrating past actions
4	Write about the best or worst experience, or the best or worst dream, or an adventure that you had	Describing/ Narrating past actions and events
5	Write a letter to a friend, accepting or rejecting his/her invitation	Expressing preference, ability-disability, likes and dislikes in order to justify a choice made or a decision taken
6	Grammar Check (taught verb tenses and wh-questions)	Finding the question when the answer is given and establishing time relations between verb tenses

Table 1: The required entries of the portfolio

Apart from the required entries a very important role in the portfolio was played by the optional entries. The students in the present study were asked to include in their portfolios any other type of writing samples, which either showed their improvement as writers, or provided a means to express themselves and share their feelings with the people who would eventually have access to their portfolios. The importance of the optional entries was also implied by Gearhart and Wolf (1995:13), who argue that if students are asked to practise specific genres (letters, diary entries etc) without writing for personal purpose, there is *'little chance for ownership of the writing'*.

Step 4: Record students' progress. An important phase in the process of portfolio development was assessing students' progress to gather information about their individual learning and help them take future action. In classroom-based portfolio assessment, as was the case in this study, the assessor of the students' writing samples was their teacher because she was familiar with the students. By being well acquainted with the instructional setting of the portfolios, the teacher found it easy to grasp the values reflected in students' work, adding thus, validity and reliability to the scoring process.

The different entries of the portfolio in the present study were assessed in different ways, using primary trait scoring and analytic rating scales (see Weigle, 2002, p. 112), correction codes or comments on first drafts, according to the aim each entry was intended to fulfil. A tabular representation of the different scoring methods used for the various entries is provided in Appendix I and examples of the scoring methods in Appendices II, III, IV.

Step 5: Get the students involved. The students were initially informed by the teacher about what the portfolio project was, and were asked to think whether they would be interested in it and committed to the whole idea. As was expected, their reaction was positive. To make students believe that the purpose of the portfolio was a joint decision between both the teacher and students, the teacher handed out a questionnaire (see Appendix V) aiming at surveying the students' writing interest and awareness. The language level of the students dictated the writing of the items is in Greek, so as to ensure understanding, and, therefore, validity of the conclusions drawn.

The students' involvement in the portfolio project also entailed personal selection and inclusion of optional entries with a view to either prove their improvement in writing, or share something with their teacher or classmates. The optional entries were accompanied with a self-assessment form (Appendix VI), which required information on the following points:
1. Why did you choose the specific entry?
2. What did you manage to do well in the particular entry?
3. Set your goals for future writing improvement and be specific about one or two things that you wish to be better at.

Self-assessment forms (see Appendices VII and VIII) were completed where the students were guided towards the development of metacognitive strategies in order to reflect on kinds and number of mistakes they made, the reason(s) why they made them and any measures that they would take in the future so as to avoid such mistakes. The self-assessment of paragraph writing strategies (Appendix IX) guided the students into checking whether the topic of each paragraph was clear, whether they had used details to help the reader understand their point, whether they had used a variety of vocabulary to make their writing 'richer' and whether they had double-checked their papers for mistakes before submitting them. Similar purposes were served with the self- assessment form of writing strategies (see Appendix X), which the learners had to consult and fill in for each entry they included in their portfolios. This form provided them with useful techniques, which the learners could use before, during and after their writing so as to improve it.

Peer-assessment, which can be defined as a *'response in some form to other learners' work'* (Puhl, 1997:7) was very smoothly introduced with the follow-up work. There, the students were presented with an excellent composition (see Appendix XI), written by one of their classmates (Alexandra), and were asked to reflect on what they had been taught about paragraph writing, and assess their peer by finding what made her work particularly exceptional and by mentioning what exactly appealed to them most. The students were also asked to participate in the process of defining the criteria that would determine the success of their performance. At the end of the project they were provided with the list of criteria (see Appendix XII) they had themselves set in class and were asked to evaluate their classmates' portfolios. The last question in this peer-assessment list required the student-assessor to select the entry that s/he enjoyed most and to record their peer's strongest and weakest point. However, the validity of the students' peer assessment could be questioned, as perhaps different friends would want to please or displease each other for reasons other than fair evaluation of writing.

The students' involvement in the procedure culminated with their evaluation of the portfolio project as a whole. What they had to do was answer the two open-ended questions below:
1. How do you assess the portfolio as a means for better learning and fairer student assessment?
2. Do you believe that your writing has improved this year? Would you like to continue the portfolio project next year?

Step 6: Communicate purpose and results. Apart from sharing decisions at every stage of the experiment with the students, it was also considered important to communicate with the parents. A letter was sent to parents at the beginning of the study (Appendix XIII), where the purposes of the portfolio development were presented. It was stressed that the portfolio would not only assess what, but also how the students

learn.

Concerning the results of the study, these were communicated to parents both with the actual portfolios, which they could access at home with the help of a cover sheet (Appendix XIV), including the number and titles of all entries, as well as with the presentation of the portfolios in class. During the presentation the students had a chance to expose their best and worst entries and to explain to their parents what they enjoyed most from their portfolio work and why. At the same time parents provided oral feedback to the teacher on the effectiveness of portfolios concerning raising their children's motivation and school life in general.

At the end of the presentation the teacher provided each one of the students with a certificate (Appendix XV), congratulating them on their participation in the portfolio project and commenting on their effectiveness and their development as writers during the study. Suggestions were also provided for areas that required further improvement.

Portfolios also proved helpful when parent-teacher conferences were held. As Grace (1992) correctly argued, with the portfolios as the basis for discussion the teacher and the parents were able to review concrete examples of their children's work rather than tried to discuss the children's progress in the abstract.

Discussion of Results

Positive aspects of the use of the portfolio as an alternative assessment tool of the skill of writing

The first positive aspect of the writing portfolio was that it became an effective means of improving students' writing skills. Guided by their teacher's instruction the students in the present study learnt to generate ideas in class, to plan and organize their work and finally to revise their work, based on their teacher's feedback.

When the students were asked to set the criteria for assessing the whole portfolio, they said it should have a variety, it should be well organized, the paragraphs should be clear in every sample of writing, spelling and grammar mistakes should be non-existent. This showed that the students internalized the strategies involved in developing good writing habits, were able to express them and could, therefore, try to apply them when they produced any piece of writing. Writing became for the students *'an on-going process of discovery'* (Raimes, 1983:142). Careful observation (Appendix XVI) of the students' revised texts showed that the students took advantage of the teacher's comments on the first drafts and made changes to help the reader understand their point better and share their experiences.

The students were given an opportunity with the optional entries in the portfolio to practise their writing skills beyond class. The optional entries in the portfolios not only enabled a considerable number of students to produce a plethora of writing texts because they actually enjoyed writing, but also to practise other genres, such as poems, fairy tales, simple stories through the use of drawings and comic strips as well as simple biographies from their favourite football players, which were not included in the syllabus. The optional entries were also enlightening for the teacher, who had a chance to learn more about particular students' feelings and hidden talents and developed a more intimate relationship with each student.

Portfolios provided a fairer way of the students' assessment of writing. Some of their comments were the following (translated from Greek):

> *"The portfolio is a better way to learn. The teacher can understand more than she can understand from a single test and she can also see the students' character"* (Alexandros, 11)

> *"The portfolio is very good and fair because you can insert assignments only when you have studied"* (Catherine, 11)
> *"I consider the portfolio a smoother way of assessment, much smoother than a test"* (Alkiviadis,11)

The comments above suggest that the students felt that writing is not terminal, it is not something you can prove over a day. Because of their ongoing nature, students considered the portfolios to be the most efficient assessment method because they had more opportunities to demonstrate their language skills.

Challenges in using the portfolios in the classroom

The only challenging thing that was observed during the portfolio project was the extra-managerial time that was required for its successful implementation in the classroom. Extra time was required from the teacher at home to analyse and interpret the different students' entries, and also for the relevant extra classroom instruction, intended to cater for the students' actual needs or lack of knowledge on particular areas. Moreover, procedural issues, such as storing the portfolios at a safe place, filling in the contents page, arranging the samples in chronological order and discussing with the students how to complete the various questionnaires required a considerable amount of time. However, according to the students' comments the whole project was not a waste of time, but a matter of learning *'more and better things'.*

Positive attitudes promoted with the use of writing portfolios in the classroom

The use of portfolios as an assessment tool in the classroom proved beneficial for a number of reasons:
- The students' motivation increased and even weak students became more actively involved in classroom activities.
- Students were encouraged to take a more active role in assessing and evaluating their own progress, and set their future learning goals with the teacher's help.
- Through the different stages of the portfolios, the students were given teacher support to allow them to make continuous progress.
- Continuous teacher support maintained the students' self-esteem and confidence, thus prompting further learning by strengthening their disposition to learn.
- The drafts and revised works could tell the students' story as practicing writers, and could provide evidence for all the stages of their development, namely effort, progress and achievement.
- The peer comments and self-assessment forms provided the students with a chance to develop their learning-to-learn strategies.
- Students realized that writing strategies, such as discussing ideas with someone, planning, collecting ideas in lists, paraphrasing, adding and subtracting information, checking for mistakes in spelling, grammar and punctuation would help them produce a nice piece of writing.
- Each student was trained in identifying both their strengths and weaknesses and felt valued as a learner, which encouraged lifelong learning.
- Students learnt to express what had not worked well and what action they thought should be taken.
- Students were able to recognize the kinds of mistakes they made and were willing to share the responsibility with their teachers.
- The portfolios managed to create an environment of mutual trust and understanding between the teacher and the students, which in the long run might contribute to learner autonomy.
- The learning of English at school became an interesting experience for the students and the 'status' of the teacher and the course rose. *'The classroom horizon was expanded and each child's canvas was enlarged'* (Grace, 1992:4)
- The inclusion of the parents in the project satisfied the idea of the 'open school'.

Suggestions for further research

It could be argued that since this portfolio experiment brought only gains to all the participants involved, it could be applied in different situations with similar results. Future research could address the following questions:

- How convenient and manageable can a portfolio with a broader focus (i.e. reading, speaking and listening) be?
- Could portfolios be implemented for large-scale assessment projects and if yes, how reliable would results be if used for promotion, placement or admission to universities?
- Will portfolios have the same success if applied with older students in secondary education?
- What type of training do teachers need to successfully implement portfolios with their students?
- How easy is it for in-service teachers to deal with the theories and practicalities of portfolio assessment if they are not given the opportunity to attend training seminars?

Conclusion

In conclusion, the study presented in this paper suggests that portfolios are an innovative way of assessment for young learners' writing. It also offers opportunities to evaluate learners' language development as a whole and to guide classroom instruction according to the learners' needs.

Portfolios have proved able to document a story for every student instead of just providing us with a set of scores. As a final word to this paper, I quote Huerta-Macias' (2002:342) comment, which also reflects insights I gained through this study: *'What is the ultimate goal of evaluation if not to give us the knowledge to be able to reflect on, discuss, and assist a student's journey through the learning process?'*

Hopefully, this paper provides convincing evidence that the portfolio writing assessment gives teachers the power to do all three.

Author's email: melbaramp@sch.gr

References

Elbow, P. & Belanoff, P. (1991). 'State University of New York at Stony Brook portfolio-based evaluation program'. In P. Belanoff & M. Dickson (Eds.), *Portfolios: process and product.* Portsmouth, NH: Heinemann, 3-16.

Gearhart, M. & Wolf, S. (1995). *Teachers' and students' roles in large-scale portfolio assessment: providing evidence of competency with the purposes and processes of writing.*CSE Technical Report 406, National Centre for Evaluation, Standards, and Student Testing (CRESST). Los Angeles: University of California, Los Angeles.

Genesse, F. (1994). 'Assessment alternatives.' *TESOL Matters*, 4/5.

Genesse, F. & Upshur, J. (1996). *Classroom-based evaluation in second language education*. Cambridge: Cambridge University Press.

Grace, C. (1992). 'The portfolio and its use: developmentally appropriate assessment of young children.' *ERIC clearinghouse on elementary and early childhood education Urbana IL*, accessed at http://searcheric.org/digests/ed351150.html on 29 Nov 2002.

Hamp-Lyons, L. & Condon, W. (2000). *Assessing the portfolio: principles for practice, theory, and research*. NJ: Hampton Press.

Huerta-Macias, A (2002). 'Alternative assessment: responses to commonly asked questions'. in J. C. Richards & A. W Renandya (Eds.), *Methodology in language teaching: an anthology of current practice*. New York: Cambridge University Press, 338-343.

Jongsma, K. S. (1989) 'Portfolio assessment'. *The Reading Teacher*: 264-265.

O'Malley, M. J. & Pierce, L. V. (1996). *Authentic assessment for English language learners: practical approaches for teachers*. New York: Addison-Wesley.

Perone, V. (1991). *Expanding student assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Pierce, V. L. (1998). 'Planning portfolios'. In D. J. Brown (Ed.), *New ways of classroom assessment*. USA: TESOL.

Puhl, A. C. (1997). 'Develop not judge continuous assessment in the EFL classroom.' *English Teaching Forum*, 4-9.

Raimes, A. (1983). *Techniques in teaching writing*. Oxford: Oxford University Press.

Smith, K. (1996). 'Action research on the use of portfolio for assessing foreign language learners'. *IATEFL Testing Newsletter*, Issue of April: 17-24.

Στεφανάκη, Ε. Η. (1997). 'Η δύναμη του portfolio: ένας τρόπος αξιολόγησης'. *Ανοικτό Σχολείο*, 24.

Trim, J. (1997). 'A European language portfolio: some questions relating to its nature, function, form, preparation and distribution'. In Council of Europe, *The European language portfolio. proposals for development*. Strasbourg: Council of Europe, 3-12.

ΥΠΕΠΘ (1999). *Funway English 2 Student's Book*. Αθήνα: ΟΕΔΒ.

Φωτιάδου, Τ. (2001). 'Το portfolio ως μέσον ανάπτυξης δεξιοτήτων των μαθητών: μια πρόταση'. *Επιθεώρηση Εκπαιδευτικών Θεμάτων*, 5: 128-138.

Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.

Appendix I
Different ways of scoring the entries in the portfolio

Entry 1	Primary trait scoring	Provision of pictures to help students with ideas and vocabulary
Entry 2	Analytic rating scale	Assessment done on the revised form after comments have been provided
Entry 3	Analytic rating scale	Marginal comments in a code form have been provided for revision
Entry 4	Analytic rating scale	Follow-up work on the previous entry has preceded and planning instructions have been provided (no revision of work)
Entry 5	Analytic rating scales	The letter of invitation has been provided and analysed in class (no revision allowed)
Entry 6	Allocation of marks for each correct answer	No help provided

Appendix II
Different Scoring Methods

The student:	Consistent Control (4)	Reasonable Control (3)	Inconsistent Control (2)	Little/No Control (1)
Uses the simple present correctly (particularly 3rd person singular)				
Knows and spells correctly the days of the week				
Uses the correct prepositions eg. on + days				
Can spell and punctuate correctly				
Organises a fluent paragraph giving additional information beyond the sentence level				

Appendix III
ANALYTIC RATING SCALE

	Task Achievement	Language Use	Cohesion	Mechanics
5	Πολύ καλή ανάπτυξη του θέματος. Όλες οι σχετικές πληροφορίες δίνονται με αποτελεσματικό τρόπο	Κανένα ή σχεδόν κανένα γραμματικό λάθος	Οι προτάσεις έχουν απόλυτη συνοχή	Δεν υπάρχουν ορθογραφικά ή λάθη στίξης που να χαλούν την όμορφη εικόνα του γραπτού
4	Καλή ανάπτυξη του θέματος με τις περισσότερες πληροφορίες δοσμένες με σαφήνεια	Μικροπροβλήματα στη χρήση των χρόνων, στη σειρά των λέξεων, τις αντωνυμίες, προθέσεις, άρθρα, αριθμούς κ.λ.π.	Ικανοποιητική οργάνωση και παρουσίαση του κειμένου	Μικρολαθάκια σε ορθογραφία και στίξη που δε χαλούν την όμορφη ροή του λόγου
3	Αποδίδεται μόνο η γενική ιδέα χωρίς πολλές λεπτομέρειες	Σημαντικά προβλήματα στη χρήση των χρόνων, αντωνυμιών, προθέσεων, κ.λ.π. χωρίς όμως να καταστρέφεται το νόημα	Στοιχειώδης οργάνωση κειμένου με περιορισμένη χρήση συνδέσμων που απαιτεί προσπάθεια από τον αναγνώστη να κατανοήσει το κείμενο	Ακολουθούνται οι βασικοί κανόνες στίξης αλλά γίνονται συχνά ορθογραφικά λάθη
2	Μερική ανάπτυξη του θέματος	Συχνά και σοβαρά ορθογραφικά λάθη που εμποδίζουν το νόημα του κειμένου	Ελάχιστη οργάνωση και συνοχή	Βασικές παραλείψεις σε σημεία στίξης και πολύ σοβαρά ορθογραφικά λάθη που εμποδίζουν την κατανόηση του κειμένου
1	Περιέχει ελάχιστα για να αξιολογηθούν	Το γραπτό είναι γεμάτο από γραμματικά λάθη	Ασύνδετες λέξεις ή φράσεις χωρίς συνοχή	Δεν υπάρχουν σημεία στίξης και τα ορθογραφικά λάθη κάνουν το γραπτό δυσανάγνωστο

Appendix IV
THE CORRECTION CODE FOR REVISING THE WRITING SAMPLES

Gr.	Grammar mistakes
S	Spelling mistakes
WO	Word order
WW	Wrong word
?	The meaning is not clear
+	Something is missing

Appendix V
Survey of Writing Interest and Awareness

Name: _____ **Date:** _____

Διάλεξε ένα κουτάκι για κάθε πρόταση (x)	Πολύ	Αρκετά	Λίγο	Καθόλου
1. Μου αρέσει να γράφω				
2. Βρίσκω το γράψιμο εύκολο				
3. Το γράψιμο με βοηθάει στο σχολείο				
4. Μου αρέσει να γράφω σε φίλους μου				
5. Θεωρώ το γράψιμο απαραίτητο στοιχείο της εκπαίδευσης μου				
6. Θεωρώ τον εαυτό μου καλό στο γράψιμο				
7. Μου αρέσει να μοιράζομαι τα γραπτά μου με άλλους (φίλους ή δασκάλους)				

8. Για τι θέματα ενδιαφέρεσαι να γράφεις περισσότερο;

_____

9. Πόσο έχεις βελτιωθεί ως συγγραφέας στα αγγλικά; Τι μπορείς να κάνεις καλά;

_____

10. Τι θα ήθελες να βελτιώσεις στο γράψιμό σου;

_____

Appendix VI
SELF-ASSESSMENT OF OPTIONAL ENTRIES
SETTING IMPROVEMENT GOALS

Your Name: _____Date: _____

1. Διάβασε την εργασία που διάλεξες εσύ να βάλεις στο portfolio σου ακόμη μια φορά και

απάντησε στις παρακάτω ερωτήσεις

α. Γιατί διάλεξες τη συγκεκριμένη εργασία;

_____

_____

β. Γράψε τι είναι αυτό που κατάφερες να κάνεις καλά στη συγκεκριμένη εργασία.

_____

_____

2. Τώρα σκέψου κάποιους στόχους που θα ήθελες να πραγματοποιήσεις στο μέλλον σχετικά με το

γραπτό λόγο.

Γράψε ένα ή δύο πράγματα που θα ήθελες να βελτιώσεις. Να είσαι συγκεκριμένος-η.

_____

_____

_____

Appendix VII

SELF-ASSESSMENT QUESTIONNAIRE

Title of assignment: Write in your diary how you spent your weekend.

Teacher's name: M. Barabouti

School Year: 2002-03

Class: E2

1. Τις τελευταίες τρεις εβδομάδες με ποια γραμματικά ή άλλα φαινόμενα ασχοληθήκαμε στο μάθημα των Αγγλικών;

α) _____

β) _____

γ) _____

2. Πόσο καλά κατέχεις κατά τη γνώμη σου τα παραπάνω;

	τέλεια	πολύ καλά	λίγο	καθόλου
(α)				
(β)				
(γ)				
(δ)				

3. Τι είδους λάθη έκανες στο συγκεκριμένη εργασία; Ποια ήταν τα περισσότερά σου;

_____

_____

4. Τι θα πρόσεχες περισσότερο αν σου έλεγα να ξαναγράψεις μια παρόμοια εργασία ή τι άλλα μέτρα θα έπαιρνες;

_____

_____

Appendix VIII
SELF-ASSESSMENT FORM OF METACOGNITIVE STRATEGIES

Please look at your grammar check sheet and tick the correct box

	Always	**Sometimes**	**Never**
I can use the Simple Present correctly			
I can use the present Continuous correctly			
I can use the Simple Past correctly			
I can form questions correctly			

<u>How many mistakes did you make? Specify the area.</u>

Simple Present (affirmative)	
Simple Present (interrogative)	
Simple Present (negative)	
Simple past (regular verbs)	
Simple Past (irregular verbs)	
Present Continuous (affirmative)	
Present Continuous (interrogative)	
Present Continuous (negative)	
Question Words	
Formation of Questions	

Why did you make the mistakes above and what is your future goal/goals?

_____

_____

_____

Appendix IX
PORTFOLIO WRITING ASSIGNMENT

PLANNING WRITING

<u>Choose one of the following topics:</u>

A) Write about the best or worst experience in your life.

B) Write about the best or worst dream you had.

C) Write about an adventure that you had.

Before you start writing you need to plan your work. Please think well, answer the questions – and then start writing.

1. How many paragraphs will there be in your writing?

_____

2. What will you write about:

    in the first paragraph:

_____

    in the second paragraph:

_____

    in the third paragraph:

_____

    in the fourth paragraph:

_____

3. List your ideas about each paragraph on a separate piece of paper and use those as a plan to help you write your assignment.

SELF-ASSESSMENT OF PARAGRAPH WRITING STRATEGIES

After you finish your writing, re-read what you wrote and check the following:

4. Is the topic of each paragraph clear?_____

5. Did you use enough details to help the reader understand your point?_____

6. Did you use a variety of vocabulary to make your writing 'richer'? Give examples. _____

7. Did you check your writing for spelling mistakes, language use and grammar mistakes? Did you correct any? _____

Appendix X
Self-Assessment of Writing Strategies

Name: _____ Date: _____

Διάλεξε ένα κουτάκι για κάθε πρόταση

Πριν γράψω:	Ναι	Όχι
1. Άκουσα με προσοχή όσα συζητήσαμε στην τάξη	☐	☐
2. Συζήτησα το θέμα με κάποιον (φίλο ή γονιό)	☐	☐
3. Συγκέντρωσα τις ιδέες μου πάνω στο θέμα σε μια λίστα	☐	☐
4. Έκανα ένα προσχέδιο	☐	☐

Ενώ έγραφα:

5. Άφηνα τις λέξεις που δεν ήξερα για να ξαναγυρίσω σ'αυτές λίγο αργότερα ☐ ☐

6. Αντικαταστούσα λέξεις που δεν ήξερα με τις αντίστοιχες ελληνικές ☐ ☐

7. Έγραφα κάτι που δεν ήξερα με περιφραστικό τρόπο ☐ ☐

Αφού τελείωσα το γράψιμο:

8. Έλεγξα το γραπτό μου για να δω αν ήταν 'μέσα' στο θέμα ☐ ☐

9. Ξαναδιάβασα το γραπτό μου για να δω αν όλες οι προτάσεις μου 'βγάζουν νόημα' ☐ ☐

10. Πρόσθεσα ή αφαίρεσα κάποιες πληροφορίες ☐ ☐

11. Έλεγξα το γραπτό μου καλά για λάθη στην ορθογραφία, τη στίξη και τη γραμματική ☐ ☐

Άλλες τεχνικές που χρησιμοποίησα:

Appendix XI
Follow up work on paragraph writing and introduction to peer assessment

Put the following paragraphs in the right order to have a very good assignment.

Dear Diary,

a) After that, I saw my sister jumping . She was great. I'm sure that her horse was the best one. The race was hard. Some people fell down and one lady hurt her head very bad. My sister finally managed to finish fourth.

b) My weekend was fantastic! I went to the horse club twice because my sister would take part in racing games. But not only my sister. Many of my friends took part too and jumped with their horses on high jumps.

c) I must say I had a great time and I hope I will go to racing games again soon.

d) When we reached the horse club, we saw a great deal of big and small horses. There were some white, brown, black, grey, beautiful horses. There was also a little pony, which was not like the others. It was brown and white and it had blue eyes. You don't believe it. It could jump up to 1.10 metres. It's name was Lucky.

Paragraph order

1	2	3	4

A form of peer assessment

Θεωρείς ότι η παραπάνω εργασία ήταν πετυχημένη ή ιδιαίτερα ξεχωριστή; Σχολίασε την άποψή σου γράφοντας συγκεκριμένα σημεία που σου άρεσαν.

_____

_____

_____

_____

Appendix XII
PEER ASSESSMENT FORM

NAME OF THE STUDENT ASSESSED: _____

NAME OF THE ASSESSOR: _____

DATE: _____

Αφού εξετάσετε καλά το portfolio του συμμαθητή σας, αξιολογήστε το σύμφωνα με τα παρακάτω κριτήρια που εσείς οι μαθητές θέσατε στην τάξη.

Tick one of the boxes and write your comments where necessary
Ο/Η συμμαθητής/τρια μου

	🙂	😐	🙁
1) Γενικά νοιάστηκε αρκετά για το portfolio			
2) Έβαλε πολλές και ποικίλες εργασίες			
3) Βελτίωσε τα δείγματα εργασίας του σύμφωνα με τις υποδείξεις της δασκάλας και τις συζητήσεις στην τάξη			
4) Οι εργασίες είχαν πλούτο λεξιλογίου			
5) Οι εκθέσεις είχαν σωστό σκελετό (introduction, main body, conclusion)			
6) Οι παράγραφοι ήταν καλογραμμένες με εισαγωγική πρόταση, λεπτομέρειες και μεταβατική πρόταση			
7) Δεν υπήρχαν πολλά γραμματικά λάθη			
8) Δεν υπήρχαν πολλά ορθογραφικά λάθη			
9) Ο συμμαθητής μου έχει κάνει αυτοαξιολόγηση και φαίνεται να ξέρει που πρέπει να προσπαθήσει περισσότερο			

10) Γράψε τώρα κάποιες πιο ειδικές παρατηρήσεις για το portfolio του συμμαθητή σου (ποιο δείγμα σου σου άρεσε; τι συγκεκριμένα θα ήθελες να βελτιώσει; ποιο είναι το πολύ αδύνατο και ποιο το πολύ δυνατό του σημείο;)

_____

Appendix XIII
Initial letter of communication with the parents concerning the portfolio project

Αγ. Παρασκευή, 6/2/03

Αγαπητοί γονείς,

Φέτος τα παιδιά σας θα εργαστούν προκειμένου να συγκροτήσουν ένα **'φάκελο υλικού'** ή όπως ονομάζεται στα αγγλικά ένα **'portfolio'**. Ο φάκελλος αυτός θα περιλαμβάνει δείγματα ορισμένων γραπτών εργασιών των παιδιών που θα αποσκοπούν στο να δείξουν, μετά από ένα διάστημα τεσσάρων μηνών, **τι αλλά και πώς μαθαίνουν τα παιδιά στο μάθημα των Αγγλικών**. Τα portfolios θα χρησιμοποιηθούν για τους εξής λόγους:

- Για να προσδιοριστούν τόσο οι δυνατότητες όσο και οι αδυναμίες των μαθητών μου **στο γραπτό λόγο**

- Για να σχεδιάσω κατάλληλες δραστηριότητες που θα οδηγήσουν στη βελτίωση της απόδοσής τους

- Για λόγους **αξιολόγησης** των μαθητών από εμένα και τους συμμαθητές τους αλλά και δικής τους **αυτοαξιολόγησης**

Στο τέλος της σχολικής χρονιάς ευελπιστούμε ότι θα καταφέρουμε να σας παρουσιάσουμε σε μια συνάντησή μας τα αποτελέσματα της προσπάθειάς μας, όπου και θα σας ζητήσουμε τα σχόλιά σας για την πρόοδο των παιδιών σας στο συγκεκριμένο τομέα και πιθανόν τη δική σας γνώμη για τη χρήση του 'portfolio' ως **εναλλακτικού τρόπου αξιολόγησης στο σχολείο.**

Προσωπικά πιστεύω ότι μέσω των 'portfolios' τα παιδιά σας θα γίνουν περισσότερο **υπεύθυνα για τη γνώση τους** και θα αναπτύξουν εκείνες τις διαδικασίες μάθησης που θα τους οδηγήσουν σταδιακά στην **αυτονομία τους ως μαθητές.**

Με εκτίμηση,

Μπαραμπούτη Μελίνα

(εκπαιδευτικός Αγγλικών)

Appendix XIV
ELEMENTARY WRITING PORTFOLIO COVERSHEET

Student: _____School Year: _____

Teacher: _____Class: _____

Native Language: _____ School: _____

CONTENTS	DATE	REQUIRED	OPTIONAL
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			

Appendix XV
Sample of Certificate given to the learners at the end of the project

July 2003

To Lydia

Congratulations
on participating in the portfolio
project

Your portfolio was a **very good one**.
It contained a variety of samples, which
showed that you are a very talented
writer.

During the project you made **a lot of
effort** and you also **improved a lot**
concerning the organization and editing
of your writing samples.

Keep on the good work

Your English teacher
Melpomeni Barabouti

Appendix XVI
Teacher's Observation Notes (from second drafts)

Steven: He corrected all his mistakes. He didn't deal with enriching his paragraph

Vicky: She tried hard to produce more words in English. Vicky is a very weak learner but she managed to get involved in the project even with minor contributions.

Francesca: She corrected her spelling mistakes but she made new ones because she added more details in her writing. The structure of the paragraphs remains insufficient.

Alkiviadis: He corrected the grammar mistakes but he made some new ones. He didn't change the content.

Henry: He corrected all the mistakes he had been taught. He extended the content slightly

Chris: He corrected very few, basic mistakes. Problems with prepositions and word order in the sentence

Alexandra: She corrected all her mistakes, which she accepted were all 'careless mistakes' and she said that her writing needed some editing.

Lydia: She made some corrections, she organized her writing and made her work more concrete. Problems with the use of prepositions.

Daniel: He didn't correct his draft at all.

Dimitris: He corrected a lot of spelling and word order mistakes. He restructured his sentences when he was in doubt about something.

Marina: She didn't manage to correct her mistakes, which were most grammar ones.

Konstantinos: He corrected his paragraph and extended it considerably

Nektaria: She corrected the grammar mistakes she had made in the first draft and enriched the content too.

Fivos: He seemed rather lazy to rewrite his first draft but after the relevant prompting he managed to make basic corrections.

Catherine: Very conscientious work She tried hard and managed to organize her work in a more coherent way.

Alexandros: Most of his mistakes were made because he was too much in a hurry to submit his work. Therefore, he managed to correct them all when they were indicated to him

Pantelis: A weak student willing to try when given the relevant guidance. He managed to correct some of his mistakes

Gloria: She made very few changes. Generally she contributed very little top the project. Rather a weak, not very motivated student.

John: He didn't seem to enjoy the portfolio or any other work much. Contributed very few samples.

Despina: She found it very hard to follow the portfolio work. She felt very unsure for herself and stated that she preferred her mistakes to be corrected by the teacher.

Investigating Portfolio Assessment with Learners of the 3rd Grade in a Greek State Primary School

[Διερεύνηση του Φακέλου Αξιολόγησης με Μαθητές Τρίτης Δημοτικού σε ένα Ελληνικό Δημόσιο Σχολείο]

Sophia Kouzouli

Assessment is a field increasingly explored in relation to the parameters it involves. The special characteristics of the learners and the interactive relationship between instruction and assessment lead to the use not only of traditional assessment techniques but also of alternative methods such as the portfolio. This study intends to investigate the implementation of a process portfolio in a Greek state primary school with a class of third graders aged between 8-9, concentrating on integration of skills. The findings show that this technique is appropriate for young learners and that it meets specific pedagogical and assessment criteria. It also exerts positive impact on metacognitive awareness, learner autonomy and positive attitude towards learning. Finally, the findings give insight to emerging problems and issues requiring further research.

༼ༀ༽

Η αξιολόγηση των μαθητών αποτελεί ένα τομέα που όλο και περισσότερο ερευνάται σε σχέση με τις εμπλεκόμενες παραμέτρους. Τα ειδικά χαρακτηριστικά των μαθητών και η διαδραστική σχέση διδασκαλίας και αξιολόγησης συντελούν στην εφαρμογή εναλλακτικών τρόπων - σε σχέση με τις παραδοσιακές μεθόδους- αξιολόγησης, όπως του φακέλου αξιολόγησης. Αυτή η έρευνα στοχεύει στη διερεύνηση μιας μορφής φακέλου αξιολόγησης, του Φακέλου Διαδικασίας, σε ένα Ελληνικό δημόσιο σχολείο, σε μια τάξη με παιδιά τρίτης δημοτικού, 8-9 ετών, εστιάζοντας στη διαπλοκή δεξιοτήτων. Από τα ευρήματα καταδεικνύεται ότι είναι κατάλληλη μέθοδος για τη συγκεκριμένη ηλικία των μαθητών, ότι ανταποκρίνεται σε συγκεκριμένα κριτήρια και ότι έχει θετικό αντίκτυπο στη μεταγνωστική τους δεξιότητα, στη μαθησιακή τους αυτονομία και στη διαμόρφωση θετικής προδιάθεσης προς τη μάθηση. Τα συμπεράσματα καταλήγουν στη διαπίστωση προβλημάτων και θέτουν προοπτικές για τη συνέχιση της έρευνας.

Key words: young learners' characteristics, language acquisition, alternative assessment, metacognitive awareness, portfolio assessment

Introduction

The study reported in this paper investigates the use of alternative assessment with young learners. In particular it focuses on the implementation of a process portfolio with the 3rd class of a Greek state primary school in order to assess young learners' linguistic competence and integration of skills, and gain formative feedback.

The paper discusses the learning characteristics of children (8-9 years old), which can contribute to a better understanding of their needs, the underlying learning philosophy and appropriate teaching practices for this particular age group. It defines the concept of assessment and identifies and examines its parameters in relation to the specific age group. Then, the paper focuses on establishing an alternative assessment framework relevant to the diverse needs of the particular primary school pupils. The analysis generates specific findings considering the practice of portfolio assessment in the classroom, the evidence gathered from the active involvement of the children and recommends ways of improving it. Finally, the paper raises practical issues that need further research.

Literature review

Young learners and aspects of learning

Assessing young language learners requires, as stated by Mckay (2006), the consideration of the special characteristics of young language learners in parallel with the learning principles and teaching practices in L1 as these are also reflected in foreign language teaching. Thus, the focus of this part lies on three components that need consideration when implementing assessment procedures: learners, principles of learning and appropriate practices.

Meggitt (2006) and Hobart and Frankel (2004) suggest that young learners, from five to twelve years of age, are different from other learners, teenagers or adults, due to certain special features which determine the way they think and learn. According to Piaget's classification (discussed in Boden, 1994) the participants of the present study fall in the third stage, the 'concrete operational stage'. At this stage logic develops and young learners undergo cognitive, social, affective and, due to their age, physical development.

As for their cognitive development, learners are in the process of developing basic cognitive and reflecting skills -perception, memory, concept formation, symbolization and critical thinking. Their attention span is generally short and they are likely to get distracted and bored rather easily. They need to be involved in active, stimulating, cognitively challenging and problem solving activities; this can be done through play, which constitutes an innate need for young learners. Young learners also begin to self correct and evaluate their performance. They develop the ability to read both aloud and silently as well as to read for information or pleasure. Also, although they combine drawing and writing, their writing can convey meaning on its own.

As far as their social development is concerned, learners at this age have already started to detach themselves from their egocentrism. They are in the process of developing an understanding of their own character as well as of their self in relation to others. Vygotsky's (1978) zone of proximal development (ZPD) and Bruner's (1983) notion of scaffolding postulate that children should be helped to gradually build up their understanding and skills, to interact with their peers and to be involved in pair or group work.

As for their affective development, children at this age are spontaneous and motivated; they need to deal with familiar genres and cooperate with familiar adults. When young learners are exposed to modeling and demonstration through rewards, their confidence and self-esteem are positively influenced and learning can take place.

Children's physical growth constitutes another major issue. Children over seven years have not adequately developed hand-eye coordination; yet they have developed the ability to move around and have increased their fine motor skills. They get tired more easily from sitting still rather than from participating in energetic activities.

Another dimension of children's theory of learning lies in the theory of multiple intelligences introduced by Gardner (1983). The notion of intelligence is not limited to one general abstract idea but is distinguished into multiple types involving special kinds of abilities and strengths. Every child has a special form of intelligence which should be matched to activities used in primary school lessons.

Because the context of this study is English as a foreign language it should be pointed out that comprehension and acquisition take place when learners are exposed to forms and structures which are just beyond their current level of competence in the language, referring to this relation as "i+1", input level plus one (Krashen and Terrell, 1988). According to Krashen and Terrell's affective filter hypothesis (ibid) relates affective factors to successful acquisition.

Assessment in language learning

Language assessment is typically distinguished into summative and formative. Summative assessment - assessment of learning - is described by Black and William (1998) as any assessment implemented at the end of a course to evaluate learners' competences. Black et al. (2003) explain that formative assessment - assessment for learning- is more frequent, involves a variety of methods and provides information which is used to adapt the teaching work to meet learning needs.

The teaching and learning models which focus on the communicative use of language in everyday life situations, i.e. singing the lyrics of a song or playing a game, and the fact that conventional testing squeezes out the joy and motivation which are inherent characteristics of young learners have caused a gradual change from norm-referenced to criterion based and to performance- based assessment. Rixon (2004) suggests that alternative assessment may involve a variety of practices: learner diaries, journals, interviews, observations, learner-teacher conferences, peer and/or self-assessment and portfolios. Portfolios, as Jones and Coffey (2006) postulate, involve a variety of methods and allow continuously recording achievement. Thus, they can offer a valuable assessment framework for primary school.

Portfolio assessment

The literature concerning portfolio assessment provides several definitions; Simon and Forgette-Giroux (2000) advocate that portfolio assessment is a cumulative and ongoing collection of entries selected according to a given framework and aiming at assessing development of a specific competency. The Council of Europe (2001) has defined three types of European Language Portfolios for primary school, secondary school and young adult life. There are different forms of portfolios, each of which serves a specific purpose, but in practice they are interrelated and overlap.

The value of portfolio assessment lies in the pedagogical and reporting qualities it possesses. The most dynamic characteristic of the portfolio is that it requires learners to assemble real evidence which provides an authentic description of what learners can do. Hamp-Lyons (1996) and Caudery (1998) argue that the contents of a portfolio are not limited to conventional testing activities such as multiple choice, true/false, matching or gap filling but involve activities which offer a complete picture of the learners' overall performance. All this evidence can be used to help young learners' language, cognitive, social, affective and physical development, and to detect problems in these areas.

Portfolios are a powerful means which encourage learners to be more involved in reasoning processes, an issue which Harmer (2006) considers as the touchstone for learning. Learners are involved in the organization of the selection of their work; they need to be critical about the work they collect as well as compare earlier and present work and, finally, draw inferences about their development. Hebert (1998) points out that this process of reflecting, developing descriptive language for the particular work selected and using metalanguage to talk about language contributes significantly to the child's metacognitive growth.

Learners are motivated to collaborate and interact with each other as well as with their teacher in order to implement self- and peer- assessment and realize their strengths and the actual areas that need improvement. Ioannou-Georgiou and Pavlou (2003) suggest that teachers have a concrete and tangible reason for arranging regular conferences with their learners. Both teachers and learners have the opportunity to develop their social skills in a cooperative atmosphere by getting to know each other better and by establishing a strong relationship which will yield beneficial educational and pedagogical effects.

The student product is highlighted not as an outcome per se; it is subjected to a certain creative procedure which is distinguished into three phases (Kemp and Toperoff, 1998). In the first phase, the collection, learners are responsible for collecting the samples needed for the compilation of the portfolio. Learners, especially young learners, are not accustomed to documenting their work. In Greek state primary schools learners use folders in order to save their class work but they are not obliged to follow a particular organizational pattern or a specific chronological order; therefore, they may have difficulty in getting used to adopting a more disciplined way. Thus, this phase requires thorough preparation and negotiation with the learners. The second phase involves the selection of the samples which is based on specific criteria related to the general purpose of portfolio. The third phase, reflection, is of great importance as it distinguishes portfolio from the mere collection of work in folders. Learners are asked to reflect upon and respond to the actual process of the lesson, to their performance and to the performance of their peers; this can be done in writing as well as orally, particularly with younger children. This final phase is a skill in itself. Teachers need to help learners master reflective skills and teach the practice of self- and peer-assessment by providing instruction with a lot of practice and feedback.

Baume (2002) concludes that portfolio assessment is a valid vehicle for both ongoing and terminal assessment. Learner achievement is judged against the intended outcomes of the course as these are presented in the portfolio itself. Berk (2002) suggests that the wide range of procedures and measures gathered over a long period of time can ensure the soundness, trustworthiness and legitimacy of it. He finds evidence of content validity as the outcomes being measured are representative of the teaching practices. Construct validity is apparent as the concrete evidence of learner performance reflects the underlying skills assessed and can support the inferences based on their assessment. Predictive validity is catered for as the evidence of the learners' performance can predict future use of the

language. Hamp-Lyons and Condon (2000) advocate that portfolios possess face validity since the samples that are collected give a picture of the abilities or knowledge of the learner along with the predetermined objectives.

Another important merit of this alternative way of assessment is the profound positive impact it exerts. Schneider and Lenz (2001) advocate that learners and their environment, teachers and language teaching in general benefit from portfolio assessment. Learners, who are helped to understand the learning process, to develop metacognitive skills and to self monitor, feel a sense of achievement and are eager to continue, while teachers obtain ample and clear feedback which they can use for a multitude of purposes, for example to focus on developing specific reading or speaking strategies, and in a multitude of ways, for example with certificates of excellent student performance.

Research context

In most Greek state primary schools English starts in the 3rd grade and is taught on a three 45 minute lesson basis per week. Until September 2011, when a new book was introduced for the 3rd grade, teachers of English were obliged to choose the course book they wanted for the third grade from a list of books from the EFL market. For grades 4 to 6 they have to use a textbook series prescribed by the Ministry of Education in collaboration with the Pedagogical Institute.

Primary schools do not involve formal assessment of pupils' language learning; pupils move up from one class to the next without examinations on the condition that they have attended at least half of the lessons and that the teacher judges that the knowledge and skills they have developed are appropriate to their class – appropriate is specified by the curriculum objectives, the class syllabus and the particular subject. In practice, assessment is based on the overall performance of students in class, oral work and homework and the revision exams pupils sit towards the end of each trimester. The frequency of paper-and-pencil tests and the use of other techniques such as self- and peer-assessment, observations, projects and portfolios are at the teachers' discretion.

Methodology of the study

The present research study set out to investigate whether portfolio assessment is an appropriate assessment technique for young learners in a Greek state primary school, pertinent to the characteristics of young learners and examine the implications for teaching and assessment as described in the literature review. The study took place over a short period of time which lasted three months.

The class consisted of fifteen learners, six boys and nine girls, aged between eight and nine, who lived in a village in the province of Elia in the Peloponnese, in the South of Greece. The learners used an English textbook, '*Zoom a'* by Mitchell and Parker (2000), accompanied by workbook and companion book. The teacher considered the specific textbook series appropriate for the needs of the pupils and compatible with peer- and self- assessment techniques.

The collection of evidence is based on methods of triangulation which, as Brown and Rodgers (2002) argue, can be used to refer to the attempt to understand some aspects of the learners' behaviour by studying it from more than one standpoint. Thus, the study uses multiple data gathering procedures with a focus not on statistics but on practical

significance. The sources of evidence include portfolio entries comprising tape-recorded performance and written tasks, open-ended questionnaire items and closed-response items involving self- and peer-assessment, a letter written by parents, and an evaluation form concerning the overall project completed by the learners.

Research questions

The study addressed the following research questions:
- Can portfolio assessment provide a young learner-centred perspective to assessment?
- Can portfolio assessment assess skills reflecting real life language use?
- Can portfolio assessment create a metecognitive environment?

The implementation of portfolio assessment

The procedures for the implementation of portfolio in this study were based on Kemp and Toperoff (1998) and Tsagari (2005) because they were manageable and applicable in this particular context.

Establishing Purpose

The teacher took into consideration the parameters of teaching English to young learners described in the literature review and realized that they needed to improve their overall achievement in English and develop in all four skills in an integrated mode. Therefore, the general aim of the portfolio was to assess integrated skills reflecting real-life language use.

Establishing Portfolio Format and Content

The teacher familiarized learners with the format of their portfolio. She explained to them that this would be divided in two parts, containing compulsory entries which would provide the basis for assessment purposes, and optional work which would allow pupils to show their talents and their best work.

Compulsory work involved eight tasks the pupils would carry out in class, including tape-recorded material based on listening and speaking activities, playing a game, writing, reading, drawing and colouring. It included three peer-assessment questionnaires (Appendix I) in L2 and eight self-assessment questionnaires (Appendix II), most of them in L1. Optional items would include two parts: 'My Reading Log' (Appendix III) and 'My Choice' (Appendix IV).

Establishing Ownership

The teacher helped the learners with the organization of the portfolio and encouraged them to gradually assume responsibility for its completion, from filling in the dates to arranging them in a chronological order.

Introducing the Idea of Portfolios

To avoid confusing learners with the word 'portfolio' the teacher explained to them where it derives from and what it actually means. She also showed them a file folder in which she had compiled a few of the compulsory activities they would need to do in the future along with self- and peer-assessment checklists in order to illustrate what the portfolio would look like.

Notifying Other Interested Parties

The teacher announced to the teaching staff and the principal that the third class of the primary school would be engaged in the compilation of a portfolio in the framework of portfolio assessment. Before the beginning of the portfolio assessment the teacher sent an informative letter to the parents. She also sent a second letter in the middle of the project (Appendix V) and asked parents to reply (Appendix VI).

Setting the Guidelines for Portfolio Presentation

A week before the compilation started the class discussed how the portfolio would be organized. They also discussed that learner-teacher conferences and reflection on their work would be necessary for learning and pedagogical purposes. They rehearsed the techniques and discussed success criteria to develop their confidence. Finally, the teacher placed the guidelines on the wall (Appendix VII) written in Greek for all the pupils to understand.

Preparation Period

As the participants were at a young age and they may easily be confused, lose interest or even lose confidence in their abilities pupils needed ongoing understanding, immediate guidance and timely feedback as far as the reflection part and the general organization of the portfolio were needed. They particularly liked 'playing teacher' and were able to make non-threatening, supportive and direct comments.

Assessment of the Portfolios

The teacher used four checklists, one of which is presented in Appendix VIII, and four global rating scales, an example of which is presented in Appendix IX. The global rating scales were more elaborate in consistency as she wanted to assess a variety of parameters but did not want the rating scales to be impractical for her to use. Additionally, the teacher took notes of the mistakes pupils made during the tasks and informed the learners about them and also modified teaching by using among others more realia, simplifications and Total Physical Response activities in order to help them improve.

The negotiation that was required for the completion of the self- and peer- assessment reports, the reflection cards for their reading logs and the reports for the optional collection of other activities fostered their metacognitive skills; students were able to reflect on their performance, evaluate themselves and set personal targets. Fisher (1989) postulates that this procedure encourages learners to be impartial and sincere.

Before the presentation of the portfolio the teacher discussed with each learner separately their final product and encouraged them to reflect on the quality of their work. This facilitated the assessment of the portfolio as a whole. After completing a questionnaire for the portfolio presentation as a whole, learners wrote the cover letter (Appendix X).

Follow-Up

At the end of the portfolio period learners and their family members were invited in the classroom to look at the complete version of portfolios. Learners were awarded a certificate which congratulated them on their effort and their work. The teacher prepared a letter in L1

for every pupil which was added in their portfolio and which outlined the weaknesses and stressed strengths, generating an individual profile for every learner.

Use of Portfolio Results

The portfolio contributed to diagnosing the strengths and weaknesses of learners and to monitoring their progress. It also assisted the development of learning and reasoning skills. It provided feedback both to the teacher and to the learners. Finally, it succeeded in involving parents in the teaching and learning process.

Discussion

The discussion of the results is based on the interpretation of the evidence collected and on the observations made during the study. Portfolio assessment proved an efficient means of assessing young learners' linguistic competences on a meaningful and contextualized basis in a variety of natural situations which they could face in their everyday life either in Greek or in English and succeeded in providing an authentic description of what these learners can do. The tasks employed reflect Piagetian thinking: they creatively stimulated the learners' imagination, related them with reality and involved an element of play. Pupils were able to show their competence in one individual skill as well as their ability to use more than one skill to achieve the overall aim of a task. The exposure to the recordings of their interactions raised their awareness of speaking and listening skills. The variety of writing tasks helped them develop their writing skills. The reading tasks and the reading of other stories enabled them to progress from reading aloud to reading silently.

The reflection over the criteria involved in self- and peer- assessment enabled students to internalize strategies that helped them develop their metacognitive awareness. Additionally, the reflection over the completion of the self- and peer-assessment reports helped them have a clear and well-marked way to successful performance, as also discussed by Gottlieb (2006). The optional items provided the learners with the chance to read several stories, reflect on them and draw a part or a picture of the story that impressed them. They were also able to review all the work they had done at school or at home, reflect on their performance and the learning targets and then select the ones they considered important for specific reasons which they wrote on their comment cards. Thus, students were given the opportunity to feel a sense of independence and autonomy, focusing on the learning process rather than on the aspiration to achieve a better grade. Learning was encouraged through experience or personal discoveries.

Materials, routines and relationships involved repetition, recycling and cooperation. An enjoyable and motivating environment was established, which supported emotional and social development. Pupils took pride in their accomplishments, were not embarrassed or afraid to be sincere with themselves and their peers and, finally, felt mutual respect for each other's work. This can be illustrated in their cover letters, in which students mentioned that they were happy with their tasks and that they wanted to repeat them and use them for other school subjects.

Parents had the opportunity to experience concrete examples of what their children did at school and to obtain useful insight into learners' weaknesses, strengths, preferences and attitudes. They were not only recipients of knowledge about their children but also active members in the learning process of their children thanks to the feedback they provided to the teacher.

The teacher was able to diagnose learners' strengths and weaknesses. For example, on one occasion, she heard a boy, who was until then considered shy and unable to participate in activities that required interaction with other learners, whisper correct utterances but not talk loud enough for the other learners to hear. Also, she was able to monitor learners' development and do remedial teaching when necessary. The other teachers, who were skeptical in the beginning, were taken aback by the learners' enthusiasm and wide spectrum of strengths it revealed. Yet they insisted on regarding it as a demanding task which would be difficult for them to undertake.

Overall, the results of the study showed that portfolio assessment is a valid assessment tool for assessing learner performance for a number of reasons. Initially, it gives an account of the performance it intends to assess. Secondly, the systematic procedure and the materials used correspond to the teaching practices. Thirdly, it predicts effective use of the language. Furthermore, it is a reliable method as there are clear assessment criteria and marking schemes. Last but not least, it is a fair method involving work over a period of time and allowing learners to revise and to comment on their work.

Limitations and suggestions for further research

Time management, storage and the financial issue concerning the expenses for purchasing folders and cassettes, were parameters which required consideration. The age of the learners was an important factor for using Greek in some self- and peer-assessment activities and in the comment cards for the optional tasks they selected, yet not in their reading logs. This happened in order to facilitate learners reflect on their work. Some might disagree as they would expect pupils to use only the target language.

The findings raise several important issues and challenges for further research:
- Portfolios could be used in primary school as a reporting, pedagogical as well as the main assessment tool. In this case, the rating scales or checklists used by the teacher might be included in order to provide a profile of the learner that incorporates both the learner's and the teacher's perspective.
- Portfolio projects could be incorporated in all grades of primary school to ensure academic consistency and to be further used as a longitudinal tool in documenting learners' achievement as well as self- and peer-assessment.
- The paper load learners have to complete should be reduced without eliminating the beneficial dynamic of portfolio.
- Portfolio assessment requires proper and continuous professional development and support.
- The questions asked to the parents should be handled with clarity in order to establish that their involvement is balanced and will facilitate the educational process.

Conclusion

This study provided evidence of the validity of portfolios for the assessment of young learners. Initially, portfolio assessment contributed to the cognitive, social and affective development of the learners. Moreover, it served as a common reference for communication between the teacher, learners and their parents. Finally, the learners involved developed metacognitive strategies and a positive attitude towards learning. This final issue is of critical importance as negative attitudes formed at this age are likely to impede not only language acquisition but also successful interaction with the students' environment when encountering challenges.

Author's email: sophiakouz@yahoo.com

References

Baume, D. (2002). 'Portfolios, learning and assessment.' Accessed at http://www.recordingachievement.org/downloads/Assess_Portfolios.pdf on 10 Oct 2005.

Berk, R. (2002). 'Teaching portfolios used for high-stakes decisions: you have technical issues!' accessed at http://www.nesinc.com/PDFs/2002_06Berk.pdf on 5 Nov 2005.

Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. (*2003*). *Assessment for learning - putting it into practice*. Buckingham: Open University Press.

Black, P. & William, D. (1998). 'Inside the black box: raising standards through classroom assessment'. *Phi Delta Kappan,* 80/2: 139-148.

Boden, M. (1994). *Piaget*. London: Fontana Press.

Brown, J. D. & Rodgers, T. S. (2002). *Doing second language research*. Oxford: Open University Press.

Bruner, J. (1983). *Child's talk: learning to use language*. New York: Norton.

Caudery, T. (1998). 'Portfolio assessment a viable option in Denmark?' *Sprogforum*, 11/4: 51-54.

Council of Europe (2001). *Common European Framework of Reference for languages: learning, teaching, assessment.* Strasbourg: Council of Europe.

Fisher, R. (1998). 'Thinking about thinking: developing metacognition in children.' *Early Child Development and Care,* 141/1: 1-15.

Gardner, H. (1983). *Frames of mind: the theory of multiple intelligences*. New York: Basic Books.

Gottlieb, M. (2006). *Assessing English language learners: bridges from language.* California: Corwin Press.

Hamp-Lyons, L. (1996). 'Applying ethical standards to portfolio assessment of writing in English as a foreign language.' In M. Milanovich & N. Saville (Eds.), *Performance testing and assessment: selected papers from the 15th language testing research colloquium*. Cambridge: Cambridge University Press, 151-164.

Hamp-Lyons, L. & Condon, W. (2000). *Assessing the portfolio: principles for practice, theory and research.* Cresskill: Hampton Press.

Harmer, J. (2006). 'Engaging students as learners.' *English Teaching Professional,* 42: 4-6.

Hebert, E. (1998). 'Lessons learned about student portfolios (benefits of student portfolios).' *Phi Delta Kappan*, 79/8: 583-585.

Hobart, C. & Frankel, J. (2004). *A practical guide to child observation and assessment.* Cheltenham: Nelson Thornes.

Ioannou-Georgiou, S. & Pavlou, P. (2003). *Assessing young learners*. Oxford: Oxford University Press.

Jones, J. & Coffey, S. (2006). *Modern foreign languages, 5-11: a guide for teachers*. London: David Fulton Publishers Ltd.

Kemp, J. & Toperoff, D. (1998). 'Guidelines for portfolio assessment in teaching English.' *English Inspectorate Ministry of Education*, accessed at http://www.anglit.net/main/portfolio/default.html on 15 Jan 2006.

Krashen, S. D. & Terrell, T. (1988). *The natural approach: language acquisition in the classroom*. Oxford: Pergamon.

McKay, P. (2006). *Assessing young language learners*. Cambridge: Cambridge University Press.

Meggitt, C. (2006). *Child development*. Oxford: Heinemann.

Mitchell, H. Q. & Parker, S. (2000). *Zoom A*. London: MM Publications.

Rixon, S. (2004). 'Assessment of young learners of English: keeping track without turning them off.' Accessed at http://www.eltforum.com/forum/pdfs/aym04_papers.pdf, on 15 Nov 2005.

Schneider, G. and Lenz, P. (2001). *European Language Portfolio: a guide for developers*. Strasbourg: Council of Europe.

Simon, M. & Forgette-Giroux, R. (2000). 'Impact of a content selection framework on portfolio assessment at the classroom level.' *Assessment in Education*, 7/1: 83-101.

Tsagari, C. (2005). 'Portfolio assessment with EFL young learners in Greek state schools'. *ELT News*, 196: 28.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Appendix I

Peer-assessment questionnaire for the 2nd entry

Διάβασε το τραγούδι που αντέγραψε ο συμμαθητής ή η συμμαθήτριά σου και συμπλήρωσε τον πίνακα.

	ΝΑΙ	ΟΧΙ	ΜΕΡΙΚΕΣ ΦΟΡΕΣ
Κάνει ωραία γράμματα.			
Λείπουν γράμματα από τις λέξεις.			
Τα γράμματα είναι πάνω στη γραμμή (δεν χορεύουν στον αέρα).			
Ο γραφικός του /της χαρακτήρας διαβάζεται εύκολα.			
Χρησιμοποιεί κεφαλαία γράμματα όπου πρέπει.			
Έχει ορθογραφικά λάθη.			
Χρησιμοποιεί τα σημεία στίξης: τελεία, κόμμα, ερωτηματικό και θαυμαστικό.			

Appendix II

Self-assessment checklist for the 3rd entry

Διάβασε προσεκτικά τις προτάσεις και συμπλήρωσε τα κουτάκια με τα οποία συμφωνείς.

		ΝΑΙ	ΟΧΙ
1)	Μπορώ να θυμηθώ τα ονόματα των ζώων που έμαθα σήμερα.		
2)	Μπορώ να θυμηθώ τα ονόματα των χρωμάτων που έχουν αυτά τα ζώα.		
3)	Μπορώ να προφέρω σωστά τα ονόματα των ζώων που έμαθα.		
4)	Μπορώ να προφέρω σωστά τα χρώματα αυτών των ζώων.		
5)	Μπορώ να ταιριάξω ονόματα ζώων που ακούω με τις αντίστοιχες εικόνες και να το εκφράσω		
6)	Μπορώ να ταιριάξω ονόματα χρωμάτων που ακούω με τις αντίστοιχες εικόνες και να το εκφράσω		
7)	Μπορώ να κάνω ερωτήσεις: Are you a…….?.		
8)	Μπορώ να απαντήσω σε ερωτήσεις: Yes, I am. / No, I am not.		
9)	Μου αρέσει να παίζω παιχνίδια στα οποία μιλώ Αγγλικά.		
10)	Μου αρέσει να σηκώνομαι από το θρανίο μου για να συμμετέχω σε δραστηριότητες με τους συμμαθητές μου.		

Appendix III
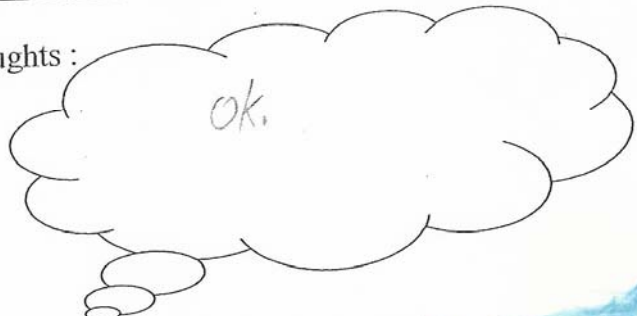
My Reading Log
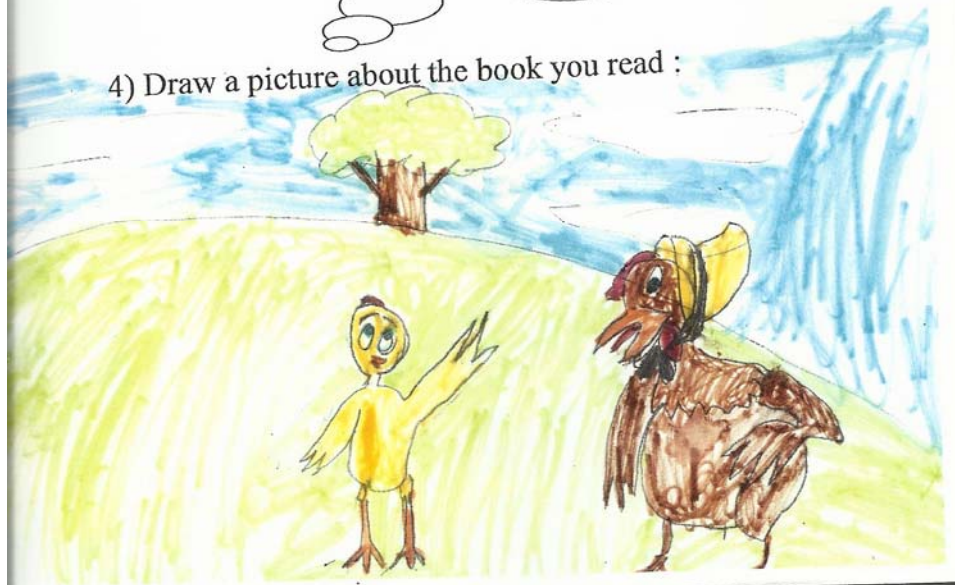
MY READING LOG

NAME : Anastacia Kousandra DATE : 7/9/2006

1) Book title : Chichen Licken

2) Main characters

Chichen Licken Henny—Penny Cocky—Locky Ducky—Lucky Drakey—Lakey foulish birds, Goosey—Loosey Gandar Lndre

3) My thoughts :

Ok.

4) Draw a picture about the book you read :

Appendix IV

My Choice

MY CHOICE

Γιατί διάλεξες αυτήν την άσκηση ;

_____

_____

Έμαθες κάτι από αυτήν την άσκηση ;

_____

_____

Πώς ένιωθες καθώς έκανες αυτήν την άσκηση;

_____

_____

Appendix V

A second letter to the parents in the middle of the portfolio project

---

03/05/2006

Αγαπητοί γονείς,

Η παρούσα επιστολή έχει σκοπό να σας ενημερώσει σχετικά με την πρόοδο των παιδιών σας όσον αφορά το μάθημα της Αγγλικής γλώσσας αλλά και για τις δικές τους σκέψεις σχετικά με την πρόοδό τους μέσα από το *Φάκελο Επιτευγμάτων* (το *Portfolio*) που ετοιμάζουν.

Θα σας παρακαλούσα *να διαβάσετε μαζί με τα παιδιά σας* το γραπτό υλικό που έχουμε συγκεντρώσει μέχρι τώρα και να ακούσετε τις δραστηριότητες που έχουμε μαγνητοφωνήσει. Έπειτα, θα ήθελα να αναφέρετε τις εντυπώσεις σας και να γράψετε λίγα σχόλια σχετικά με ό,τι διαβάσατε και ό,τι σας άρεσε από τη δουλειά των παιδιών σας. Τέλος, θα σας παρακαλούσα να βάζατε το σημείωμα αυτό μέσα στο *Φάκελο Επιτευγμάτων*.

Σας ευχαριστώ για το χρόνο που θα διαθέσετε και για τη συνεργασία σας.

Η καθηγήτρια της τάξης

Σοφία Κουζούλη

Appendix VI

A parent's reply to the 2ⁿᵈ letter

91

Αγαπητή κ. Σοφία, πιστεύω πως με αυτές τις δραστηριότητες, τα παιδιά εξασκούνται στο μάθημα των Αγγλικών και με αυτόν τον τρόπο γίνονται πιο κατανοητά.

Εντύπωση μου έκανε που διορθώνει ο ένας τα λάθη του άλλου και οι ερωτήσεις που κάνει ο ένας στον άλλο.

Επίσης μου άρεσε η μαγνητοφώνηση των παιδιών με κάποιες ερωτήσεις που κάνει κ.λ.π. Αυτό άρεσε πολύ και στο ίδιο το παιδί, γιατί άκουσε την προφορά της στην γλώσσα των Αγγλικών.

Mrs Demenaga/ Bessy's mother

Appendix VII

The guidelines for the Implementation of Portfolio

Τι είναι το portfolio;

Είναι η συλλογή συγκεκριμένων εργασιών και δραστηριοτήτων που θα κάνουμε στα Αγγλικά.

Γιατί θα κάνεις το portfolio;

Για να δείξεις την πρόοδό σου:
- τι μπορείς να πεις στα Αγγλικά
- τι μπορείς να γράψεις στα Αγγλικά
- τι μπορείς να διαβάσεις στα Αγγλικά
- τι μπορείς να καταλάβεις από κάτι που άκουσες στα Αγγλικά

Τι θα βάλεις στο portfolio;

- Τις υποχρεωτικές εργασίες
- Ερωτηματολόγια που θα αξιολογείς τον εαυτό σου ή θα διατυπώνεις τις σκέψεις σου για κάθε μια από τις υποχρεωτικές εργασίες.
- Σχόλια για τις εργασίες ορισμένων συμμαθητών σου.
- Σχόλια για τα βιβλία που θα διαβάσεις: "My Reading Log".
- Ασκήσεις που εσύ θα θελήσεις να βάλεις στο portfolio: "My Choice".
- Σχόλια που θα δικαιολογούν γιατί διάλεξες να βάλεις κάποιες ασκήσεις στο portfolio.
- Περιεχόμενα
- Μια παράγραφο που θα γράψουμε στο τέλος και θα διατυπώσεις τις σκέψεις σου για το portfolio (Cover Letter).
- Ένα ερωτηματολόγιο που θα αξιολογείς όλο το portfolio.
- Ένα ερωτηματολόγιο που θα συμπληρώσεις για να δείξεις τι έμαθες κάνοντας το portfolio.

Appendix VIII

Assessment checklist for copying (writing mechanics) skills for the 2nd entry

1	Has produced legible handwriting	
2	Has copied the format of the song correctly	
3	Has left gaps between words	
4	Has left gaps between stanzas	
5	Writing is properly aligned	
6	Has not forgotten any words	
7	Has not forgotten any letters	
8	Has no spelling mistakes	
9	Has used capital letters where necessary	
10	Has copied punctuation marks correctly	
	• full-stops	
	• commas	
	• question marks	
	• exclamation marks	

Appendix IX

Assessing speaking global rating scale for the 1st entry

EXCELLENT	• Carried out the task successfully and with relative ease • Very good pronunciation/ intonation€ • Fairly easy to understand • Very few pauses
VERY GOOD	• Carried out the task successfully • Pronunciation slightly influenced by L1 • General meaning fairly clear • Not many pauses
GOOD BUT CAN DO BETTER	• Carried out the task with some difficulty • Pronunciation influenced by L1 • Meaning is understood with some effort • Longer pauses to search for words
TRY HARDER	• Had great difficulty carrying out the task. • A lot of serious pronunciation / intonation errors • Almost impossible to understand • A lot of unnaturally long pauses

Appendix X

The cover letter

122

NAME: Anastasia Koutsandrea DATE: 13/6/06

MY PORTFOLIO

Το portfolio ήταν μια πολύ ωραία
ιδέα τη κυρίας. Μου άρεσε που γράψαμε
τα κασέτα, που γράψαμε την έκθεση,
που ζωγραφίσαμε το τεραζάκι που
πέξαμε με τα ζωάκια κ.ά. Θέλω
να συνεχίσω και του χρόνου το
portfolio και σε άλλα μαθήματα.
Το portfolio μου άρεσε γιατί κάναμε
πολλές δραστηριότητες.

NAME: Demi Kanuela DATE: 13/06/2006

MY PORTFOLIO

Μου άρεσε πολύ που γράψαμε το
τραγούδι στην κασέτα. Δεν κουράστηκα
καθόλου. Θέλω να πάρω κι σε άλλο
μάθημα το portfolio.

164

The viability of alternative assessment methods in the Greek upper secondary school: the oral portfolio

[Η βιωσιμότητα εναλλακτικών μεθόδων αξιολόγησης στο Ελληνικό λύκειο: ο φάκελος αξιολόγησης για τον προφορικό λόγο]

Angeliki Daphni

The final examination of the English language subject, in the context of the Greek state upper secondary education, is a traditional paper-and-pencil test which does not include any assessment of oracy skills. This article explores the viability of the oral portfolio as an alternative assessment and pedagogic method that can facilitate the assessment of speaking and listening skills and create a more motivating learning environment. To this effect, three methodological tools were designed, namely, a questionnaire addressing upper secondary English teachers in Greek state schools, a case study involving an oral portfolio implementation and finally, a questionnaire for students to record their experience. The study demonstrates that implementation of the portfolio contributed to a successful assessment of oracy skills and that it was a stimulating experience for students. The results of the study also showed that the pedagogical value of the portfolio counterbalanced its practical constraints. The paper concludes by putting forward recommendations for the future application of this assessment technique in state school education.

છ

Η τελική αξιολόγηση του μαθήματος της Αγγλικής γλώσσας, στα πλαίσια της Ελληνικής δημόσιας ανώτερης δευτεροβάθμιας εκπαίδευσης, διεξάγεται μέσω μίας παραδοσιακής γραπτής δοκιμής η οποία δεν περιλαμβάνει καμία αξιολόγηση του προφορικού λόγου. Το άρθρο αυτό διερευνά τη βιωσιμότητα της αξιολόγησης του προφορικού λόγου βάσει φακέλου ως μία εναλλακτική μέθοδο αξιολόγησης και διαπαιδαγώγησης που μπορεί να διευκολύνει τη διαδικασία αξιολόγησης της παραγωγής και κατανόησης προφορικού λόγου και των ακουστικών δεξιοτήτων και να δημιουργήσει περισσότερα κίνητρα για μάθηση. Για το σκοπό αυτό, σχεδιάστηκαν τρία μεθοδολογικά εργαλεία και συγκεκριμένα, ένα ερωτηματολόγιο απευθυνόμενο σε καθηγητές Αγγλικής γλώσσας της ανώτερης δευτεροβάθμιας εκπαίδευσης, μία περιπτωσιολογική μελέτη που αφορούσε στην εφαρμογή ενός φακέλου αξιολόγησης με αντικείμενο τον προφορικό λόγο, και τέλος, ένα ερωτηματολόγιο που δόθηκε στους μαθητές για να καταγράψουν την εμπειρία τους. Η έρευνα καταδεικνύει ότι η εφαρμογή του φακέλου μπορεί να συμβάλλει στην επιτυχή αξιολόγηση του προφορικού λόγου και αποτελεί μία ενθαρρυντική εμπειρία για τους

μαθητές λυκείου. Τα αποτελέσματα επίσης κατέδειξαν ότι η παιδαγωγική αξία του φακέλου εξισορρόπησε τους πρακτικούς της περιορισμούς. Το άρθρο τελειώνει κάνοντας υποδείξεις για τη μελλοντική εφαρμογή αυτής της τεχνικής στα δημόσια σχολεία.

Key words: alternative assessment, traditional paper-and-pencil tests, oral portfolio, oral assessment, oracy skills.

Introduction

Recently, there has been a shift in students' assessment towards alternative assessment methods. This shift is of great importance in education and, in particular, it has beneficial implications for the teaching of English as a Foreign Language (EFL) learning in the Greek upper secondary school. To be promoted to the next class in the present context, students need to take a traditional paper-and-pencil test at the end of the school year. The format of the test is specified by a Presidential Decree (60/2006/F.E.K 65 F.E.K A'). This achievement test excludes the assessment of oracy skills, i.e., listening and speaking. The current study attempts to explore the viability of the oral portfolio as a representative alternative assessment and pedagogic method. It is argued that the portfolio is an efficient and stimulating assessment tool of oracy skills.

The paper begins with a discussion of the concerns that have been raised about traditional testing in contrast to the favourable learning conditions that alternative assessment, and, more specifically, portfolio, creates. It presents the teaching context of the study and its methodological instruments, namely, a questionnaire addressing upper secondary EFL teachers, a case study, and a students' questionnaire. The final section discusses issues in the implementation of the portfolio and provides practical guidelines for its implementation.

Testing and Alternative Assessment

'Testing' and 'alternative assessment' techniques have been both used to assess Foreign Language Learning (FLL). In the Greek state educational context, 'testing' refers to official, standardized measurement procedures, administered on a 'one-off' basis at specified times of a school year, which take the form of paper-and-pencil tests. On the other hand, 'alternative assessment' models collect assessment evidence with informal procedures over a period of time under no time constraints, which makes them low-stakes in terms of consequences (Alderson and Banerjee, 2001, p. 228). Moreover, alternative assessment results are qualitative and take the form of a profile. This is much more useful and informative compared to a single test score as it provides details concerning each student's growth, attitude and current needs. Furthermore, alternative assessment offers opportunities to evaluate the effectiveness of instruction (Stiggins, 2001, p. 468). In this way, alternative assessment serves a diagnostic and 'formative' function, i.e., it involves an ongoing process of gathering information on the extent of learning and on strengths and weaknessess, which teachers can feed back into their course planning and into the actual feedback they give to learners (Council of Europe, 2001, p. 186). In addition, McNamara (2000, p. 3) argues that tests are conceived to catch people out on what they do not know and consequently they can affect the psychological state of students as they are associated with feelings of anxiety and powerlessness which impede learning. On the contrary, alternative assessment is influenced by assumptions of humanistic education that

"education should deal with both dimensions of humans – the cognitive or intellectual and the affective or emotional" (Moskowitz, 1978, p. 18).

The next section discusses the characteristics of portfolios as an alternative assessment method.

Portfolio

"Portfolios in classrooms today are derived from the visual and performing arts tradition in which they serve to showcase artists' accomplishments and personally favored works" (Sweet, 1993, p. 1). Brown and Hudson (2002, p. 81) define the portfolio as a 'purposeful' collection of any aspect of a student's work that displays the efforts, skills, abilities, achievements, and contributions to a given class. Paulson et al. (1991) support that portfolios permit instruction and assessment to be woven together in a way that more traditional approaches do not. For Genesee and Upshur (1996, p. 99) the primary value of the portfolio assessment is that it can reveal students' achievement seen as improvement and accomplishment against both an earlier point of development, and a goal that provides a target and a direction. Noticing improvement when comparing work over time motivates students (Stiggins, 2001, p. 469). For Murphy (1994), portfolio assessment relieves students who suffer from 'test anxiety' of a 'test-like situation', and allows the assessment process to be done as a 'normal part of the classroom routine'. Similarly, Reineke (1998, pp. 83-84) argues that the portfolio is sensitive to students. According to Brown and Hudson (2002, pp. 81-82), the portfolio can:

- build and capitalize on the actual work done in class;
- increase student responsibility for the learning processes;
- encourage collaboration between teachers and students and change the teacher's role from adversary to coach.

Finally, Venn (2000, p. 538) argues that portfolio assessment enables teachers and students to share the responsibility for setting learning goals and for evaluating progress towards meeting these goals. The most comprehensive and widely reported initiative of large-scale portfolio applications is the European Language Portfolio.

Motivated by the literature reviewed above, the present student set out to investigate the viability of portfolios as an alternative and pedagogic method in Greek state schools.

Teaching context

English as a school subject in the Greek senior high school curriculum has a very low status for two main factors which have set a barrier against students' motivation (Pedagogical Institute, 2011):

- It is not connected to any high-stake language examination or certificate that would have a direct effect on students' professional or educational life. This tempts parents and students to trust FLL to private language centres.
- It does not contribute to the total grade that determines university entrance.

Although the national curriculum for the teaching of EFL in the senior high school (Government Gazette – F.E.K 3994/11-10-99) recognizes that learners need to be able to produce spoken language and emphasizes the instruction of EFL, the format of the final

achievement test, which is specified by the Presidential Decree 60/2006 /F.E.K 65 F.E.K A', assesses only writing and reading skills. There is no current official test requirement that assesses oracy skills, i.e., listening and speaking.

This mismatch between curriculum objectives and tests is likely to lead students to study only whatever is on the tests (stressed also by Brown and Hudson 2002, p. 48). Consequently, since the focus of the final achievement test for the English language in the senior high school is on reading and writing, students are not given any incentive to practise their oracy skills, which has a negative effect on the development of comprehension and production of realistic spoken language. Moreover, it has negative consequences for teachers, as they have to adapt their teaching practices to the objectives of the final achievement test and focus only on skills that are included in it.

Research design and research questions

The current study aims to explore alternative ways to assess the oracy skills of senior high school students through the use of an oral portfolio. The aim of the study was approached empirically through the adoption of three instruments:
- a questionnaire designed to portray the beliefs, experiences and attitudes of upper secondary EFL teachers towards assessment principles and techniques;
- a case study that involved the observation of a senior high school English class where an oral assessment portfolio was implemented for a period of time;
- a questionnaire given to the students of the teaching group to assess their experience.

The research questions of the study were as follows:
1. Can the portfolio assess oracy skills effectively?
2. Is the oral portfolio an efficient pedagogic tool?
3. Can the oral portfolio create a more motivating FLL classroom in the Greek state senior high school?

The teacher questionnaire

The teacher questionnaire (Appendix I) contained nine questions. It was designed to target the teachers':
   a) opinions concerning the parameters of effective assessment;
   b) assessment practices;
   c) personal perceptions and preferences concerning oral assessment;
   d) attitudes towards portfolio assessment.

Question 1 collected the teachers' biodata. The teacher sample included 22 EFL teachers of senior high schools in Athens, capital of Greece, as well as in Cyclades, Evia and Arcadia, provinces of Greece. Among the respondents there was only one male teacher. Moreover, there were 5 holders of Masters Degrees and 2 teachers who had attended several teaching seminars. The respondents' teaching experience in years was as follows:
- 1-8 years: 5 respondents;
- 9-17 years: 5 respondents;
- 18-25 years: 6 respondents;
- 25-35 years: 6 respondents.

Question 2 involved the teachers' opinions concerning the parameters of effective assessment and consisted of four sub-questions which made enquiries about four different

aspects of assessment: a. recipients of assessment, b. purposes of assessment, c. agents of assessment, and d. marking procedures. The types of responses given were scale items that consisted of alternative responses representing degrees of importance. According to the findings of sub-question a (see Table 1), 67% of the respondents considered that it was very important for effective assessment to provide feedback to the teacher and the learner, while 47% of the respondents considered that effective assessment was important for instructional procedures.

a. Effective assessment provides feedback for the…

Recipients of assessment	1	2	3	4	5
a. teacher	67%	28%	5%	0%	0%
b. learner	67%	16.5%	16.5%	0%	0%
c. instructional procedures	47%	23.5%	23.5%	6%	0%

(1 = very important; 5 = not at all important)

Table 1. Recipients of assessment

Table 2 demonstrates the respondents' beliefs concerning the purposes of assessment. Identifying the specific needs of individual students and monitoring the effectiveness of instruction were considered very important purposes by 68.5% and 63% of the respondents respectively, while assessing and understanding students' performance in class was considered very important by 53% of the respondents. Keeping students alert was considered very important by 44% of the respondents. However, 11% of the respondents considered that this purpose was not important at all. The other two purposes of assessment were considered very important by fewer respondents. In particular, placing students into levels of ability and decisions about advancement or promotion were considered very important purposes of assessment by 23,5% and 21% of the respondents respectively.

b. The purpose of effective assessment is to...

Purposes of assessment	1	2	3	4	5
a. assess and understand students' performance in class	53%	21%	26%	0%	0%
b. identify the specific needs of individual students	68.5%	26%	5%	0%	0%
c. monitor the effectiveness of instruction	63%	16%	10.5%	10.5%	0%
d. place students into levels of language ability	23.5%	12%	41%	23.5%	0%
e. make decisions about advancement or promotion of individual students to the next level of instruction	21%	37%	21%	21%	0%
f. keep students alert	44%	17%	17%	11%	11%

Table 2. Purposes of assessment

As shown in Table 3, the primary agent of assessment should be the teacher as this agent was considered to be very important by the highest percentage of respondents (83%). Students were considered very important agents by 44,5% of the respondents, while a governmental body or institution was considered to be very important by 17% of the respondents only.

c.   The agent of effective assessment is…

Agents	1	2	3	4	5
a. the teacher	83%	17%	0%	0%	0%
b. a governmental body or institution	17%	11%	17%	27.5%	27.5%
c. the student	44.5%	44.5%	11%	0%	0%

Table 3. Agents of assessment

Table 4 shows the findings of sub-question 'd' which was concerned with marking procedures. According to the results, 53% of the respondents believed that students' profiles were a very important marking procedure. Classroom performance was regarded as very important by 42.5% of the respondents, while a general impression was believed to be very important by 29% of the respondents. No respondent considered performance on tests to be a very important criterion of awarding marks.

d.   Marks should be awarded on the basis of…

Marking procedures	1	2	3	4	5
a. a general impression	29%	41%	6%	6%	18%
b. performance on tests	0%	39%	50%	11%	0%
c. classroom performance	42.5%	31.5%	21%	5%	0%
d. Students' profiles that describe their performance at a range of different levels and in different areas	53%	29%	12%	6%	0%

Table 4. Marking procedures

The above results highlighted the EFL teachers' belief that traditional paper-and-pencil tests should not be the primary form of assessment. The results showed that assessment should: 1. provide feedback for the teacher, the learner and the instructional procedures, 2. identify the students' needs and monitor the effectiveness of instruction without focusing only on categorization and promotion, 3. take into consideration students' opinions concerning assessment processes and practices, and 4. award marks based on students' profiles.

Question 3 investigated the kind of assessment practices that were currently in use. This is a closed question that provides a list of assessment methods and requires respondents to report how frequently they use each one of them. Respondents had to select among four degrees of frequency: never (N), rarely (R), often (O), and nearly always (NA). As displayed in Table 5, all respondents had used traditional tests. Also, there were small percentages of respondents who had never used alternative assessment techniques such as, self-reports (9,5%), portfolio (14,5%) and peer appraisal (20%).

Assessment methods	N	R	O	NA
1. Traditional pencil-and-paper tests	0%	24%	28%	48%
2. Learner self-reports (self-appraisal, diary, record-keeping)	9,5%	24%	52%	14,5%
3. Portfolio	14,5%	33%	38%	14,5%
4. Peer appraisal	20%	45%	25%	10%

Table 5. Assessment methods currently used

Questions 4, 5 and 6 shifted the focus of the questionnaire to teachers' perceptions and preferences concerning effective oral assessment, which were expected to provide useful information for the development of the oral portfolio of the current study. Question 4 is closed and required respondents to judge the idea of assessing students' performance on subjectively-scored tasks, such as speaking and writing tasks, based on criteria rather than on a general impression of the students' performance, as 'very useful', 'OK', or 'not useful'. All respondents considered this idea to be useful. This is very important as in the context of EFL teaching in the Greek educational system, no official assessment criteria have been set for the assessment of speaking and writing.

Question 5 probed the most widely accepted criteria for students' oral performance. It required ranked responses as it asked respondents to rank their preferences from a given list of oral criteria. Respondents placed the criteria in the following order (see also Table 6): 1. ability to get the message across, 2. effort to interact, 3. fluency, 4. accuracy of language use and range of vocabulary, and 5. pronunciation. Overall, the results show that communicative criteria such as 'ability to get the message across' or 'effort to interact' were regarded more important than structural criteria, such as 'accuracy' and 'pronunciation'.

Oral assessment criteria	1	2	3	4	5
Accuracy of language use and range of vocabulary	14%	7%	14%	51%	14%
Pronunciation	0%	7%	7%	14%	72%
Fluency	0%	36%	28%	36%	0%
Ability to get the message across	72%	28%	0%	0%	0%
Effort to interact	22%	28%	50%	0%	0%

(1= the most important … 5 = the least important)

Table 6. Oral assessment criteria

Table 7 demonstrates the findings of question 6 that contains statements concerning issues of marking procedures, objectives, and techniques of oral assessment and invites respondents to state their agreement or disagreement with them or to take a neutral position. According to the results, 52% of the respondents believed that assessment criteria should be known to candidates and 67% believed that they should vary according to the type of the speaking task. For 71% of the respondents speaking tasks should promote social interaction. Finally, nearly three quarters of the respondents were favourable to the use of the portfolio as an assessment method of the speaking skill.

Statements	Agree	Neutral	Disagree
1. Students who sit a speaking test should know the criteria used.	52%	48%	0%
2. The marking criteria should vary according to the type of the speaking task.	67%	24%	9%
3. Speaking tasks should promote social interaction and assess students in pairs.	71%	24%	5%
4. Portfolio is a good method to assess the speaking skill.	71%	24%	5%

Table 7. Oral assessment marking procedures, objectives and techniques

Question 7 aimed to confirm the accuracy of the results obtained so far. It required respondents to provide ranked responses by answering whether they agreed 'a lot', 'quite a lot', 'a little' or 'not at all' with three statements. The findings, displayed in Table 8, demonstrated that EFL teachers were critical of traditional paper-and pencil tests, supported

that oracy skills had to be assessed officially, and considered that the portfolio could create motivating learning conditions.

Statements	A lot	Quite a lot	A little	Not at all
1. Traditional pencil-and-paper tests judge students' learning without assisting it.	24%	57%	14%	5%
2. The absence of the speaking skill from the official format of the final achievement test does not encourage teachers to teach it extensively and students to practise it.	57%	33%	5%	5%
3. The use of portfolio as an innovative instruction and assessment method can raise students' motivation towards the English language subject in the senior high school.	38%	62%	0%	0%

Table 8. Summary of results

Questions 8 and 9 explored various portfolio assessment issues. Question 8 consists of an initial closed question which required respondents to tick either 'Yes' or 'No' depending on whether they had ever used portfolio assessment systematically. This question was answered positively by only 7 respondents out of 22. These respondents had received further training since their graduation.

Question 8 is also followed up by two open questions. The first open question required from those respondents who had answered positively to the previous question to refer to the benefits of this type of assessment. The respondents reported that the portfolio promotes autonomy, self-fulfillment and greater involvement, and provides evidence of progress on an on-going nature. The second open question consists of two parts and required the rest of the respondents to refer to factors that a. could encourage them to use portfolio assessment and b. impede the use of the portfolio. The factors that could encourage teachers to use portfolio in their classrooms were: adequate portfolio knowledge, fewer students and more teaching time. Factors which impeded the use of the portfolio were: lack of portfolio knowledge, lack of students' interest, lack of time and material resources, the heavy workload that the portfolio entails, and mixed ability classes.

Question 9 consists of scale response items. This question addressed only teachers with portfolio experience and required them to state how strongly they agree or disagree with the existence of some advantages and drawbacks of the portfolio that have been reported in literature in the field. The data gathered (see Table 9) confirmed the beliefs held by many educational experts concerning the potential of the portfolio to provide evidence on students' abilities, attitudes and developmental processes, as well as diagnostic information about corrective action. Moreover, 67% strongly disagreed with the statement that the portfolio assessment provides totally unreliable results and 50% agreed that the portfolio involves a time-consuming and costly procedure.

Statements	1	2	3	4	5
1. Portfolios carry an optimal amount of information about students' abilities.	58.5%	33%	8.5%	0%	0%
2. Portfolios reflect students' attitudes and developmental processes that take place in their learning over time.	58.5%	33%	8.5%	0%	0%
3. Portfolios provide diagnostic information about corrective action to be taken by students and for	41.5%	50%	8.5%	0%	0%

	1	2	3	4	5
instruction by revealing students' improvement or lack of it over time.					
4. Portfolios help students get engaged in their learning and make them more autonomous.	50%	33%	8.5%	8.5%	0%
5. Portfolio assessment provides totally unreliable results for students' performances.	0%	8.5%	8.5%	16%	67%
6. Portfolio involves a time-consuming and costly procedure.	8%	8.5%	50%	8.5%	25%

(1 = strongly agree …. 5 = strongly disagree)

Table 9. Advantages and drawbacks of the portfolio

Overall, the analysis of the questionnaire data provides evidence of the teachers' belief that oracy skills should be assessed using an alternative assessment method such as portfolios. The next section presents the implementation of an oral portfolio for the purposes of a case study according to the findings of the current questionnaire.

Case study: a portfolio implementation

The case study involved the implementation of an oral portfolio (5 month duration) in the first grade of the senior high school located in Dionysus, a suburb of Athens, capital of Greece during the academic year 2006-2007. The participants were 18 students. They were all sixteen years old. The syllabus was based on the coursebook 'Get real 2' by Helbling languages (Hobbs & Keddle, 2006), which targeted a B2 proficiency level on the Common European Framework of Reference (CEFR; Council of Europe, 2001).

During the implementation period observations were conducted to provide empirical data on the effectiveness of the oral portfolio not only as an assessment but also as a pedagogic tool. The intention here was that, apart from assessing, portfolios were used as a way to facilitate English language learning, to motivate learners, to activate reflection on learning processes, and to promote self-assessment and autonomy. Moreover, the portfolio aimed to gain information about the oral portfolio practicalities so as to guide teachers towards the inclusion of this technique in their own classrooms.

Implementation of the portfolio

The portfolio took the form of a plastic folder. It contained:
- an audio-cassette with four obligatory speaking core tasks;
- an audio-cassette with four optional speaking tasks;
- documents such as: a teacher's assessment report and a student's self-assessment report for each core task, and a student's reflective report for each optional task.

Successful presentation of the portfolio entailed that it contained all the recorded tasks and paper work, and was worth 20% of the total portfolio grade. The portfolio was given two points that were added to the students' first semester grade.

To implement the portfolio in class, several steps were taken. To begin with, in early September 2006, the school principal and students were informed of a new, student-friendly assessment method. Students were given written guidelines to clarify the teaching goals, format, content and deadline of the portfolio. A preparation period followed to familiarize students with spoken tasks and to practise self-assessment skills. Four core tasks were performed. They were intended to make the assessment process a learning process and

involved pairs of students randomly formed to enhance meaning negotiation. The teacher set up the task, ensured that students knew how to proceed and then recorded students' performance. Students listened to their recorded performance at home and produced a list of errors to serve as a basis for individualized instruction and to elicit self-correction.

All core tasks followed the principles of communicative language tests, which *"are intended to be a measure of how the testees are able to use language in real life situations"* and reflect the communicative situations in which testees are likely to find themselves in the future (Kitao & Kitao, 1996). The tasks displayed many of the communicative features identified by Nation (1989, pp. 24-29), such as split information, steps, assignment of roles, and existence of a goal and an outcome.

Core tasks were 'graded' i.e., presented in an order that permitted demands on the learners to increase gradually. Also, skills assessed in prior activities were deemed necessary for succeeding ones. Appendix II presents the fourth task of the portfolio, which involves story-telling based on picture stimuli that create an information gap leading to interaction. Task 4 was the last task the testees had to perform because it was considered to be the most difficult as, according to Ellis (2003, p. 206), tasks requiring construction of interpretations of visual stimuli generate great complexity.

Core tasks were marked through an assessment form (Appendix III) that was filled in immediately after each performance. The assessment form was a 'rating scale'. This is a short description of different levels of language ability that aims to describe what the learner can do at each level and to help the assessor decide what score to give (Underhill, 1987, p. 98). The rating scales of the portfolio were 'analytic', i.e., they contained five separate categories of criteria, and separated out three levels of language ability to encourage the assessor to give a number of scores, which makes scoring more reliable (Hughes, 1989, p. 94). The assessment form provided the necessary metalanguage for giving feedback to the students in the form of a profile that indicated areas of srength and weakness.

Optionally, students could record a personal oral performance of their choice, such as a monologue or an intercultural experience. Optional tasks satisfied the teacher's need for information about students' performance outside classtime, and gave students the possibility to maintain ownership and responsibility of their work. Students with four optional tasks gained an extra point to the total portfolio grade, but there was no marking penalty if students did not manage to record any.

Students were also required to fill in a self-assessment report for each core task (Appendix IV) and a second report for each optional task (Appendix V). Self-assessment motivated learners to transform weaknesses into learning goals and encouraged them to reflect on their progress and the quality of their performance in relation to known criteria. Additionally, learners reported their feelings while making use of higher-order thinking skills, such as analysis and observation. These are lifelong learning skills that enhance autonomy.

A teacher-student, ten-minute, private conference was conducted after each core task to discuss the list of errors and reports establishing a good teacher-student rapport and promoting joint goal-setting and negotiation of grades. Conferences provided students with supportive comments that helped them recognize and enjoy their accomplishments allowing them to develop positive self-images. These are strategies that increase learners' satisfaction and maintain their motivation (Dörnyei, 2001).

During the conference students were asked the following questions:
- Did you like your performance?
- In what ways have you improved?
- What was most difficult for you?
- What did you do when you couldn't find the appropriate word / expression to use?
- Did you do anything to help your partner when he / she couldn't continue?
- Name three things you learned about …
- What do you think you can improve in your next performance?
- What do you think of your teacher's assessment report? Have you been fairly judged? What are the areas you disagree with?

A final conference followed after all the portfolio tasks had been performed to discuss optional tasks and their reports and also check on the portfolio contents. Students were asked some of the following questions:

- What kinds of spoken performances have you included in your portfolio?
- Choose one task you are most proud of and say why …
- Do you solve problems the same way you did earlier in the year? How do you solve them now?
- What would you like to do next, using what you learned from this task?

A follow-up event was organized at the end of the portfolio period. This involved a competition for the most interesting optional task and an award of certificates to offer praise for effort.

The student questionnaire

The student questionnaire (Appendix VI) aimed to shed light into the students' own voice about the portfolio impact. Students were invited to fill in the questionnaire anonymously during classroom time after the portfolio period had been completed. Data were collected by 16 students.

Question 1 looked at the qualities of portfolio. The results indicated that 56% of the students had become more interested in the English language, 69% realized their strengths, 88% realized their weaknesses, 69% focused their study on their weak points, and 63% gained more confidence in their speaking skills. However, the portfolio helped only 37% of the students to become well-organized and only 31% to become more responsible. Question 2 required students to compare and contrast portfolio with traditional tests. Students expressed preference towards portfolio assessment when considering that it provided easier (81%), fairer (57%), and more interesting assessment (75%), while traditional tests were regarded as less useful (81%), less helpful (56%), and more stressful (75%). Questions 3 to 7 considered specific components of the portfolio. Concerning the core tasks (Question 3), students found them interesting and guiding (94%), not stressful (63%), and not tiring (69%). However, students expressed their dissatisfaction towards performing the tasks in public, which implied that it was face-threatening.

The impact of self-assessment and conferences, which both constituted new experiences for students, was dealt with by questions 4 and 5 respectively. In particular, the majority of the students regarded self-assessment as fair (94%) and believed that it made them think about their strengths and weaknesses (81%). However, 56% admitted that it did not motivate them to study more at home. Regarding conferences, they were considered pleasant by 94%,

interesting by 75%, and helpful by 69%. Conferences were seen as a chance for better co-operation with the teacher by 94%. Furthermore, although conferences were tiring for only 19%, they were stressful for 63%.

Regarding the impact of the optional tasks (Question 6), 69% considered them a challenging experience and 81% a good opportunity to show what they could do outside classroom. Although only eight students (50%) stated that recording was difficult, only four students out of sixteen recorded optional tasks. These students admitted that they had enjoyed the recording procedure.

According to the results of question 7, neither the content nor the form of the reports complicated students, as 75% found them guiding and 81% found them easy to fill in. Finally, 87% answered positively when asked whether they would like to be assessed with a portfolio again.This was very encouraging as the oral portfolio was a novice and brief experience for the specific teaching group.

Conclusion and recommendations

In conclusion, the findings of the study indicate positive changes in the students' attitude towards English as the oral portfolio can provide the driving force to sustain their active involvement in the often tedious learning process. Nevertheless, it would be useful to attempt to highlight some implementation pitfalls as well as some recommendations for teachers who are willing to experiment with new assessment techniques in accordance with their professional judgment.

Assessment is a rather unpleasant and stressful experience whatever form it may take. Portfolios are demanding and may be seen as threatening and confusing or even as another awkward and arduous trial. The students of the current teaching group seemed more comfortable with the familiar true-false test format and disliked challenging activities that entailed creativity and original work, like optional tasks, or that required more extensive responses, like reflective reports. This finding is in agreement with Nunan (1989) who suggested that learners show a preference for teacher-centred over learner-centred structures. Therefore, initial motivation has to be generated. This can be done by increasing students' expectancy of success. To maintain and protect motivation, students have to be presented with stimulating, enjoyable, and relevant oral tasks (Dörnyei, 2001). However, deadlines have to be set and enforced, while students need to be convinced that the tasks are worthwhile (Ellis, 2005, p. 25). As Foster (1998) reports, if students fail to take tasks seriously, they will view them as 'games' and eschew meaning negotiation because it detracts from the 'fun' element. Finally, although portfolio assessment aims to create a relaxing and non-judgmental setting, assessment standards should remain high.

Portfolios place extra logistic demands. Specifically, they are uneconomical in terms of material resources, such as tape recorders, folders, cassettes and photocopies, as well as of the time required to be constructed, administered and marked. Moreover, scheduling tasks in pairs and individual conferences for large classes may interfere with other instructional activities. Additionally, administering oral tasks and conferences during classtime complicates teachers as they have to maintain order while attending to the students' performance. Finally, students' portfolios require storage space.

Portfolios have to be fully integrated into the curriculum alongside traditional tests and teaching materials so as not to be seen as separate from learning and to be accepted as a

formal assessment method. Moreover, an e-portfolio platform designed by the Ministry of Education with recorded or videoed samples of students' speech would greatly attract and excite upper secondary students. At the same time, an e-portfolio would abolish the ephemeral nature of spoken discourse, as it would provide easy access to students' speech, and enable them to notice their progress as well as evoke a critical evaluation of each others' skills at various times of the school year. Finally, an e-portfolio requires minimal storage space and increases capabilities in using technology to support lifelong learning.

Finally, EFL teachers are obviously concerned with improving their assessment methods but appear to lack the opportunity, time and resources to revise and update their assessment approaches. Therefore, sufficient, in-service assessment training has to be planned and provided by the state.

Author's email: adafni@sch.gr

References

Alderson, J. C. & Banerjee, J. (2001). 'Language testing and assessment (Part 1).' *Language Teaching,* 34/4: 213-216.

Brown, J. B. & T. Hudson (2002). *Criterion-referenced language testing.* Cambridge: Cambridge University Press.

Council of Europe (2001). *Common European framework of reference for languages: learning, teaching, assessment.* Strasbourg: Council for Cultural Co-operation, Education Committee. Cambridge: Cambridge University Press.

Dörnyei, Z. (2001). *Motivational strategies in the language classroom.* Cambridge: Cambridge Univerity Press.

Ellis, R. (2003). *Task-based Language learning and teaching.* Oxford: Oxford University Press.

Ellis. R. (2005). *Instructed second language acquisition: a literature review.* New Zealand: Research Division, Ministry of Education.

Foster, P. (1998). 'A classroom perspective on the negotiation of meaning.' *Applied Linguistics,* 19, 1-23.

Genesee, F. & Upshur, J. A. (1996). *Classroom-based evaluation in second language education.* Cambridge: Cambridge University Press.

Government Gazette-FEK 3994 (1999). 'The Curriculum for the Teaching of the English Language in the Unified Lyceum', 24163-24192.

Hobbs, M. & Keddle, J. S. (2006). *Get real.* Helbling Languages.

Hughes, A. (1989). *Testing for language learning.* Cambridge: Cambridge University Press.

Kitao S. K. & Kitao, K. (1996). 'Testing communicative competence'. *The Internet TESL Journal*, 11/5, accessed at http://iteslj.org/Articles/Kitao-Testing.html on 6 Nov 2011.

McNamara, T. (2000). *Language testing.*Oxford: Oxford University Press.

Moskowitz, G. (1978). *Caring and sharing in the foreign language class.* Rowley, MA: Newbury House.

Murphy, S. (1994). 'Writing portfolios in K-12 Schools: implications for linguistically diverse students.' In L. Black, D. Daiker, J. Sommers & G. Stygall (Eds.), *New directions in portfolio assessment*. Portsmouth, NH: Boynton/Cook & Heinemann, 140-156.

Nation, P. (1989). 'Speaking activities. five features.' *ELT Journal,* 43/1: 24-29.

Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge: Cambridge University press.

Paulson, F. L., Paulson, P. R., & Meyer, A. C. (1991). 'What makes a portfolio a portfolio?' *Educational Leadership*, 48: 60-63.

Pedagogical Institute (2011). 'Unified programme of studies for foreign languges'. Accessed at http://rcel.enl.uoa.gr/xenesglosses/sps.htm on 6 Nov 2011.

Reineke, A. R. (1998). *Challenging the mind, touching the heart.* California: Corwin Press.

Stiggins, J. R. (2001). *Student-involved classroom assessment*. New Jersey: Prentice Hall.

Sweet, D. (1993). 'Student portfolios: classroom uses.' *Education Research Consumer Guide*, 8: 1-5.

Underhill, N. (1987). *Testing spoken language.* Cambridge: Cambridge University Press.

Venn, J. J. (2000). *Assessing students with special needs.* Upper Saddle River, NJ: Merrill.

Appendix I

Questionnaire for Teachers

1. Biodata
a. Sex: Male ☐
      Female ☐
b. Teaching experience: 1-8 years ☐   9-17 years ☐
          18-25 years ☐   25-35 or more ☐
c. Current position of work: _____
d. Further training: PEK ☐
          MAs ☐
          Other (please specify) _____

2. Based on your experience, state your ideas about what an effective assessment policy should be like by placing a tick or a cross under the appropriate column for each statement.
(1 = very important … 5 = not at all important)

a. Effective assessment provides feedback for the…

Recipients of assessment	1	2	3	4	5
a. teacher					
b. learner					
c. instructional procedures					

Other (please specify) _____

b. The purpose of effective assessment is to…

Purposes of assessment	1	2	3	4	5
a. assess and understand students' performance in class					
b. identify the specific needs of individual students					
c. monitor the effectiveness of instruction					
d. place students into levels of language ability					
e. make decisions about advancement or promotion of individual students to the next level of instruction					
f. keep students alert					

Other (please specify) _____

c. The agent of effective assessment is…

Agents	1	2	3	4	5
a. the teacher					
b. a governmental body or institution					
c. the student					

Other (please specify) _____

d. Marks should be awarded on the basis of…

Marking procedure	1	2	3	4	5
a. a general impression					
b. performance on tests					
c. classroom performance					
d. Students' profiles that describe their performance at a range of different levels and in different areas					

Other (please specify) _____

3. Consider the assessment methods that you use and place a tick or a cross under the appropriate column to indicate how frequently you use each one.
(N =never, R = rarely, O = often, NA = nearly always)

Assessment methods	N	R	O	NA
1. Traditional pencil-and-paper tests				
2. Learner self-reports (self-appraisal, diary, record-keeping)				
3. The building of students' profiles of abilities (portfolio)				
4. Peer appraisal				

Other (Please specify) _____

4. How do you find the idea of assessing students' performance on tasks for which there is not a single correct answer, such as speaking and writing tasks, based on criteria rather than on a general impression of the students' performance?
(Please tick)
- Very helpful   ☐
- OK   ☐
- Not helpful   ☐

5. What should the marking criteria for a speaking task be? Answer the question by numbering the following in the order you prefer.
(1 = the most important   ….   5 = the least important)
- Accuracy of language use and range of vocabulary   ☐
- Pronunciation   ☐
- Fluency   ☐
- Ability to get the message across   ☐
- Effort to interact   ☐
- Other (please specify) _____

6. Do you agree with these statements? Put a tick in the column that shows your response.

Statements	Agree	Neutral	Disagree
1. Students who sit a speaking test should know the criteria used			
2. The marking criteria should vary according to the type of the speaking task			
3. Speaking tasks should promote social interaction and assess students in pairs			
4. Portfolio is a good method to assess the speaking skill.			

7. Consider the following statements and indicate the degree of your agreement by circling the relevant number.
1. Traditional pencil-and-paper tests judge students' learning without assisting it.

A lot	quite a lot	a little	not at all
4	3	2	1

2. The absence of the speaking skill from the official format of the final achievement test   does not encourage teachers to teach it extensively and students to practise it.

A lot	quite a lot	a little	not at all
4	3	2	1

3. The use of Portfolio as an innovative instruction and assessment method can raise students' motivation towards the English language subject in the senior high school.

A lot	quite a lot	a little	not at all
4	3	2	1

8. Have you ever used portfolio assessment in your classroom / with your students?

Yes ☐

No ☐

If Yes,

What do you consider the benefits of portfolio assessment?	

If No,

a. What would encourage you to use portfolio assessment in your teaching?	
b. What factors impede the use of assessment portfolio in your teaching?	

9. If you have ever had any portfolio experience, consider the following statements and indicate the degree of your agreement by putting a tick or a cross under the appropriate column.
(1 = strongly agree   ….   5 = strongly disagree)

Statements	1	2	3	4	5
1. Portfolios carry an optimal amount of information about students' abilities					
2. Portfolios reflect students' attitudes and developmental processes that take place in their learning over time					
3. Portfolios provide diagnostic information about corrective action to be taken by students and for instruction by revealing students' improvement or lack of it over time					
4. Portfolios help students get engaged in their learning and make them more autonomous					
5. Portfolio assessment provides totally unreliable results for students' performances					
6. Portfolio involves a time-consuming and costly procedure					

Thank you for your cooperation!

Appendix II

Date: ……………………………………………
Task 4: Story-telling

Student A

> **Part A**
> You are going to tell the story of Mr. Pea, a very careless driver. Look at picture A   and tell your partner what Mr. Pea did in the morning, while he was driving to work. You have to speak for one minute. You start first.
>
> When you finish, your partner will tell you what Mr. Pea did in the afternoon, when he arrived home.

> **Part B**
> Work in pairs. Look at Picture C and decide about what happened to Mr. Pea on another day. Construct a story about it. Include information on the following:
>
> When and where
>
> Who he was with
> What he did
>
> What happened next
>
> The ending
>
> Make your story as interesting as possible. You have three minutes. Remember that you have to speak in English at all times.

> **Part C**
> Narrate the second half of your story to your teacher. You have to agree with your partner at which point he / she stops and you take over. You have to speak for one minute.

Picture A

Picture C

Date: ……………………………………………….

Task 4: Story-telling

Student B

Part A

You are going to tell the story of Mr. Pea, a very careless driver. Look at picture B and tell your partner what Mr. Pea did in the afternoon, when he arrived at home. You have to speak for one minute. You start second.

Before you start, your partner will tell you what Mr. Pea did in the morning, while he was driving to work.

Part B

Work in pairs. Look at Picture C and decide about what happened to Mr. Pea on another day. Construct a story about it. Include information on the following:

When and where

Who he was with
What he did

What happened next

The ending

Make your story as interesting as possible. You have three minutes. Remember that you have to speak in English at all times.

Part C

Narrate the first half of your story to your teacher. You have to agree with your partner at which point you stop and he / she takes over. You have to speak for one minute.

Picture B

Picture C

Appendix III

Assessment Report

Student's name :……………………………..…Date: ………………….

Criteria	Rating scale			Mark
	1.5-2	0.5-1.5	…-0.5	
Pronunciation	Correct pronunciation and appropriate intonation.	Pronunciation slips and inappropriate intonation occasionally.	Incomprehensible pronunciation and colourless intonation impede communication.	
	4-6	2-4	…-2	
Task achievement	Provides a credible interpretation of the picture, contributes to the construction of the story, narrates half of the story successfully, tries to make the story sound as interesting as possible, knows where to stop / start, produces two long turns lasting approximately one minute each.	Provides an interpretation of the picture with some support, limited contribution to the construction of the story, narrates half of the story with some support, the story produced is rather dull, has to be reminded when to stop / start, produces two long turns lasting approximately half a minute each.	Fails to provide an adequate interpretation of the picture even with a lot of support, does not contribute to the construction of the story, fails to narrate even a small part of the story although a lot of support is provided, produces only short turns of limited length.	
	3-4	1.5-3	…-1.5	
Fluency	Can speak coherently and clearly with few intrusive hesitations.	Speaks hesitantly with pauses.	Fragmentary speech and incomplete sentences impede communication.	
Accuracy	A fairly wide range of structures and vocabulary, errors minimal in number and gravity, communication of the message is achieved.	Limited vocabulary, structures with little variety, frequent errors that do not prevent communication of the essential message.	Use of only basic vocabulary and structures that may render communication of the message difficult.	
Interactive communication	Initiates discussion, listens to his/her partner's contribution, prompts his/her partner, is co-operative and polite.	Difficulties in turn-taking, limited evidence of co-operation, is not always polite.	Monopolizes the dialogue, no turn-taking, no signs of co-operation, impolite.	
Total				

Appendix IV

Task 4
Student's name:……………………………………………………………………….

Date:…………………………………………………………..

	Self-assessment Report	**Circle your score** (1=not good, 2=good, 3=very good, 4= excellent)
1	My pronunciation was correct and my intonation was appropriate.	1 2 3 4
2	My interpretation of the picture was credible, I contributed to the construction of the story equally, I narrated my part of the story successfully, my story was quite interesting.	1 2 3 4
3	I spoke clearly without pauses.	1 2 3 4
4	I used a variety of expressions correctly and made few grammar errors	1 2 3 4
5	While constructing the story with my partner, I spoke when I had to, listened to him / her and tried to help him / her understand me when I spoke.	1 2 3 4
Total		

- How do I feel about my performance on this task?
- Have I ever narrated a story before?
- What were my strong points?
- What were my weak points?
- What should I try to improve in the future?

Appendix V

Optional Task Report

Dear Teacher,

What you are going to listen to is a / an_____

_____

_____.

When I recorded it, I was alone / with _____

at _____

_____.

I selected to record this type of speaking experience because _____

_____

_____.

This recording shows that I can _____

_____

_____.

I feel that my performance was _____

_____

_____.

In the future, I need to improve _____

_____

_____.

Appendix VI

Dear Student,
Now that you have completed your portfolio, I would like you to answer the following questions anonymously. Please, be sincere as your answers will help me to improve our lessons.

1. (Please tick)

Has your portfolio helped you to…	Yes	No
1. realize your strengths?		
2. realize your weaknesses?		
3. focus your study on your weak points?		
4. gain more confidence in your speaking skills?		
5. become well-organised?		
6. become more responsible about your study and learning?		
7. become more interested in the English language lesson?		

Other _____

2. (Please tick)

			Yes	No
	more difficult			
	more interesting			
The portfolio is	fairer	than a traditional test.		
	more useful			
	less stressful			
	more helpful			

Other _____

3. (Please tick)

The core tasks	Yes	No
1. should be conducted in front of the whole class.		
2. were guiding and interesting.		
3. were a stressful experience.		
4. were a tiring experience.		

Other _____

4. (Please tick)

Having the opportunity to self-assess your performance	Yes	No
1. is fair.		
2. motivates you to study more.		
3. makes you think about your strengths and weaknesses.		

Other _____

5. (Please tick)

Were the conferences	Yes	No
1. pleasant?		
2. stressful?		
3. interesting?		
4. tiring?		
5. helpful?		
6. a chance for better co-operation with your teacher?		

6. (Please tick)

The optional tasks	Yes	No
1. were a good opportunity to show your teacher what you can do outside the classroom.		
2. were difficult to record.		
3. were a challenging experience.		
4. were difficult to design.		

7. (Please tick)

The self-assessment and optional tasks reports were	Yes	No
1. guiding.		
2. easy to fill in.		

Finally, would you like to be assessed with a portfolio again? (Please tick)

Yes ……..   No ……..

Thank you for your co-operation!

Your Teacher

Prospects of Using the European Language Portfolio as Pedagogical and Assessment Tool in Greek Schools

[Προοπτικές του Ευρωπαϊκού Portfolio Γλωσσών ως Εργαλείο Μάθησης και Αξιολόγησης στο Ελληνικό Σχολείο]

Eirini Bompolou

This paper focuses on the implementation of the European Language Portfolio (ELP), a document that plays a central role in the Council of Europe's language policy. More specifically, this paper investigates if the systematic use of the Greek version of the ELP, in the foreign language classroom could benefit the learners from a pedagogical point of view and at the same time serve as an assessment tool for the EFL teacher. A small scale study was conducted using questionnaires and interviews to explore the attitudes of Greek foreign language teachers and students. The paper discusses the benefits that the ELP has to offer to both teachers and learners and the steps that need to be followed to put the ELP into practice and to disseminate its use. The paper concludes that the introduction of the ELP in the Greek educational system could bring radical changes to the way students are taught and evaluated.

೫

Η μελέτη εστιάζει στην εφαρμογή του Ευρωπαϊκού Portfolio Γλωσσών, το οποίο είναι ένα έγγραφο καθοριστικής σημασίας για την γλωσσική πολιτική του Συμβουλίου της Ευρώπης. Πιο συγκεκριμένα, ερευνά αν η συστηματική χρήση της ελληνικής απόδοσης του Ευρωπαϊκού Portfolio Γλωσσών, στη διδασκαλία των ξένων γλωσσών θα μπορούσε να ωφελήσει τους μαθητές από παιδαγωγική άποψη και ταυτόχρονα να χρησιμεύσει ως εργαλείο αξιολόγησης για τους καθηγητές Αγγλικής ως ξένη γλώσσα. Μια έρευνα μικρής κλίμακας διενεργήθηκε με τη χρήση ερωτηματολογίων και συνεντεύξεων για να εξετάσει τη στάση των καθηγητών και των μαθητών. Η μελέτη συζητά τα οφέλη του Ευρωπαϊκού Portfolio Γλωσσών για τους μαθητές και τους καθηγητές και τα βήματα που πρέπει να ακολουθηθούν ώστε να εφαρμοστεί και να εξαπλωθεί η χρήση του. Η μελέτη καταλήγει στο ότι η εισαγωγή Ευρωπαϊκού Portfolio Γλωσσών στο ελληνικό εκπαιδευτικό σύστημα θα μπορούσε να φέρει ριζικές αλλαγές στον τρόπο που διδάσκονται και αξιολογούνται οι μαθητές.

Introduction

The Council of Europe (CoE) has introduced two inter-related reference instruments in order to "promote international cooperation in the field of modern languages" (Council of Europe, 2001, p.1): 'Common European Framework of Reference for Languages: Learning, Teaching and Assessment' (CEFR; Council of Europe, 2001) and the European Language Portfolio (ELP; Council of Europe, 2004). The former provides illustrative descriptors and defines levels of proficiency in a coherent and transparent manner, so that the learners' language qualifications can be compared with those of other learners of the language in other parts of the world. The latter is a document in which people who learn or have learned a foreign language can record and reflect on their language learning and cultural experiences.

History and current status of the ELP

The idea of the ELP emerged in 1991, in a Council of Europe's intergovernmental symposium entitled "Transparency and Coherence in Language Learning in Europe: Objectives, Evaluation and Certification" (Centre of Language Teaching and Research, 2002; also in North, 1992; North, 2000; Sheils, 1996). From 1998 to 2000 the ELP went through a pilot phase in which educators from fifteen countries and from all kinds of educational settings took part on a voluntary basis (Little & Perclová, 2001; Schärer, 2000; Ushioda and Ridley, 2002). Finally, in 2001, on the occasion of the European Year of Languages, the CEFR and the ELP were officially launched (Schärer, 2000, p.13). By the end of 2010, the Validation Committee[1] had accredited 113 ELP models from more than thirty different countries and organisations.

In Greece, there are two validated models. The "European Language Portfolio for learners aged 12 to 15 in Greece" (accreditation number 43.2003)[2] and the newest one entitled "Model for Primary Education (young Learners aged 9 to 12)" (accreditation number 110.2010). Both ELP models have undergone a pilot phase with encouraging results (see Kaga-Giovoussoglou, undated; Kaga, 2010).

Presentation of the ELP

The term "portfolio", evokes the idea of an "artist's portfolio" where the artists select and display their "best work" (Little & Perclová, 2001). A language teacher may think of different types of portfolios such as writing portfolios in composition classes (Baak, 1997), working portfolios, display portfolios, and assessment portfolios (Danielson and Abrutyn, 1997), portfolios for assessment and learning purposes (Klenowski, 2000), teaching portfolios (Bastidas, 1996), portfolios providing a framework for a process writing course (Rea, 2001) and even electronic or digital portfolios (Ali, 2005; Woodward and Nanlohy, 2004). However, the ELP is an innovative tool, which is far removed from any of the portfolios described above because of its three-part structure. According to the "Principles and Guidelines" (Council of Europe, 2004) valid ELPs are made up of the *"Language Passport"*, the *"Language Biography"* and the *"Dossier"*. The *"Dossier"* is the only part of the ELP that resembles the portfolios used in L2 teaching. In the *"Dossier"* the learners select and put samples of their own work in order to provide evidence for their achievements and illustrate their profile as it is sketched in the other two parts of the ELP (Council of Europe, 2004).

The *"Language Passport"* is a standard component of all ELP models, apart from the ones addressing very young learners, which promotes pan-European recognition of the ELP and facilitates student mobility (Council of Europe 2004). In the Language Passport a student's

"overall L2 proficiency is periodically summarised against the self-assessment grid from the CEFR (see Council of Europe 2001; Council of Europe, 2004; Little 2009). In addition the learners record their formal qualifications such as exam results or language certificates and illustrate any significant intercultural experiences. It is evident that the Language Passport is meant to be updated regularly by the learners (Council of Europe 2004).

Finally, the *"Language Biography",* the most extensive component of the ELP, is the place where learners use checklists in the form of grids to assess their knowledge and set goals for the future. From a pedagogic point of view, the Language Biography has a "pivotal" role since it provides "a focus for the reflective processes that mediate between the Language Passport and the Dossier" (Little & Perclová, 2001, p.2).

The ELP helps "to motivate learners by acknowledging their efforts to extend and diversify their language skills at all levels" and provides "a record of the linguistic and cultural skills they have acquired" (Council of Europe, undated). In other words it fulfils both a *pedagogical* function and a *reporting* function (Little and Perclová, 2001). In this respect, the three components of the ELP are complementary since its *pedagogical* function is fulfilled by the Language Passport and the Dossier while its *reporting* function is largely fulfilled by the Language Biography and the Dossier (Council of Europe, 2004; Ushioda & Ridley, 2002).

Basic principles governing the ELP

Plurilingualism is an underlying language competence which can be transferred to the learning of new languages. Interculturality, on the other hand, means "knowledge of the Otherness" (Sheils, 1996, p. 93). The notions of "plurilingualism" and "interculturality" are central to the ELP where intercultural experiences are considered as integral part of the learning process. Apart from that, it is essential for the learners that the ELP is recognisable worldwide and it is valued highly in as many countries as possible. For this reason all ELP models, alongside with the elements addressing the particular teaching context, should retain some common core elements otherwise the reporting function of the ELP will have no value (Schärer, 2000).

A core value of the ELP is self-assessment. According to Nunan (1988, p.116), "in a learner-centred curriculum model both teachers and learners need to be involved in evaluation". The self-evaluation, however, should not be random, but based on certain criteria, and should be performed under the guidance of an educator (Little 2004, p. 14). This is also advocated by the CEFR, (Council of Europe, 2001) which states that self-assessment is more accurate if it is related to clear descriptors of language proficiency.

Self-assessment in the ELP is both formative and summative (Little and Perclová, 2001). Formative assessment takes place in the Language Biography, where the learners evaluate their present knowledge in five communicative skills using the self assessment grid of the CEFR and setting goals for the future: *listening, reading, spoken interaction, spoken production, writing*. In the Dossier, where learners choose material that, according to them, can demonstrate their knowledge, assessment is both formative and summative. Finally in the Language Passport, where they fill in a self-assessment grid, after the end of a teaching cycle, estimating to which degree they have achieved their goals, assessment is summative (Little, 2002).

The main benefit of this criterion-reference system as opposed to traditional norm-referenced examinations is that when learners set goals and then estimate how much they

have progressed in achieving them, even the weakest learners will feel proud of their progress and will gradually manage to reach their goals (Little and Perclová, 2001). When students try to find out if they meet certain criteria, they gradually become autonomous by learning not only to evaluate their learning, but also to plan and monitor it more effectively (Nunan, 1988, p.116). The earlier the students are exposed to autonomous learning the less they will resist it (Little, 1991). To show this in practice, Dam (1995) redesigned her teaching approach to offer her learners the opportunity to become more autonomous and she demonstrated that in their effort to achieve the goals they had set and to understand why, what and how they learn her learners became more and more autonomous.

In the case of the ELP, the criteria the learners have to apply are represented by the language proficiency descriptors that are included in each ELP model. By trying to reach these criteria and by becoming conscious of the process followed to achieve their goals the learners who use the ELP gradually become more autonomous.

Methodology of the study

Although there is a general agreement (Little, 2009; Kohonen, 2001; Stoicheva et al., 2009) that the ELP is an effective tool for the teaching of foreign languages, research studies need to confirm that it is applicable to local systems. This is the focus of a small-scale study reported here (for more details, see Bompolou, 2007). The main research question of the study was: "Does the use of the Greek ELP have a positive impact on the teaching of foreign languages in the Greek educational system?"

To answer this question, some supplementary questions emerged:
- Is the "ELP for learners aged 12 to 15 in Greece" a suitable document? Is it compatible with the Greek curriculum?
- Are the Greek teachers and learners willing to work with the ELP?
- What course of action should be followed and what obstacles should be overcome in order to make the ELP part of the daily teaching at schools?

At first, it was investigated whether the first ELP model was compatible with the Greek 'Cross-Thematic Curriculum Framework for Modern Foreign Languages' and the Individual Subject Curriculum (ISC) (Pedagogical Institute, 2003a; 2003b). It was found that the ELP and the Greek Curriculum, are based on the same principles, as both seek among other things: *"to provide access to life-long learning"* and *"to assist the development of European citizenship awareness, while preserving national identity and cultural awareness"* (Pedagogical Institute, 2003a, p. 11). The fact that both documents are governed by the same principles is more evident in the ISC for Modern Foreign Languages where the focus is not only on "foreign language literacy" but also on "multilingualism and multiculturalism". Moreover, the ability to use and acquire "skills and abilities necessary for lifelong learning like the ability to 'learn how to learn' are also valued. Regarding assessment, the ISC instructs teachers to opt for alternative assessment methods such as "portfolio assessment" and "self-assessment" along with more traditional ones (2003b, p. 381).

The questionnaire devised in order to elicit the teachers' attitudes towards the ELP and its suitability for use in the Greek educational system was distributed to a sample of fifty-four (54) foreign language teachers (47 women and 7 men) who came from different parts of Greece. The questionnaire included scanned extracts from the three parts of the "European Language Portfolio for learners aged 12 to 15 in Greece" (model 43.2003), as the degree of

the teachers' familiarity with the document was not known. The pages from the ELP which were scanned and included in the questionnaire are presented in the following table:

Component of the ELP	Pages
Language Passport	5, 9,10,15,16,19,20 (my numbering)
Language Biography	6,7,9,10,11,37
Dossier	1 (my numbering)

Table 1. Pages from the ELP model 43.2003 included in the questionnaire

Six learners (three boys and three girls), from different levels in the Greek educational system, were also interviewed. Their age ranged from 11 to 18 and they all came from Athens, the capital of Greece. The sample included one learner from primary school and four students from secondary education (two from junior high school, one from senior high school and one from vocational high school).

The learners were at first asked some questions concerning their attitudes towards language learning and the way they had been taught English and French at school. To understand what the ELP is, the students were shown either the Greek ELP for secondary education, or the Irish Junior Version of the ELP (70.2006), depending on the educational sector they belonged to.

The interviews were semi-structured (see Appendix 1) and took the form of an informal discussion in Greek with each of the students. To avoid intimidating them and to make it easier for them to express themselves. The interviews were audiotaped but not transcribed and the analysis of the students' opinions was based on notes taken during the interviews as well as on the audiotaped material. These interviews were not meant to give an account of the attitudes of all Greek students towards the use of the ELP in their classroom; they aimed at sketching the first impressions of a sample of students who examine an ELP for the first time and recording their reactions.

Finally, the teachers and learners' opinions that were expressed in the two Greek pilot project reports (see Kaga, undated; Kaga, 2010) and the final report on the pilot project (Schärer, 2000, see also Little and Perclová, 2001) were also compared to the responses of participants.

Results

Greek teachers' stance towards the ELP

Teachers and students agreed that the use of the ELP in the Greek classrooms is not only applicable, but would also be most welcome. The teachers were asked to fill in anonymous questionnaires. In the first part of these questionnaires they were asked to give personal data concerning their teaching situation and teaching experience (Question 1 and Question 2). The results analysed in Table 2 show us that the randomly selected sample includes teachers who represent all kinds of educational settings.

As far as their experience is concerned most of the teachers (42) had been teaching for five to fifteen years at the time they answered the questionnaire; three teachers had less than five years of teaching experience and nine were very experienced with more than fifteen years of teaching.

Sector	Number of teachers
Primary only	6
Junior High School only	9
Senior High School only	17
Vocational High School only	6
More than one sector	13
Administrative Position	3

Table 2. Educational sector of the teachers

Before being presented with parts of the ELP, the teachers were asked if they had received any information about it (Question 3). Thirty-three of the teachers answered positively and twenty-one negatively. The teachers who said that they were familiar with the ELP were asked to describe it briefly. Their descriptions were very precise proving that teachers keep in pace with the new developments.

Teachers were then asked if they considered the Greek ELP for secondary education appropriate for their teaching situation (Question 4). Only fifteen teachers in our survey responded negatively. The most frequent reasons for not considering it suitable are that they did not know anything about the ELP, that "*they have not received any training in using it*" and that "*the level of their students is very low*", followed by the argument that "*their learners are not used to self-evaluation*". This argument, however, could be contradicted by the fact that learner autonomy is a gradual process and time and effort is needed on behalf of the teachers to train their students on how to become more autonomous (Little et al, 2002). Although the teachers did not consider the portfolio appropriate for their teaching situation, all of them responded that it is "*useful tool for the learners*" (Question 5) and "*useful tool for the teachers*", (Question 7) *w*hereas the European average is 70% and 78% respectively (Schärer, 2000:10).

Regarding the main benefits stemming from the use of the ELP for the learners (Question 6) the most popular answer was: "*it enables students to learn how to learn*" (all of the teachers responded positively) followed by "*it promotes the use of the target language in this process of reflection*", "*it enables students to monitor their learning*" and "*it helps learners become more autonomous*" which were selected by fifty-one Greek teachers. The least popular statement was that the ELP "*can facilitate student mobility*" as eighteen teachers disagreed with it, twenty-four agreed and only twelve agreed strongly.

As far as the benefits for the teachers are concerned (Question 8), all of the teachers agreed that the ELP "*helps the teachers focus on different aspects of the language*" and that by using it "*teachers can be more creative*". Fifty-two teachers agreed with the statements that the ELP "*enables teachers to monitor the progress of the individual learners*" and that "*it is flexible for the teachers*".

The great majority of the teachers (44) believed that the particular ELP is a suitable document for Junior High School students (Question 9) and agreed that a similar ELP should be introduced to other levels of the Greek educational system (Question 11) (fifty-two teachers gave positive answers). Surprisingly, only eighteen teachers (and only three among the twelve teachers who work in the primary education) replied that a version of the ELP should be introduced in the primary school. The results, however, would probably be different if the teachers were also shown extracts from the new version of the ELP for primary education because the version shown is too complicated for very young learners. In general, forty-five teachers wanted a similar ELP to be introduced to the senior high school,

thirty-three teachers wanted one for the vocational high school and twenty-seven for the university (Question 12).

Question 10 asked teachers about the ability of the Greek Junior High School students to assess themselves. Only nine of the teachers believed that the Greek learners who are between 12 and 15 years old are not mature enough to evaluate themselves. This either implies that these teachers thought that the learners of the lower secondary education were much more mature than those who still attend primary education, or is again explained by the fact that the ELP that was included in the questionnaire addresses secondary education learners.

As far as the "Dossier" is concerned, the vast majority of the teachers (48) agreed that it should include extracts illustrating the students' ability to produce written language and thirty teachers believed that it should include samples of the students' oral production. Finally, ten teachers suggested "other" things such as projects, creative work or anything "*that the learner him/herself would feel is worthy of keeping or is proud of*" (Question 13).

Question 14 focused on the innovations that the ELP introduced. According to the teachers, the most important innovation, selected by thirty-eight teachers is that the ELP is "*taking account in a positive way of all learning regardless of whether gained in or outside of formal education*" followed closely by "*the central role of self-assessment*" and "*the development of self-directed learning and learner autonomy in a life-long perspective*" (selected by thirty-six teachers). "*Its transnational dimension, which provides Europe-wide transparency and comparability*" was selected by only twenty-three teachers. Finally, teachers agreed that the ELP could facilitate student mobility (Question 15).

Greek learners' attitude towards the ELP

As mentioned earlier, six students who learn English as a foreign language in different aducational sectors in Greek schools were interviewed. These students were not familiar with the ELP. In order to solve this problem, I showed each student a photocopied version of an Irish ELP addressing children (No. 70.2006) or the Greek ELP for learners 12 to 15 depending on their age. The analysis of the students' answers concerning language learning and the way they were used to be taught showed that most of them were not very happy about the way they were taught foreign languages at school and that their teachers followed a more or less teacher-centred approach, although they sometimes allowed some student initiative.

The students' answers concerning the ELP were very interesting. All of them were impressed by it and said that they would like to have their own. In the words of a senior high school student the ELP "*shows exactly what we can do with the foreign language*". The learners of French who had taken part in the Greek pilot project reported by Little (2002) and Kaga-Giovoussoglou (undated) expressed exactly the same positive attitude with the learners interviewed for this study.

As far as the contents of the Dossier are concerned, all students said they would include projects or written texts, while the ones attending senior and vocational high school would also include foreign language certificates. Finally, the learners unanimously asserted that they would keep and update their ELP regularly after finishing the educational sector that they attended. The older learners also felt that it would be very useful for them if they wanted to study abroad and that it would help them find a job in Greece or abroad.

Summary of the results

The ELP is a tool which is recognised in all countries of the CoE and offers many benefits to both foreign language teachers and learners. This small scale study recorded the attitudes of a number of learners and teachers who claimed that its implementation would be beneficial for them. More specifically, participants of the study agreed that the ELP is a very useful tool for both teachers and learners. They also agreed that it would increase the learners' motivation for two reasons. The first reason related to the pedagogical function of the ELP: it could make low ability learners feel less neglected as it gives every student a sense of achievement (Little et al, 2002). The second reason was that its reporting function would enhance the learners' extrinsic motivation because it would improve their employment prospects, in all the countries of the CoE.

Most learners believed that using the ELP would be a pleasant experience for them, which means that their intrinsic motivation would be enhanced. Teachers also agreed that they would be more creative and that the ELP would be a good motive for their students. This is highlighted by Little et al. (2002) too, for example "*By making learners responsible for their own learning, we challenge them to fuel their learning with the intrinsic motivation that underpins their out-of-school activities*" (ibid, p. 17).

Suggestions and Conclusion

In the past few years, two ELP models have been created in Greece. School advisors in several parts of Greece have organised seminars to introduce the ELP to teachers of primary education (Kaga, 2010). The scope of the seminars however, should be broadened to include teachers of secondary education because a very limited number of secondary school teachers have attended seminars concerning the use of the ELP. It would be useful if the Ministry of Education started a teacher-training program in which the use of ELP would be demonstrated by teachers who have already used it. At the same time teachers should have the opportunity to use it in their classroom and share potential problems they encounter with other participants in the program (Little, 2006).

Moreover, short and long-term evaluation projects need to be conducted, to investigate the impact of the ELP on teachers and learners and to amend the mistakes that may occur at the beginning of its implementation. As far as students are concerned, the organisation of an ELP competition that could also lead to an exhibition granting a kind of prize to the best Portfolios, would provide an additional motive for the learners to work harder (Ushioda and Ridley, 2002).

 In spite of all the benefits of the introduction of the ELP to the Greek educational system, there will be many obstacles to overcome. It should be realised that its introduction needs time even if the Ministry of Education solves the problem of covering the financial cost of the ELP copies and the teacher training that is demanded. However, according to the results of this study, its systematic implementation in the Greek education could offer a rewarding experience for both foreign language teachers and learners.

Author's email: mepr4@yahoo.gr

Notes

1. The Validation Committee is a sub-committee of the Steering Committee for Education of the Council of Europe which was responsible to assure the conformity of European Language Portfolio models to the common European Principles and Guidelines.
2. This means that it was accredited in 2003 and it was the 43rd model to be recognised.

References

Ali, S.Y. (2005). 'An Introduction to electronic portfolios in the language classroom'. *The Internet TESL Journal*, 11/8.

Baak, E. (1997). 'Portfolio development: an introduction'. *Forum,* 35/2.

Bastidas, A. J. (1996). 'The teaching portfolio: a tool to become a reflective teachers'. *Forum,* 34/3.

Bompolou, E. (2007). 'Implementation of the European Language Portfolio in the Greek Secondary Education'. Unpublished M.Ed Dissertation. Patras: Hellenic Open University.

Centre of Language Teaching and Research (2002). *"Language Passport" – "Learner Passport" – "Language Portfolio": the story of an idea.* Feriburg Centre of Language Teaching and Research.

CILT (2006) *My languages portfolio – ELP junior version: revised edition*. London: CILT.

Council of Europe (2001). *Common European framework of reference for languages: learning, teaching, assessment.* Cambridge: Cambridge University Press.

Council of Europe (2003). *European Language Portfolio for learners aged 12 to 15 in Greece.* Athens: Ministry of Education.

Council of Europe (2004). *European Language Portfolio (ELP): principles and guidelines with added explanatory notes (version 1.0)*. Strasbourg: Council of Europe, Language Policy Division.

Council of Europe. (2010). *European Language Portfolio for primary schools (for pupils aged 9 to 12).* Athens: Ministry of Education.

Council of Europe. (undated). *Guide to compiling an ELP*. Accessed at http://www.coe.int/t/dg4/education/elp/elp-reg/Source/Guidetocompile/Guide_to_compiling_ELP_EN.pdf on 30 Nov 2011.

Dam, L. (1995). *Learner Autonomy 3: from theory to classroom practice.* Dublin: Authentik.

Danielson, C. & Abrutyn, L. (1997). *An introduction to using portfolios in the classroom.* Alexandria: Association for Supervision and Curriculum Development.

Kaga-Giovoussoglou, E. (undated). 'The ELP pilot project in Greece'. In D. Little (Ed.), *The European Language Portfolio in use: nine examples*. Accessed at http://www.coe.int/t/DG4/Portfolio/documents/ELP%20in%20use.pdf on 1 Dec 2011.

Kaga, E (2010). *Emploi du Portfolio Européen des langues à l' échelle de l' établishment scolaire - troisiéme rapport.* Accessed at http://elp-wsu.ecml.at/LinkClick.aspx?fileticket=1mSR5of0x7c%3d&tabid=1637&language=en-GB on 13 May 2011.

Klenowski, V. (2000). 'Portfolios: promoting teaching'. *Assessment in Education,* 7/2: 215-236.

Kohonen, V. (2001). *Developing the European language portfolio as a pedagogical instrument for advancing student autonomy*. Accessed at https://www.uta.fi/laitokset/okl/tokl/projektit/eks/pdf/kohonen2001.pdf on 11 Dec 2011.

Little, D. (1991). *Learner autonomy 1: definitions, issues and problems.* Dublin: Authentik.

Little, D. (2002). *European Language Portfolio: a seminar sponsored by MIUR-Directorate General for International Relations Directorate General for Piedmont Pininfarina Schoo*l.

Turin: Italy. Accessed at www.coe.int/t/DG4/Portfolio/main_pages/Turin_Report1rev.doc on 2 May 2011.

Little, D. (2004). *European Language Portfolio: Council of Europe seminar sponsored by the Ministry of Education, Spain 30 September – 2 October 2004* DGIV/EDU-LANG (2004) 14. Accessed at http://www.coe.int/Portfolio on 4 April 2011.

Little, D (2009). *The European Language Portfolio: where pedagogy and assessment meet.* 8th International Sminar on the European Language Portfolio, Graz, 29 September-1 October 2009. DGIV EDU LANG (2009) 19. Accessed at http://www.coe.int/t/dg4/education/elp/elp-reg/Source/Publications/ELP_pedagogy_assessment_Little_EN.pdf on 11 Feb 2012.

Little D, Goullier F., Margonis-Pasinetti R. & Andronescu E. R. (undated). *The European Language Portfolio: A guide to the planning, implementation and evaluation of whole-school projects.* Accessed at http://elp-wsu.ecml.at/LinkClick.aspx?fileticket=tdAJOhiWZ9I%3D&tabid=2374&language=en-GB on 5 Jun 2011.

Little, D. & Perklová, R. (2001). *The European Language Portfolio: a guide for teachers and teacher trainers.* Strasbourg: Council of Europe.

Little, D., Ridley, J. & Ushioda, E., (2002). *Towards Greater Autonomy in the Foreign Language Classroom*. Dublin: Authentik.

North, B. (Ed) (1992). *Transparency and Coherence in Language Learning in Europe: Objectives, Assessment and Certification*. Symposium held in Róschlikon, 10-16 Nov 1991. Strasbourg: Council for Cultural Co-operation.

North, B. (2000). 'Defining a flexible common measurement scale: descriptors for self and teacher assessment'. In G. Ekbatani & H. Pierson (Eds.), *Learner-directed assessment in ESL.* Mahwah, New Jersey: Laurence Erlbaum Associates, 13-27.

Nunan, D. (1988). *The learner-centred curriculum*. London: Cambridge University Press.

Pedagogical Institute (2003a). *Cross-Thematic Curriculum Framework for Compulsory Education* (translated from the Official Gazette issue B nr 303/13-03-03 and issue B nr 304/13-03-03). Accessed at http://www.pi-schools.gr/programs/depps/index_eng.php on 29 Apr 2011.

Pedagogical Institute (2003b). *Αναλυτικό Πρόγραμμα Σπουδών Αγγλικής Γλώσσας.* Accessed at http://www.pi-schools.gr/lessons/english/pdf/14depps_XenonGlosson-Agglika.pdf on 29 Apr 2011.

Rea, S. (2001). 'Portfolios and process writing: a practical approach' *The Internet TESL Journal,* 7/6.

Schärer, R. (2000). *A European Language Portfolio: final report on the pilot project (1998-2000)*. Strasbourg: Council of Europe.

Sheils, J. (1996). 'The Council of Europe and language learning for European citizenship'. *Evaluation and Research in Education*. 10/2 & 3: 88-103.

Stoicheva, M. Hughes, G. & Speitz, H. (2009) *The European Language Portfolio: an impact study*. Strasbourg: Council of Europe.

Ushioda, E. & Ridley, J. (2002). 'Working with the European Language Portfolio in Irish post-primary schools: report on an evaluation project'. *CLCS Occasional Paper*, 61.

Woodward, H. and Nanlohy, P. (2004). 'Digital portfolios in pre-service teacher education'. *Assessment in Education,* 11/2: 167-178.

APPENDIX

SEMI-STRUCTURED INTERVIEWS WITH STUDENTS

BEFORE PRESENTING THE ELP TO THE STUDENTS:
1. Do you like learning foreign languages?
2. What do you specially like (dislike) about the way you are taught?
3. Did your last language teacher ever ask you to work in pairs or groups?
4. Did your last teacher ever used extra materials or just the coursebook?
5. Did your teacher ever ask for your opinion about what to do in the lesson or how you would like to learn?

AFTER PRESENTING THE ELP TO THE STUDENTS:
6. Have you ever seen such a document before?
7. Imagine that your teacher told you that you would start using this document from next year. What are your feelings about it?
     A) Something new.
     B) It seems interesting.
     C) It won't make any difference.
     D) Extra work
8. If you started by the dossier, what would you like to include in it?
9. Take a look at the self-assessment grid. Do you feel able to assess yourself in English? Try to do so for your favourite skill.
10. The ELP will help you become more autonomous in the sense that you can now set your own goals and try to achieve them. How does this make you feel?
     A) I'm very happy about it.
     B) I don't know.
     C) I feel scared about it.

11. Will you show your portfolio to your parents?
12. Would you show it to your future employers?
13. Will you keep your portfolio after finishing (primary school, high school etc)?
     - Is it OK if part of this interview will be shown to other teachers?
     - Do you have any questions yourselves?

Thank you very much for taking part in this interview!

Portfolio Assessment of Speaking Skills in English as a Foreign Language in Primary Education

[Αξιολόγηση μέσω Φακέλου Μαθητή των Δεξιοτήτων Προφορικού Λόγου στα Αγγλικά ως Ξένη Γλώσσα στην Πρωτοβάθμια Εκπαίδευση]

Georgia Efthymiou

This study focuses on the assessment of speaking skills with reference to young learners. This is achieved by using an alternative method of assessment, namely portfolios. The general aim is to introduce learners' to portfolio assessment of their speaking skills and to promote further learning and autonomy making, thus, learning and assessment coexist in a non-threatening mode. Three methodological tools are used for this research; a needs analysis questionnaire addressing the pupils' needs of the fifth grade of a Greek primary school, the European Portfolio of Languages (ELP) - used in tandem with the oral portfolio Dossier - and a final evaluation questionnaire given to the pupils after the completion of the oral portfolio project. Based on the statistical analysis of pupils' evaluation results and the teacher's observation throughout the school year, it is evident that the oral portfolio denotes a time-consuming and laborious assessment process. Nevertheless, the pupils see it as an interesting experience and are willing to use it again in the future. In conclusion, students' portfolios are an innovative method of assessment that can actually promote the development of speaking skills and young learners' metacognitive strategies in the EFL classroom, and raise their interest in learning.

CB

Η παρούσα μελέτη επικεντρώνεται στην αξιολόγηση των προφορικών δεξιοτήτων με αναφορά στους νεαρούς μαθητές. Αυτό επιτυγχάνεται χρησιμοποιώντας μία εναλλακτική μέθοδο αξιολόγησης, συγκεκριμένα τους φακέλους μαθητών (portfolios). Ο γενικός στόχος είναι η εισαγωγή των μαθητών στην αξιολόγηση των προφορικών τους δεξιοτήτων μέσω του πορτφόλιο και η προώθηση της περαιτέρω μάθησης και αυτονομίας κάνοντας έτσι την μάθηση και την αξιολόγηση να συνυπάρχουν με έναν μη απειλητικό τρόπο. Τρία μεθοδολογικά εργαλεία χρησιμοποιούνται για αυτή την έρευνα: ένα ερωτηματολόγιο ανάλυσης αναγκών που απευθύνεται στις ανάγκες των μαθητών της πέμπτης τάξης ενός Ελληνικού δημοτικού σχολείου, το Ευρωπαϊκού Πορτφόλιο Γλωσσών (ΕΠΓ) - το οποίο χρησιμοποιείται παράλληλα με το Ντοσιέ (φάκελος) του προφορικού φακέλου - και ένα τελικό ερωτηματολόγιο αξιολόγησης που δίνεται στους μαθητές μετά την ολοκλήρωση του προγράμματος με το προφορικό φάκελο. Η στατιστική ανάλυση βασισμένη στα

αποτελέσματα της αξιολόγησης από τους μαθητές και την παρατήρηση του δασκάλου καθ' όλη την διάρκεια του σχολικού έτους, οδηγούν σε μερικά πολύτιμα συμπεράσματα σχετικά με την βιωσιμότητα του προφορικού φακέλου. Είναι προφανές πως το προφορικό πορτφόλιο αποτελεί μια χρονοβόρα και απαιτητική διαδικασία αξιολόγησης. Παρόλα αυτά, οι μαθητές το βλέπουν σαν μια ενδιαφέρουσα εμπειρία και είναι πρόθυμοι να το χρησιμοποιήσουν ξανά στο μέλλον. Συμπερασματικά ο φάκελος μαθητή είναι μια καινοτόμος μέθοδος αξιολόγησης που μπορεί πραγματικά να προάγει την ανάπτυξη των προφορικών δεξιοτήτων στην τάξη της Αγγλικής γλώσσας, να αναπτύξει τις μεταγνωστικές στρατηγικές των νεαρών μαθητών και να αυξήσει το ενδιαφέρον τους για μάθηση.

Key words: alternative assessment, young learners, needs analysis, oral portfolio *Dossier,* European Language Portfolio, evaluation questionnaire

---

Introduction

Although formal assessment is a mainstay of educational programmes (Butterfield et al., 1999), there has been a recent shift in pedagogy to alternative methods of assessment, which, among other things, is believed to enhance learners' metacognitive knowledge and strategies leading to the development of lifelong learning skills (Council of Europe, 2001). One of the most prominent forms of alternative and authentic assessment in the language field is portfolio assessment.

Speaking skills assessment in English as a Foreign Language (EFL) in Greek primary education is undervalued, because alternative methods of assessment suggested by the Greek Government Gazette (2003, p. 381) are scarcely used, as witnessed by the class teacher and author of this paper. The focus of this study will be to define young learners' speaking skills needs and difficulties through needs analysis and then try to improve and develop them by using an alternative method of learning and assessment, i.e. oral portfolios, following, thus, the National curriculum specifications both for tasks development and assessment. The European Language Portfolio (ELP), used for the purposes of this study, aims to enhance speaking skills in the English language, through the assignment of communicative speaking tasks that focus on describing people, narration (story-telling), explaining a procedure, and on information transfer. These tasks agree with the objectives set for the A1/A2 level of the CEFR (Council of Europe, 2001, p. 24) regarding productive skills development. Will the oral portfolio together with the communicative speaking tasks make a difference in learning and assessment in the current teaching context? This is the main research question addressed in the present study.

The official EFL curriculum for the Greek state primary schools

The Greek EFL curriculum for primary education (Greek Government Gazette, 2003) is based on the Common European Framework of Reference (Council of Europe, 2001). There are three axes on which the curriculum is set, namely, that of *literacy*, *plurilingualism* and *pluriculturalism.* On a syllabus level this means that the pupils of the 4th to 6th grade of the primary school should acquire gradually the following skills: receptive and productive skills, strategies for learning and communication, parallel use of the L1 and L2 and development of multicultural conscience.

Assessment in the curriculum is defined as evaluation of the teaching aims and objectives, the teacher and the pupils (Greek Government Gazette, 2003, p. 381). In relation to pupils, assessment does not concern only learners' language knowledge, but also their ability to use it in various and authentic situations. Assessment methods should be varied and the criteria should be based on the predetermined learning objectives avoiding the comparison of learners with each other. The use of pen-and-paper tests, which assess mostly reading and writing skills, should be combined with alternative forms of assessment, if all four skills are to be assessed properly (ibid, p. 381). Communicative tasks are at the heart of alternative assessment methods, i.e. pupils' portfolios, project work, self- and group- assessment. In the next section, the classroom material and assessment methods will be compared against the curriculum principles, in order to justify the use of portfolio assessment in this study.

Implemented syllabus and curriculum compatibility: considerations for portfolio assessment

The textbook and workbook used in the particular teaching context are *Fun Way 2* (Pedagogical Institute, 2000). The book *syllabus*, that is, "the content or subject matter of an individual subject" (White, 1988, p. 4) covers the four language skills to one degree or another. It belongs to the *Type A* syllabi[1], which are product-oriented focusing "on what is to be learnt by pre-selecting objectives and content before any consideration of the specific learners and by assessing success in terms of achievement" (White, 1988, p. 44). This text-based syllabus comes in conflict with the curriculum ideology; several inconsistencies of the syllabus methodology make it far from communicative. Namely, there is no integration of skills and writing activities are missing. Negotiation of meaning and communication are partially used. Also, activities do not exploit language in a realistic way. Dialogue - active learning - is not used in all kinds of activities. Thus, project and group work are not promoted, either. Lastly, there is no parallel use of the L1 and the L2 language in the activities of the book.

When it comes to the reality of the classroom, the assessment of the particular learners' performance in the foreign language does not fully reflect the curriculum specifications, either. The technique that is employed by many state school teachers - the author included - is teacher-made, paper and pencil tests, i.e. progress tests, which do not assess oracy skills, mainly due to time limitations. As Brown and Hudson (2002) argue, a mismatch between curriculum objectives and tests can make students want to study only whatever is on the tests.

Since the written standardised tests are the only method of assessment for these primary pupils, they should be complimented, as the curriculum suggests, too (Greek Government Gazette, 2003, p. 381). In terms of speaking skills assessment, this should be done through communicative tasks, which, in turn, call for the use of alternative assessment methods (ibid, p. 381).

Assessment and alternative assessment: Portfolio

Testing and assessing young learners

Katz (1997, p. 1) very aptly characterises young learners - age group from six to twelve years old - as "notoriously poor test-takers: perhaps because they are sometimes confused by being asked questions that they think the tester must already know the answers to"! For young learners, who come to the second language classroom without their choice and

without still recognizing the usefulness of a foreign language, a special approach to the language assessment[2] is warranted. McKay (2006, p. 25) narrows down what makes young learners' treatment special to three points: *growth*, *literacy* and *vulnerability*. Since they grow cognitively, socially, emotionally and physically at the same time, young learners are unstable. Their literacy knowledge and skills' development are a slow process for most of them, and they are particularly vulnerable to criticism or failure resulting from assessment.

Portfolios: The ELP and the Junior Portfolio

According to Trim (1997, p. 3) "a language portfolio is a document… in which individual learners… can assemble over a period of time, and display in a systematic way, a record of their qualifications, achievements and experiences in language learning, together with samples of work they have themselves produced". The portfolio used in this study is the European Language Portfolio documented to fulfill both the assessment and learning functions.

The *European Language Portfolio*[3] (ELP) is the official educational tool produced by the Language Policy Division of the Council of Europe. It was launched in 2001 at a pan-European level to celebrate the European Year of Languages. The ELP is divided in three separate but interconnected sections (Council for Cultural Cooperation, 2000), namely the *Language Passport* (language certificates), the *Language Biography* (language experiences) and the *Dossier* (samples of personal work). Kohonen (2000, p. 8) looks at the *Dossier* as "a dynamic and flexible pedagogical tool that can be used regularly in language teaching", in contrast to the Biography and the Passport sections that are more detached from the daily language classroom.

The *Junior Portfolio* (CILT, 2006) is examined in this study. It was recently designed by the Centre on Information of Language Teaching and Research (CILT) and it specifically addresses young learners; it corresponds to the linguistic development of the particular learners (past A1/A2 CEFR level) and it is child-friendly, i.e. it is colourful, illustrated and easy to follow. It is, in fact, a ring-folder with an attractive blue plastic hard cover. Just as the ELP, it is divided in the three main sections mentioned above. The *Dossier* is the most important part of the ELP for young learners, as it is a personal collection of their own work (CILT, 2006, pp. 17-18). Any special piece of their work can be filed here illustrating their experiences and achievement.

Oracy skills development in primary education and needs analysis

Communicative speaking tasks

Communicative tasks can help towards the effective assessment of oracy skills. Nunan (1993, p. 59) defines a communicative task as "…a piece of classroom work which involves learners in comprehending, manipulating, producing or interacting in the target language while their attention is principally focused on meaning rather than form". If speaking skills development is to receive more attention in the syllabus, then this has to be done through more speaking tasks that are also more communicative. Apart from the irregularities between the national curriculum and the coursebook syllabus mentioned earlier, it is assumed that a particular syllabus should match as closely as possible the needs of the particular learners. Thus, it should not be designed in a vacuum. This is feasible by analysing students' needs.

Case study: Oral portfolio implementation

Participants

This study involves a mixed-ability, mixed-sex class of sixteen Greek learners of English at the fifth grade of a state primary school. Students attend three English teaching periods per week, lasting 45 minutes each. The level of proficiency expected to be achieved by the pupils at the end of this grade roughly corresponds to the A1/mid A2 level (Basic user) of the Common European Framework of Reference (Council of Europe, 2001). With regard to oral skills, A2 level pupils are expected to be able to comprehend and produce simple phrases and sentences related to familiar topics, handle short social exchanges, simply describe people and places, ask for repetition or clarification and, generally, satisfy their most basic communication needs in everyday situations.

Needs analysis questionnaire

There are many methods for conducting needs analysis. Using a questionnaire is the most practical and thus, the most common needs analysis tool. This is so because it can be adjusted to the language and level of proficiency of the learners. It can also provide tabulated results, which are easier to analyse. The main purpose of the questionnaire (Appendix I) administered to the particular class of primary school learners was to define their *subjective* needs[4] in speaking skills. It was given to the pupils together with its Greek translation in the classroom at the beginning of a lesson, early in the school year.

Questionnaire results and evaluation

The particular questionnaire focused mostly on learners' needs in terms of skills development. This was the reason, why section C (Appendix I) was disproportioned in comparison to the other sections. The data collected from the questionnaire analysis[5] showed that pupils reported their willingness to produce language in English, although they felt that speaking is more difficult for them than writing or rote learning of vocabulary and grammar. Additionally, they felt that speaking was practiced less in class and most of them would prefer to have more oral activities. Their communicative nature to learning were also shown by their preference to work in class, either in groups or alone. The fact that they were in favour of innovative forms of teaching and assessment was consolidated by their almost absolute unanimity about self-evaluation, too. In such a traditional text-based classroom with pen-and-paper test routines, learners' answers denoted both eagerness and preference to alternative methods of teaching and assessment. It is in children's nature to be explorative and innovative, so we should not only give them roots but also wings, as Donaldson (1978) would argue, providing them with challengeable learning tasks and with self-involving assessment tools.

Rationale for using an oral portfolio as an assessment tool

As the questionnaire results showed, the particular learners needed more speaking practice in English, but they also needed an unthreatening environment, in order to view oral activities as something within their grasp. The oral portfolio can provide such an environment, where learning and assessment are a natural, experimental and harmless process. Additionally, it is an assessment and learning tool, compliant with the National curriculum specifications.

The ELP and the Dossier

The Junior Portfolio (see earlier discussion) booklet of each pupil was accompanied by a cardboard folder that constituted the *Dossier*. Both were kept in the classroom. Since this was an oral portfolio, the *Dossier* consisted – by the end of the project - of pupils' audio recordings of the three out of the four oral tasks performed throughout the year, of notes and summaries that helped them with the performance of the oral tasks, and evaluative forms. A thorough analysis of portfolio tasks and its components follows below.

Implementation of the portfolio

The oral portfolio was implemented in the academic year 2008-2009 for a period of six months. The idea of the oral portfolio was introduced to the pupils at the beginning of the school year and they were informed that they needed to be equipped with a recording equipment (e.g. a micro-tape recorder, an mp-3-/4 or a cell phone with recording function). Moreover, the parents, as well as the headmaster of the school, were notified of the new component of oral assessment.

The work with the *Junior Portfolio*, as well as with oral task performance and assessment, was ideally done once a week. By the end of the oral portfolio project an evaluation questionnaire was completed by the pupils (Appendix II). Additionally, as a gesture of appreciation, the teacher distributed to the pupils a certificate of achievement (Appendix III) to take home with their portfolio.

Portfolio tasks and components

The *Dossier* of the portfolio contained four speaking tasks, so that they could be easily and evenly distributed throughout the school year. It was also thought, that four tasks would be enough to keep young learners' enthusiasm high. Due to the transient nature of speaking, students' performance had to be captured and recorded in order to be available for review and analysis after the live performance both by the pupils themselves and by the teacher. A sample of transcripts of the taped tasks can be viewed in Appendix IV.

The four tasks were oral tasks for young learners widely suggested in the literature (McKay, 2006; Heaton, 1990; Underhill, 1987; Byrne, 1986). They were chosen because they met the National curriculum specifications about using communicative tasks (Greek Government Gazette, 2003, p. 381) and because they fell into the A1/A2 level of the CEFR (Council of Europe, 2001, p. 24). The oral assessment tasks are the following presented in the order they were performed:
a. Presentation of a person they know well (Description).
b. Story telling and re-telling based on a book (Appendix V).
c. Performance task; explaining how to make something (Procedure).
d. Information transfer task; explaining where to place household items on a worksheet according to others' description (taken from Ioannou- Georgiou & Pavlou, 2003; see Appendix VI).

All four tasks are characterized by communicative authenticity[6]. In line with Underhill (1987), all these tasks are communicative to an extent. The author explains that, "when a learner says something that is relevant and true (for himself at least), to someone else who is interested and has not heard it before (from the speaker, at least), then that act of speech is communicative" (ibid, p. 8). Accordingly, these tasks are authentic to the extent that we all

need at one time or another to describe things, to transfer information accurately and talk about something we have witnessed. These features are more evident in the fourth task, which also embodies the information transfer technique[7]. Also, a communicative feature of the fourth task is the challenge that it provides by creating suspense for the outcome of the task.

The performance of the first three tasks was tape recorded both by the pupils and by the teacher. The last task was not recorded, because of the noise level anticipated in pair work, but it was evaluated using classroom observation and self-assessment forms. Pupils were allowed to use notes/realia for their oral performances. For instance, for the third performance task the pupils could bring the materials/photos they needed for the explanation of the procedure (i.e. of a construction, a snack, a recipe).

Teacher's and learners' assessment tools

Alternative assessment should satisfy the same psychometric qualities, as do conventional tests, a fact accredited by many academics in the relevant literature (Bachman and Palmer, 1996; Brown and Hudson, 1998a/1998b; Council of Europe, 2001; inter alia). *Validity*, *reliability* and *feasibility* are the three fundamental qualities that any method of assessment needs to meet. For the teacher's assessment and the pupils' self-evaluation of the oral tasks in the current portfolio, the methods used to minimize subjectivity in scoring and increase validity and reliability are rating scale rubrics and checklists (Appendix VII). All rating scale rubrics consist of three scales of language ability and three to four evaluation criteria. For the sake of sampling and demonstrating evidence, in Appendix VIII there is an assessed oral task taken from a pupil's portfolio dossier.

Evaluation of the oral portfolio

Reflection of the assessment procedures of the oral tasks

One of the basic merits of portfolios is the possibility to collect the documentation of children's learning achievements into a coherent whole (Stiggins, 2005). The assessment of the speaking skills in the particular case study was done through two tools; through the *Junior Portfolio* booklet and through the rating scales and checklists of the oral tasks filed in the *Dossier*. The former played the role of an "organiser of learning" and of an overall assessment (self-assessment) of pupils' speaking skills and progress throughout the year. The *Dossier* of the oral portfolio, on the other hand, was the core of the study used for a scrutinised oral task assessment on the part of teachers, pupils and parents, in extent. The participation of the pupils in the portfolio assessment and their effort to perform the tasks were valued the most, so the certificate of achievement was accredited to all. Undoubtedly, the best judges to evaluate the oral portfolio are the teacher and pupils involved. Their evaluations follow below.

Teacher's evaluation and findings

For the sake of clarity, the difficulties that portfolio development and implementation presented will be differentiated to practical and technical ones. The practical difficulties related to the procedures and processes of portfolio assessment. The most obvious one was that of the workload versus time limit. Although the particular portfolio assessment was restricted to speaking skills, it was marginally accomplished within the scheduled year plan. Moreover, timelines were hard to meet, because of forgetfulness and absences on the part

of the pupils, and unexpected school activities. As a result, some taped performances were lost and some pupils abstained from the tasks.

Class observation showed that pupils took the four oral tasks seriously and they worked hard on them. Nevertheless, it was a stressful time for them when it came to recording their performance. Procedures of the oral portfolio were time-consuming disorienting some pupils to believe that portfolio analysis is a way to miss the traditional English lesson. Lastly, the recording of the three oral tasks presented some delays occasionally, since machines were not always trustworthy. In such cases, the teacher, who, in any case, recorded all performances, made a copy for the pupils who did not have their own recordings.

The technical problems stem from the nature of the oral tasks being assessed. By examining the tasks for evidence of the psychometric qualities mentioned above the following conclusions were reached. On the one hand, validity of inferences was established to an extent because of the unanimity of the oral tasks for all, and of the assessment of tasks by one teacher only as the involvement of more assessors would jeopardise the validity of inferences drawn. On the other hand, reliability, i.e. objectivity in marking (Koretz, 1998), and feasibility were enhanced with the construction of analytic rating scales rubrics and checklists, as well as the use of clear criteria. All these made marking semi-objective, consistent and practical, although some subjectivity remained, due to the nature of the assessment of the oral. However, practicality was hindered in the case of the fourth task, which could not be recorded due to classroom noise.

Self-assessment created a non-competitive environment in the classroom and that led to a beneficial washforward effect[8] of portfolio assessment. Moreover, the utility of the portfolio was high both for advanced and weaker pupils, because they could all participate. In the following section, the results of pupils' portfolio evaluation shed some more light to the impact it had on the particular context.

Learners' evaluation and aptitude to portfolios

Making pupils responsible for their own learning is one of the aims of alternative assessment. They were, therefore, asked to evaluate oral portfolio through a questionnaire (Appendix II – presented in both English and Greek to the students).
The results of the evaluation questionnaire showed that the 16 pupils not only accepted portfolio assessment and evaluation smoothly, but also embraced it and supported its interesting (13 pupils) and useful nature (8 pupils), although they had been introduced to it for a short time (Table 1).

Question 1	
Options	Frequency
more difficult	0
more interesting	13
more useful	8
fairer	5

Table 1: Portfolio assessment versus traditional tests assessment

The pupils seemed eager to move away from the traditional assessment test and typical classroom routines, since the oral portfolio could create more enthusiasm in the English lesson for 11 pupils and motivate 10 of them to improve their speaking skills (Table 2).

Question 2	
Options	Frequency
more interest in English	11
more confidence in speaking	10
awareness of weaknesses	10
more responsibility in learning	9

Table 2: Portfolio creates...

The preparation and recording of oral tasks (Table 3) presented almost half of the pupils (9 pupils) with some stress and extra work, but the majority (11 out of 16 pupils) said that the tasks were interesting and feasible; pupils performed them more diligently and laboriously, than it was ever expected, with minimum negative reactions. Nevertheless, although pupils liked the audio recorded components of their tasks, only 5 pupils saw its usefulness for their self-assessment and further improvement (Table 3).

Question 3	
Options	Frequency
understandable and interesting	11
stressful experience	7
meaningful because recorded	5
demanding in preparation	9

Table 3: Oral tasks

Additionally, self-assessment was an exciting experience to them (14 pupils) and helped most of them (13 pupils) to become aware of their language proficiency level (Table 4).

Question 4	
Options	Frequency
interesting experience	14
awareness of proficiency level	13
boring and worthless	1
difficult	0

Table 4: Self-evaluation of oral tasks

In effect, the majority (13 out of 16 pupils) favoured portfolio assessment and expressed the desire to use portfolio assessment again in the future (Table 5).

Question 5	
Options	Frequency
in favour	13
against	3

Table 5: Portfolio assessment

In the next section, conclusions will be drawn and relevant suggestions will be proposed.

Suggestions for future action

The evaluation of the teacher and pupils discussed above suggests that portfolio assessment intrigued pupils and attracted their attention. To become effective, though, portfolios have to become an undivided part of current pedagogy in the ELT classroom. This may still seem

to be a long and bumpy road, but it is a one-way road, because recent underpinning theories of learning and the curriculum design favour the implementation of portfolios in the classroom. As De Fina (1992, p. 65) argues, "any approach that involves students in their education and that stimulates and excites them to evaluate themselves and build expertise is certainly worth the effort". Undoubtedly, careful programming and deadlines should be made early in the school year making sure that they are followed strictly by all. Portfolios are a complimentary means of assessment. They could substitute formal assessment anytime, though, especially in the primary school, where there is no large-scale formal assessment of skills, like speaking elaborated in the current study. Cooperation among all those concerned in the teaching, learning and assessment process and support through adequate resources is needed, too.

Improving learning with the help of a portfolio-based assessment remains a challenge. Nevertheless, change for the change's sake is of no value if ELT teachers are not trained in how to employ the oral portfolio. The more the teacher's abilities to assess and report on young learners' progress is trusted, the less the reliance on standardised testing.

Conclusion

Portfolio pedagogy is limited because of the overreliance on standardised testing in the assessment of young learners flowing from the social demand "to prove rather than improve learning" (Klenowski, 2002, p. 76). In the present study, portfolio assessment aimed to outbalance or lessen this reliance by promoting pupils' metacognitive development, and to compensate for the luck of assessment of oracy skills in ELT in Greek primary schools, too. The *Junior Portfolio* aligned with the theoretical principles of the *CEFR* and the oral tasks were developed taking into consideration the National curriculum's aims and objectives, adjusted to the developmental and proficiency level of the particular context. As far as the assessment methods and criteria are concerned, they accommodated what constitutes effective oral and portfolio assessment.

Author's email: geoefthym@gmail.com

Notes

1. White (1988, p. 44) labels the early communicative syllabi as *Type A* syllabi and the more process-oriented ones as *Type B* syllabi.
2. Although used interchangeably, the term *assessment* refers to the judgement carried about the "learner's level of skills and knowledge" (Nunan, 1990, p. 62), whereas *testing* is a subset of assessment dealing with the evaluation of specific learning objectives on the base of standardised tests carried out at specified times of the school year (West, 2004).
3. For more about the ELP see www.coe.int/portfolio.
4. Subjective needs refer to the *process* of learning, that is, how the foreign language needs to be taught for effective learning to take place (Manolopoulou-Sergi, 2004).
5. The tabulated statistical charts can be viewed in Efthymiou (2009).
6. As West (2004, p. 244) explains, a communicatively authentic task replicates all the processes of spoken communication, namely, descriptions, narrations, explanations, instructions, regardless of whether such a task would exist in the real world.
7. In information transfer, there is an information gap between the pupils and they have to convey it to one another in order to complete the task (Johnson, 1982).
8. The washforward effect "refers to the extent to which a test includes and tests language relevant to the post- language learning situation" (West 2004, p. 88).

References

Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice.* Oxford: OUP.

Brewster, J., Ellis, G., Girard, D. (1992). *The primary English teacher's guide.* Harmondsworth: Penguin.

Brown, J. B. & Hudson, T. (1998a). 'The alternatives in language assessment'. *TESOL Quarterly*, 32/4: 653- 675.

Brown, J. B. & Hudson, T. (1998b). 'The alternatives in language assessment: advantages and disadvantages'. *University of Hawaii Working Papers in ESL*, 16/2: 79- 103.

Brown, J. B. & Hudson, T. (2002). *Criterion-referenced Language Testing.* Cambridge: CUP.

Butterfield, S., Williams, A. and Marr, A. (1999). 'Talking about assessment: mentor-student dialogues about pupil assessment in initial teacher training'. *Assessment in Education*, 6/2: 225- 246.

Bygate, M. (1987). *Speaking*. Oxford: Oxford University Press.

Byrne, D. (1986). *Teaching oral language*. New York: Longman.

CILT, The National Centre for Languages. (2nd Ed.) (2006). *European Language Portfolio – Junior version.* London: CILT Publications. Also available on line at http://www.nacell.org.uk/resources/pub_cilt/portfolio.htm

Council for Cultural Cooperation (2000). *European language portfolio: Introduction*. Strasbourg: Council of Europe, accessed at http://www.coe.int/T/DG4/Portfolio/?L=E&M=/main_pages/introduction.html on 5 Feb 2009.

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

De Fina, A. (1992). *Portfolio assessment: Getting started*. NY: Scholastics Inc.

Donaldson, M. (1978). *Children's minds*. New York: Norton.

Efthymiou, G. (2009). 'Alternative assessment of speaking skills in an EFL context in primary education using portfolio.' Unpublished M.Ed. dissertation, Patras: Hellenic Open University.

Greek Government Gazette 304, B', 13-3-2003. *The revised unified Greek curricula*. Athens: The Pedagogical Institute. Also available on line at http://www.pi-schools.gr/lessons/english/pdf/14depps_XenonGlosson-Agglika.pdf

Heaton, J. B. (1990). *Classroom testing.* New York: Longman.

Ioannou-Georgiou, S. & Pavlou, P. (2003). *Assessing young learners*. Oxford: OUP.

Johnson, K. (1982). 'Five principles in a "communicative" exercise type'. In K. Johnson (Ed.), *Communicative syllabus design and methodology,* 163-175. Oxford/London: Pergamon/ Pentice Hall.

Katz, L. G. (1997). 'A developmental approach to assessment of young children'. *ERIC Digest.* ERIC ED 407172.

Klenowski, V. (2002). *Developing portfolios for learning and assessment: processes and principles.* London: Routledge Falmer.

Kohonen, V. (2000). 'Exploring the educational possibilities of the "Dossier": suggestions for developing the pedagogical function of the European language portfolio'. In Council of Europe, *Enhancing the pedagogical aspects of the European language portfolio (ELP)*, 5-31. Strasbourg: Council of Europe, accessed at http://www.coe.int/T/DG4/Portfolio/?L=E&M=/main_pages/documents.html on 31 Jan 2009.

Koretz, D. (1998). 'Large-scale portfolio assessments in the US: Evidence pertaining to the quality of measurement'. *Assessment in Education: Principles, Policies and Practice*, 5/3: 309- 334.

Manolopoulou-Sergi, E. (2004). 'Needs analysis'. In Ayakli, C., Karavas, K. Manolopoulou-Sergi, E. & Spinthourakis, J. A., *Course Design and Evaluation*, Vol. 1: 81- 124. Patras: Hellenic Open University.

Mckay, P. (2006). *Assessing young language learners*. Cambridge: Cambridge University Press.

Modern Languages Division: Council of Europe. (2001). *Common European Framework of reference for languages: learning, teaching, assessment*. Cambridge: Cambridge University Press. Also available on line at http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf

Nunan, D. (1990). 'Action Research in the Classroom'. In J. C. Richards & D. Nunan (Eds.), *Second language teacher education*. Cambridge: Cambridge University Press, 62-82.

Nunan, D. (1993). 'Task-based syllabus design: selecting, grading and sequencing tasks'. In Grookes, G. & Gass, S. (eds.), *Tasks in a pedagogical context: integrating theory and practice.* Clevedon: Multilingual Matters, 55-68.

Pedagogical Institute of Greece. (2000). *Fun way 2. Student's book & Workbook*. Athens: Publishing Organisation Of Educational Books.

Stiggins, R. J. (4th Ed.) (2005). *Using student-involved classroom assessment to close achievement gaps*. Columbus, OH: Merrill Prentice Hall.

Trim, J. (1997). 'A European language portfolio: some questions relating to its nature, function, form, preparation and distribution'. In Council of Europe, *The European language portfolio. Proposals for development*. Strasbourg: Council of Europe, 3-12.

Underhill, N. (1987). *Testing spoken language*. Cambridge: Cambridge University Press.

West, R. (2004). *Testing and assessment in language learning, vol. 1*. Patras: Hellenic Open University, 1-193.

White, R. V. (1988). *The ELT curriculum: design, innovation and management.* Oxford: Blackwell Publishers Inc.

Appendix I

Pupils' needs analysis questionnaire:

NEEDS ANALYSIS QUESTIONNAIRE FOR 5TH GRADE LEARNERS OF A GREEK PRIMARY SCHOOL

The questionnaire that follows will help with specifying your needs better, as far as learning the English language is concerned, and with making the planning of your lessons as good as possible.

A. Background information

Please, give some information about yourself:

Name: ………………………………………………………………………………………

Class: …………………………………… Age: ………………………………

Mark your answer with a tick ☑.

A.1- Do you learn English outside school?　　　Yes ☐　　　No ☐

A.2- If yes, what class are you in?　　A class ☐　B class ☐　C class ☐　D class ☐

B. Learning attitude

English is mostly useful for you, because:

Number the boxes with 1-3 (1= not important, 2= important, 3= the most important).

- you can communicate/write letters to English speaking friends　　☐
- you get better marks at this subject at school　　☐
- you can use your computer (games, Internet, e-mail)　　☐

C. Needs

C.1- What do you consider most difficult in English? Put only one tick (✓).

- Learning new vocabulary/grammar　☐
- Writing descriptions/letters/stories　☐
- Speaking with someone　☐

C.2- Evaluate your abilities and knowledge in English in the following areas.

Put a number from 1–3 in the boxes according to the scale:
[1 = Satisfactory,　**2** = Good ,　3 = Unsatisfactory)

- Reading/Reading comprehension　　☐
- Writing (postcard, letter, story)　　☐
- Conversation/oral speech (speaking)　　☐
- Listening comprehension (listening)　　☐

C.3- What activities do you think you do <u>less</u> in the English classroom at school? Put only one tick (✔).

- Writing activities (writing)   ☐
- Reading activities (reading)   ☐
- Speaking activities/conversations (speaking)   ☐
- Listening activities (listening)   ☐

C.4- Which one of the previous four activities would you like to do <u>more</u>?

   Write only one: ……………………………………

D. Learning and assessment preferences

D.1- How do you think you learn best? Put up to four ticks (✔).

- By watching videos/ pictures/ performing?   ☐
- By reading what you want to learn?   ☐
- By listening to songs/ rhymes/ music?   ☐
- By playing games/doing role-plays/ projects? ☐

D.2- Do you learn better when you do tasks/activities:

   Put only one tick (✔).

- In class (alone or in groups)? ☐
- At home in peace and quiet? ☐

D.3- Do you like the current way of your assessment in English, i.e. with written tests? Put only one tick (✔).

- Yes ☐
- No ☐

D.4- Would to like to take part, too, in your report making your own assessment?

   Put only one tick (✔).

- Yes ☐
- No ☐

Appendix II

Students' evaluation questionnaire

Evaluation of the oral ELP & *Junior Portfolio*

Now that you have completed your own oral portfolio, I would like you to answer
sincerely to the following questions. Your answers will help improve your lessons
in the future. Put a tick (✓) where necessary.

1) Portfolio, in comparison with a traditional assessment test is:
   a. more difficult.   ☐
   b. more interesting.   ☐
   c. more useful.   ☐
   d. fairer.   ☐

2) Has the portfolio helped you to:
   a. become more confident in your oral speech in English?   ☐
   b. become more responsible for your own learning?   ☐
   c. become more interested in the English lesson?   ☐
   d. understand your weak spots in speaking in English?   ☐

3) The four oral tasks were:
   a. understandable and interesting.   ☐
   b. a stressful experience.   ☐
   c. meaningful, because of the audio recording.   ☐
   d. demanding in preparation.   ☐

4) Your self-assessment in the four oral tasks of the portfolio:
   a. was an interesting experience for you.   ☐
   b. helped you to see for yourself your weaknesses and your strengths.   ☐
   c. was boring and worthless.   ☐
   d. was difficult for you.   ☐

5) Would you like to use portfolio again for your assessment?
   a. Yes ☐
   b. No   ☐

Thank you for your cooperation.

Your teacher,

Georgia Efthymiou

Appendix III

Certificate of achievement for the oral portfolio

Speaking in English
(Front page)

ALL STAR
STUDENT
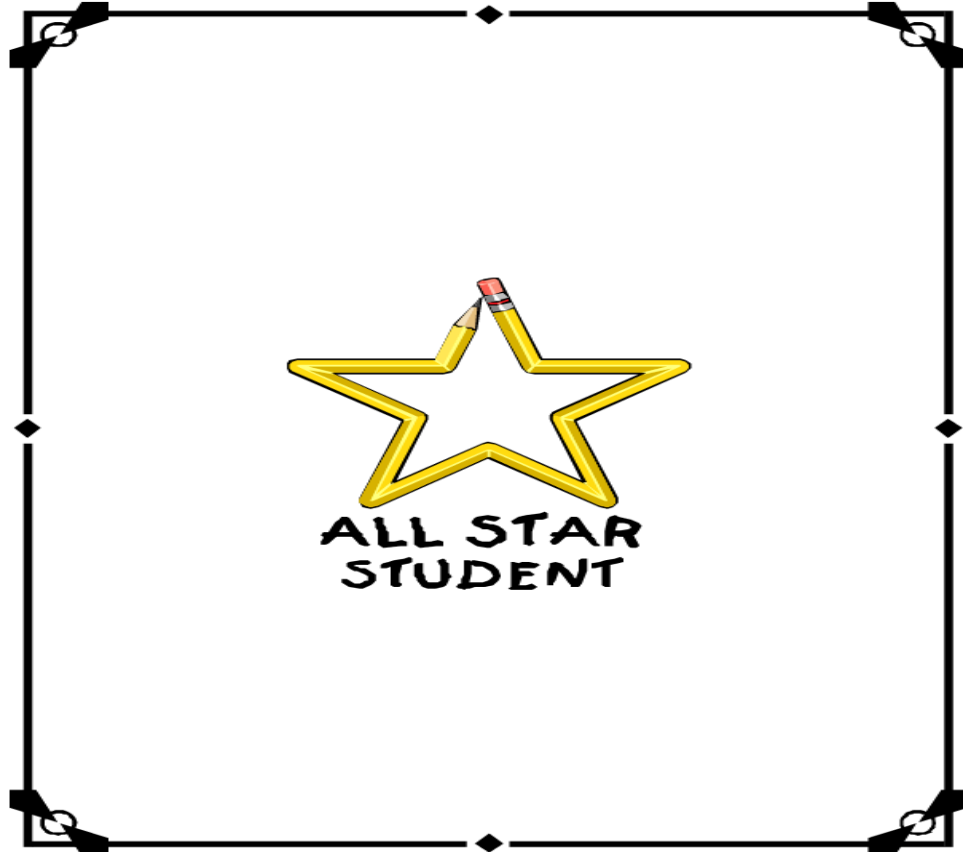
(Back page)

Certificate of Achievement

Christina has earned this certificate for participating in the oral portfolio project. She accomplished the oral tasks successfully and showed improvement in her speaking and conversational skills

Keep up the great work!

The teacher,
Georgia Efthymiou
June 2009
Class: E2

Appendix IV

Sample transcription of the three taped oral portfolio tasks:

Task 1: **Description of a person you know well**

T= Teacher, C= Chrysa (pupil)
1. T: Chrysa, tell us about the person you are going to describe.
2. C: I describe my mum. Her name is Toula. Er… she has got blue eyes and long fair hair. Em… she likes..eh… dress..eh.. she likes dress and… and drive cars. She likes cooking, too. She doesn't like false and playing chess... and the noise. I love my mum..eh.. and... I love my mum!
3. T: Okay. Good. Er… what about her… That's all?
4. C: Yes.
5. T: Her character? Have you talked about her character?
6. C: Eh, yes. Er… she is very good character and eh….she always… she always eh... good for us.
7. T: Okay. Thank you.

Task 2: **Story-telling**

T= Teacher, A= Andreas (pupil) -*The snowman*
1. T: So, Andreas. Tell us about your story.
2. A: This story is about a snowman and a child. Er… one morning the child is wake up and see the window and out is snowing. It wear his clothes and it goes out to play. Er… it makes a snowman and the night… eh… the boy brush his teeth and see the snowman. When it goes to bed, the.. the.. child is go down, open the door and it can see the snowman. They are playing all the night and the snowman takes the child from his hand and they fly in the sky and they go in a party for snowmen. Then they drink, they dance, and they do a lot of things. Then, te child is go to home, it goes to sleep and the other morning the child is…see the snowman, but the snowman doesn't… isn't there, and then the child is crying… and… and…(pause).
3. T: Okay, what happens next? That's all?
4. A: yes.
5. T: Yes? So, the story ends there. What happens to the snowman?
6. A: Er… (in greek) Πώς είναι το «έλιωσε»; Δεν ξέρω.
7. T: He melts?
8. A: Yes.
9. T: Okay, the sun is out and the snow… becomes water.
10. A: Yes.
11. T: Okay. This is it. I think something else happens at the end. The child has got something from the snowman.
12. A: Yes.
13. So, he knows that the whole thing is real.
14. A: Yes.
15. T: Not just a dream. Okay, good. Very good.

Task 3: Procedure description

T= teacher, F= Fanis, P= pupil – *How to make a rice pudding*
1. T: Fani, tell us. What are you going to make?
2. F: I am going to make a rice pudding.
3. T: Okay.
4. F: I must have rice, milk, sugar, and, if I want, cinnamon. First, we are boiling the rice. After that, we are putting the… rice in a bowl with milk.
5. T: Okay.
6. F: Eh… then… we are putting the sugar inside the milk and the rice…
7. T: Ahm.
8. F: And finally, we got our rice pudding. If we want, we can put cinnamon eh… over it.
9. T: … on top of it.
10. F: Yes.
11. T: Do you boil the milk with the rice?
12. F: No.
13. T: No?
14. F: Only the rice with water.
15. T: Okay. I did not know we make pudding like this. Have you done this?
16. F: Yes.
17. T: Yes? Right. Have you got any questions? Christina?
18. C: Is it delicious?
19. F: Yes.
20. T: Okay, anything else?
21. P: What is pudding?
22. Class: Πουτίγκα. Ρυζόγαλο.
23. T: Okay, thank you, Fanis.

Appendix V

The list of books read by the pupils for the story telling and re-telling task:

(1)   Disney, W. (1996). *Toy story.* Connecticut: Glolier Enterprises Inc.

(2)   Purnell Books. (1983). *Moshops and the mystery.* Exeter: A. Wheaton & Co. Ltd.

(3)   Disney, W. (1985). *Winnie the Pooh and the blustery day.* Leicestershire: Ladybird Books.

(4)   Escott, J. (1992). *How the rhinoceros got his skin.* Manchester: Little Owl.

(5)   Disney, W. (1997). *The fox and the hound.* Connecticut: Grolier Enterprises Inc.

(6)   Dooley, J. & Kerr, A. (2002). *The little mermaid.* Berkshire: Express Publishing

(7)   Briggs, R. (1988). *The snowman.* Leicestershire: Ladybird Books.

(8)   Cunliffe, J. (1993). *Rossie and Jim.* London: Scholastic Publications Ltd.

(9)   Purnell Books. (1983). *Moschops digs a hole.* Exeter: A. Wheaton & Co. Ltd.

(10)   Adler, H. (1988). *The cock the mouse and the little red hen.* Leicestershire: Brown Watson.

(11)   Billam, R. & Julian0Ottie, V. (1994). *Alpaca the rabbit.* London: HarperCollins Publishers Ltd.

(12)   Dooley, J. & Kerr, A. (2002). *Anna and the dolphin.* Berkshire: Express Publishing.

(13)   Hunia, F. (1993). *The elves and the shoemaker.* Leicestershire: Ladybird Books.

(14)   Mitchell, H. Q. (2005). *Robin Hood.* London: MM Publications.

(15)   Oxenbury, H. (1988). *Eating out.* London: Walker Books.

(16)   Mitchell, H. Q. (2005). *Jasper's pot of gold.* London: MM Publications.

Appendix VI

Task 4 worksheet – Taken from Ioannou-Georgiou & Pavlou (2003).

Describe your room to your partner:
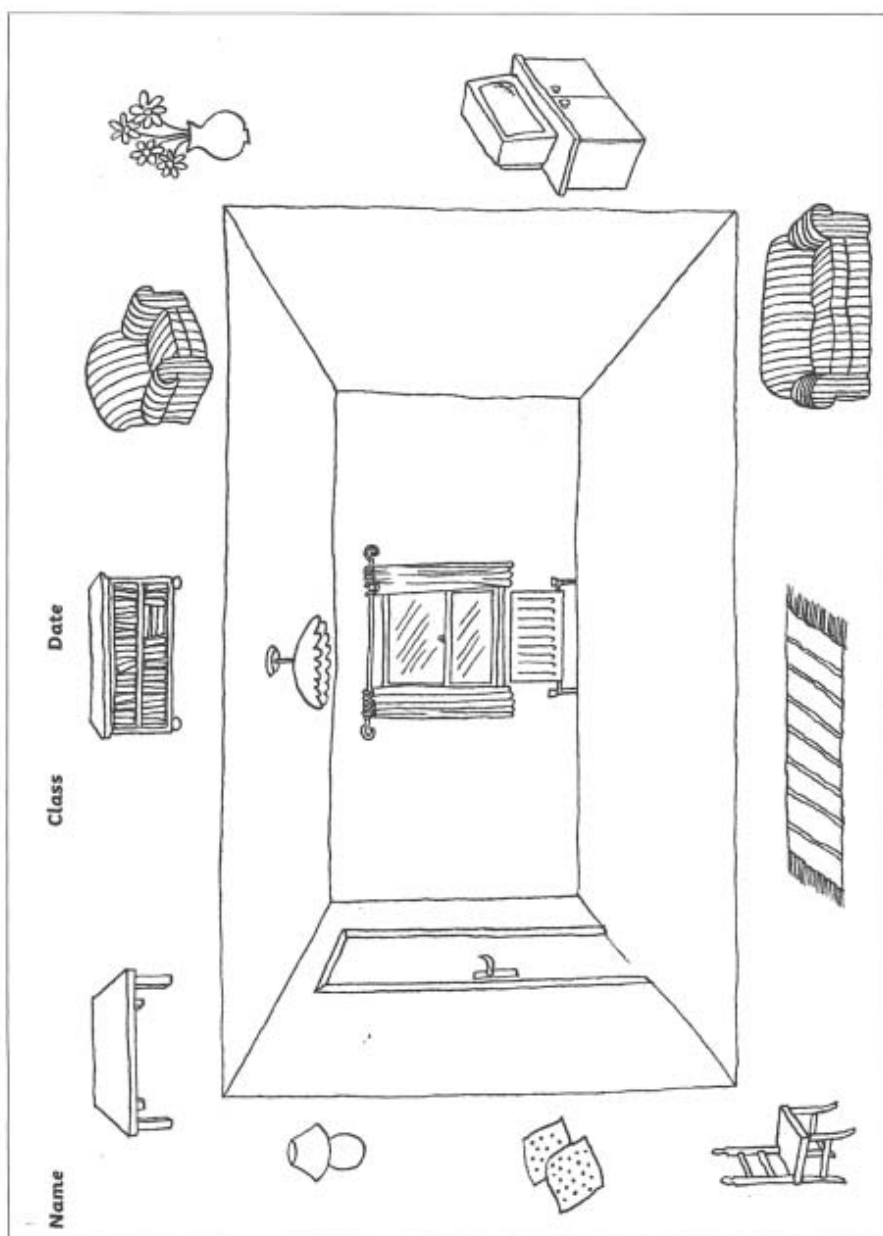
Appendix VII

A checklist form and a rating scale rubric used both for the teacher's assessment and pupils' self-assessment of the description task (Task 1):

1. <u>**Description checklist**</u>

☐ **Physical description of the person**

☐ **Character description**

☐ **Likes**

☐ **Dislikes**

☐ **Other details/information**

1. <u>**Rating scale rubric**</u>

Put a tick (✓) in the right box:

	☺	😐	☹
Vocabulary			
Grammar			
Fluency			

Appendix VIII

A sample of an assessed oral task, i.e. task 3 – story telling and retelling:

A1. Teacher's assessment form of storytelling:

Ευθύμης
Καπλάνης

Task 2

<u>Story-telling checklist- Teacher assessment</u>

Title of the story: Moschops digs a hole

A. Content and structure of the narrative:

	<u>YES</u>	<u>NO</u>
- Mentions what the title of the book/ story is.		✓
- Describes where/ who /what.	✓	
- Describes atmosphere/ feelings.	✓	

B. Language of description:

	<u>YES</u>	<u>NO</u>
- Vocabulary	✓	
- Grammar	✓	
- Fluency	✓	
- Coherence		✓

A2. Teacher's assessment form of story retelling:

"Moschops Digs a hole"

Ευθύμης Καπλάνης

Story – retelling rubric

Put one ✓ to the right box.

	Points: 8	Points: 9	Points: 10
Vocabulary growth	Simple	Developed ✓	Variant
Sentence length	Small	Medium ✓	Big
Fluency	Many pauses	Few pauses ✓	Some pauses
Sequence	No neat story line sequence	Quite clear story line ✓	Smooth sequence

B1. Pupil's summary of his story (for the story retelling part of task 3):

This story is about dinosaurs.
Two dinosaur kinds dug a hole accidenmaly
in Mrs Karty's patio but they had no idea
what this patio. When Grandpa has learnt
about this wat really angry. Then Mr
Ichtyosaurus had popped his head aut of
the water to see what happend. First
Grandpa explained what a patio is and
ask them about the hole. Moschops said
that the truth can be different thing
to different people. Everyone looked
at Mochops with puzzled faces even
Unkle Rex. Then they took to exaple Grandpa.
For Mr Ichtyosaurus Grandpa is a signal for
sleep for Ally is a wall to rest and for
unkle Rex is a hill that you can see your
enemies. Suddenly a flower began to sing
but theye didn't understand what it wanted to say
Finally Moschops understand that Grandpa is
a shadow for the flower. But Grandpa still
didn't understand what is this all about the
hole. Moschops said that they just lowered
the floor. Then she laught and went to Mrs Yatty who was
In the end Ally find a solution and they all learnt a lot
a good less

B2. Pupil's self-evaluation form in Greek for the story retelling task:

Βάλε στο κατάλληλο κουτάκι ένα ✓.

	Points: 8	Points: 9	Points: 10
Λεξιλόγιο	Απλό	Αναπτυγμένο	Ποικίλο ✓
Μέγεθος προτάσεων	Μικρό	Μεσαίο	Μεγάλο ✓
Ευχέρεια λόγου	Πολλές παύσεις	Μερικές παύσεις ✓	Λίγες παύσεις
Συνοχή	Όχι ομαλή σειρά γεγονότων	Αρκετά ξεκάθαρη σειρά γεγονότων ✓	Ομαλή σειρά γεγονότων

Moschops Digs a Hole

Self-assessment: an alternative method of assessing speaking skills

[Αυτοαξιολόγηση: μια εναλλακτική μέθοδος αξιολόγησης του προφορικού λόγου]

Ekaterini Chalkia

The present study focuses on self-assessment as an alternative method of assessing the speaking skills of a group of sixth graders of a Greek State Primary School. The paper consists of two parts. In the first part, traditional and alternative assessment approaches are compared and a literature review on self-assessment is presented. In the second part the methodology and the findings of the study are presented. The study was carried out by means of a questionnaire and observation notes. This was done in order to draw conclusions on the benefits of self-assessment, the difficulties students faced while carrying out self-assessment as well as to reveal the extent to which students improved their speaking skills after being involved in self-assessment. The findings revealed that the students were positive towards self-assessment. Although self-assessment was of limited duration, it turned out to be a worthwhile activity as it fostered motivation and sensitized the students to take a more active role in the learning process. It also enabled them to notice their strengths and weaknesses and improve their speaking skills. The study also revealed the practical difficulties the students faced in carrying out their self-assessment. Finally, the study concludes with recommendations for further research into this specific assessment method.

೮ಌ

Η παρούσα μελέτη εστιάζεται στην αυτοαξιολόγηση σαν μια εναλλακτική μέθοδο αξιολόγησης του προφορικού λόγου μιας ομάδας μαθητών έκτης τάξης ενός Ελληνικού δημόσιου δημοτικού σχολείου. Η μελέτη αποτελείται από δύο μέρη. Στο πρώτο μέρος συγκρίνονται η παραδοσιακή και η εναλλακτική αξιολόγηση και παρουσιάζεται το θεωρητικό πλαίσιο σχετικά με την αυτοαξιολόγηση. Στο δεύτερο μέρος παρουσιάζονται η μεθοδολογία και τα αποτελέσματα της μελέτης. Η μελέτη έγινε μέσω ενός ερωτηματολογίου και καταγραφής παρατηρήσεων με σκοπό να βγουν κάποια συμπεράσματα σχετικά με τα οφέλη της αυτοαξιολόγησης, των δυσκολιών που αντιμετώπισαν οι μαθητές κατά τη διάρκειά της καθώς επίσης και τα οφέλη αναφορικά με τη βελτίωση του προφορικού λόγου των μαθητών. Τα αποτελέσματα έδειξαν ότι οι μαθητές ήταν θετικοί στην αυτοαξιολόγηση. Παρόλο που η αυτό-αξιολόγηση είχε περιορισμένη χρονική διάρκεια, αποδείχτηκε μια δραστηριότητα που ενίσχυσε το κίνητρο των μαθητών και τους έδωσε τη δυνατότητα να έχουν πιο ενεργό ρόλο στη διαδικασία της μάθησης, να

εντοπίσουν τις δυνατότητες και τις αδυναμίες τους και να βελτιώσουν τον προφορικό λόγο τους. Η μελέτη έδειξε επίσης τις πρακτικές δυσκολίες που αντιμετώπισαν οι μαθητές κατά τη διάρκεια της αυτοαξιολόγησής τους. Τέλος, η μελέτη συζητά προτάσεις για περαιτέρω έρευνα σχετικά με τη συγκεκριμένη μέθοδο αξιολόγησης.

Introduction

Contrary to traditional assessment practices whereby learners individually take a test, alternative assessment is aligned with current pedagogical practices, tailored to enhance collaborative pair or group work and preparing autonomous learners equipped to function in the real world where collaboration is needed. Among such methods is self-assessment, which is the subject of this study.

The study has been undertaken for two main reasons. First, working in the Primary Education with young learners for fifteen years, student assessment is relevant to my personal and professional interests. Referring to speaking skills, observation of the everyday classroom reality has revealed a disproportionate amount of time spent on reading aloud, asking comprehension questions and performing drills as the core of instruction. Regarding the assessment of speaking, it is ignored and the only means employed are paper-and-pencil tests. Besides, this study is relevant to the current reforms to the educational system in Greece, which demands innovation and multiple approaches to assessment, a component of which is self-assessment (Ministry of Education and Religious Affairs and the Pedagogical Institute, 2002). This study is also of interest to English as a Foreign Language (EFL) teachers who might want to implement self-assessment of speaking in their own contexts. The paper begins with the theoretical background, then it describes the methodology that was followed and finally it presents the findings and the conclusions that were drawn.

Theoretical background

Self-assessment: definition and related terms

Self-assessment is the ability of individual students to judge their performance, making decisions about their selves and their abilities. Blatchford (1997, p. 2) defines it as a process involving judgments of one's own attainment in relation to other children. Montgomery (2001, p. 5) defines it as students' appraisal of their own work or learning processes. Gronlund and Cameron (2004, p. 14) emphasize its importance as a way to operationalize the principles of formative assessment with the purpose to monitor learning progress and providing corrective prescriptions to improve learning. Two terms commonly related to self-assessment include *self-evaluation*, and *self-monitoring*. According to Rolheiser and Ross (2000) the former involves the students' judgment of the quality of their work, based on evidence and explicit criteria for the purpose of future improvement. The latter is a term initiated by Dickinson (1987) referring to processes which include record keeping and can be carried out by means of fixed format record cards or sheets, learning diaries, etc.

Traditional and alternative assessment procedures

Traditional testing (dictation, multiple-choice, fill-in-the-gaps, matching activities and discrete-item tests) has been popular in most educational contexts including the Greek one and a useful instrument in improving students' knowledge and skills, clarifying the objectives of curricula, planning instruction, reinforcing teaching and learning and promoting educational development (Tsagari, 2004). Nevertheless it fails in providing information about students' attitudes, motivation, interests, and learning strategies and it has been severely criticized as inappropriate and harmful for young learners as it does not reflect their developmental changes while at the same time it distorts the curriculum in the early grades (Shepard, 1994). Alternative assessment which is a continuous process involving students and teachers in making judgments about the students' progress (O'Malley and Valdez-Pierce, 1996) seems to constitute a remedy for a number of reasons. It emphasizes the *process* by which learners produce an outcome rather than the *product* (Puhl, 1999). Focusing on the product, tests seem inadequate to provide the continuous measurement of student growth necessary for planning instructional strategies, whereas alternative assessment better reflects the developmental processes in language learning (Genesee and Hamayan, 1994) thus best reflecting the needs of young students.

Moreover, alternative assessment places emphasis on *feedback* which increases student achievement. Feedback enables students to become aware of the gaps that exist between their current knowledge and skills and their desired goal and guides them through specific suggestions to attain this goal (Boston, 2002). Traditional testing provides feedback associated with *norms*, a single, all encompassing mark which does not always give students a precise picture of what aspects of their work has been strong or weak (Mowl, 1996). Marks rank and classify students, indirectly discriminating between "good" and "bad" ones, leading to competition among them, turning learning into a threatening experience and affecting children's motivation to learn. As Smith and Rottenberg (1991, p. 10) stress, tests "cause stress, frustration, burnout, physical illness, misbehaviour and fighting, and psychological distress".

Alternative assessment, on the other hand, is *criterion-referenced*, as the students' performance is not compared to each other's "but to a set of criteria of expected performance or learning targets" (Cameron, 2004, p. 223), reported in the form of a qualitative, descriptive, profile (Lynch, 2001). The rationale behind criterion-referencing is student motivation and encouragement. According to Kane et al (1997, p. 201) "students exhibit a greater motivation to learn and a greater amount of engagement with performance tasks and portfolio assignments than with other types of assignments".

Self-assessment: literature review

One of the key concepts of educational systems claiming to be student-centered is the *active involvement* of students in the assessment of their own progress which assists in the development of their critical self-consciousness within the learning process (Nunan, 1988). Learners who are active in taking initiatives learn more things and better than people who rely on their teachers (Shepard, 2000). Self-assessment drives towards that direction since it produces learners who are more active and focused and better placed to assess their own progress in terms of communication (Harris, 1997).

Students who learn to assess their own work move "from being "other-regulated" to being "*self-regulated*" or *autonomous*" (Cameron, 2004, p. 235), able to monitor their own performance, evaluate their progress, control their learning and decide how to use the resources available to them within or outside the classroom (O'Malley and Valdez-Pierce,

1996). Brown and Dove (1993) report that through self-assessment students use higher levels of reflection through developing a questioning and self-analytic approach to their professional practice and engaging in deep rather than surface learning. This is particularly important for young learners who are in the initial stages of the development of their cognition and need constant assistance to further develop it.

Self-assessment has a *motivational effect*, which is essential when teaching young learners. Gardner (2000) claims that successful self-assessment breeds confidence which in turn enhances motivation. Additionally, Oscarsson (1989) highlights the motivational effect of self-assessment in terms of goal orientation which influences classroom activities to the direction which best serve the students' communicative goals. Being engaged in setting their own learning goals, students create their own level of pressure which results in a relaxed, anxiety-free learning atmosphere. Seeing that their goals have been attained, creates intrinsic motivation rather than extrinsic in the form of grades and general praise (Dragemark, 2000). The effect self-assessment has on the establishment of a motivating learning atmosphere and positive attitudes also derives from the fact that teachers really listen to the students and the content of what they say.

Moreover, through self-assessment procedures the teacher can figure out and observe what the student actually understands while at the same time a "dialog" occurs between them. Unlike the traditional assessment approaches which have been unidirectional, self-assessment allows for a bi-directional flow of information in which both teacher and student are involved in the progress of the student's learning (Donato, 2000). Last but not least, self-assessment is advantageous to the teachers themselves. By participating in their own evaluation, students share the assessment burden with the teachers reducing the teachers' workload and freeing them to concentrate "on developing learning materials and giving help in other parts of the learning process" (Blue, 1988, p. 101). This is particularly useful in large classrooms where the teacher has to attend to a large number of students within a limited time. Besides, as the range of assessment techniques is expanded the learners broaden their range of experience within the realm of the assessment.

Self-assessment and speaking skills

The relationship between self-assessment and speaking skills is stressed by Underhill (1987) who includes self-assessment among the general types of oral tests. He claims that in real life we continuously assess how successful our communication is by listening to ourselves when we speak, watching the effect our speaking has on the interlocutors and by their replies. However, this self-assessment is unconscious, since in real communication we do not have time to consciously monitor ourselves. Similarly, Ellis and Sinclair (1989) mention that students should be enabled to monitor their own spoken language for problems and disfluencies as they will need to do so in real-life situations when they no longer have a teacher to rely on. All learners are able to ascertain the degree of their oral proficiency within certain limits and what they need is the experience which derives from training in monitoring and assessing themselves (Underhill, 1987).

Methodology

Aims and research questions

Undertaking this particular study, the intention was to investigate the possible benefits of self-assessment for primary school learners and the difficulties they might face while

carrying out self-assessment. The study also aimed at revealing the extent to which students can benefit with regards to their speaking skills. The research questions posed were:

- What is the students' reaction to self-assessment?
- What are the benefits to be gained from self-assessment concerning speaking skills?
- What difficulties do students encounter when they carry out self-assessment?

To answer the above research questions I gleaned data from a questionnaire distributed to students on completion of the study. This constituted the *quantitative instrument* of the study. Along with that, observation notes were also used throughout the study, which constituted the *qualitative instrument*.

The study tools

The format used in constructing the questionnaire was that of *closed items* arranged in a scale. Students had to tick a "smiley", a "neutral" or a "sad" face corresponding to the statements "I strongly agree", "I do not agree completely", "I totally disagree" respectively (Appendix 1). The closed items related to the benefits of self-assessment for the students including the usefulness of the self-assessment checklist, the discussion they had with the teacher after each lesson, goal-setting as well as the usefulness of peer-assessment. Besides, there were statements depicting the students' perception of the progress made in speaking and statements referring to their perceived difficulties in carrying out self-assessment. Additionally, there was an *open-ended item* comprising four sub-statements where students had the opportunity to elaborate more on the progress they made in speaking, their reaction to self-assessment and the difficulties they faced carrying out self-assessment. Considering the students' language level, both the instructions and the statements were in Greek, the native language of the students and the wording of the statements was as simple as possible. The questionnaire was anonymous to encourage students to answer as sincerely and freely as possible.

In conjunction with the students' questionnaire, classroom observations, namely "procedures for recording classroom events in a way that can be studied" (Allwright, 1988: xvi), were used in order to gather data regarding the progress students made in speaking. These observations were used due to the fact that recording students' interactions was impractical for all the pairs or groups of students. Actually these were *real-time observations*, taking place without using any electronic means of recalling the data but through taking notes on pre-determined aspects of the students' progress in speaking (Wallace, 2000). In particular, through the observations I looked into the proportion of mother tongue and target language use, the quality of target language use (complexity of structures, length of utterances, use of formulaic expressions, range of expressions used to express opinion, agree and disagree) as well as fluency.

The participants of the study

The study took place in the sixth-grade class of the 7th Primary School of Arta, a city in the north west of Greece which consisted of eighteen monolingual, Greek-speaking students. The class consisted of eight girls and ten boys, aged eleven to twelve, all of whom shared the same national and socio-cultural background. They had already been taught English for three years at school. Despite homogeneity in terms of age and the time spent studying English at school, the class was actually of a mixed-ability profile ranging from elementary to pre-intermediate, since the majority attended extracurricular English classes at different levels. The activities in the coursebook currently in use (Fun Way English 3) usually restricted

speaking to developing accuracy and generating discourse up to the sentence level. Regarding assessment, I had never systematically kept any records of how the students' language was progressing nor had students been involved in self-assessment procedures before. Teacher-made tests assessing Grammar and Vocabulary constituted the core assessment procedures.

The study started in January 2007 and extended over a period of three months during which the students participated in six speaking activities each followed by self-assessment. Special attention was taken not to deviate from the weekly, three-hour schedule. The activities were thematically or linguistically linked to the coursebook material. They were conceptually appropriate and visually attractive for the particular age group of students and they were constructed according to criteria underpinning effective teaching and assessment activities (Williams, 1991), approximating real-life, authentic tasks, stimulating genuine interaction and eliciting authentic communication. They included role-plays, problem-solving tasks, group discussions and communication games capturing a range of sub-skills and styles and allowing students to demonstrate their full potential in speaking.

After each activity the students filled in a self-assessment checklist which usually took them five to ten minutes to complete. This checklist was divided into two parts. In the first part students had to rank their preferences by ticking a "smiley", a "neutral" or a "sad" face, to show the degree to which they agreed with the written statements. These statements pertained to the students' attitude towards the activity and co-operation with their partners, as well as the extent to which they displayed the speaking skills we intended to assess. In the second part, I asked students to answer some questions in the hope that they would elaborate on the statements in the first part and consequently provide me, as the classroom teacher, with even more information that would be useful to instructional practice.

Results

Questionnaire

The findings of the questionnaire demonstrated that 13 students felt that self-assessment was a positive experience for them, whereas only 5 strongly agreed that they would like to do self-assessment in every lesson. 10 students strongly agreed that the self-assessment checklist was useful in helping them focus on the points they needed in order to improve their speaking skills. All the students found the discussion with the teacher after the lesson useful, while 12 of them said that they participated more in the lesson after having set a goal and trying to achieve it. 13 students felt that self-assessment was more enjoyable than tests and 15 said that they were less stressed when engaged in self-assessment comparing to taking tests. 16 students felt they benefited from engagement in peer-assessment and 14 that peer-assessment helped them think and understand the assessment criteria better.

Regarding the progress students made in speaking, 13 students strongly agreed that self-assessment helped them realize their strengths and weaknesses in speaking and 10 felt more confident in speaking. 12 students improved their speaking skills. 14 students became aware that speaking is more than answering the teacher's questions, they could participate in a conversation, handle turn-taking, initiate and close a conversation. However, only 6 students could keep the conversation going, even when they did not have the necessary vocabulary at their disposal, through employing paraphrase. Concerning the difficulties students faced when they had to assess themselves, 12 students disagreed that it was easy

to self-assess and 10 also disagreed that it was easy to understand corrections when speaking.

The students' responses to the open-ended items were qualitatively analyzed and salient features were identified to examine correspondence to the findings of the closed items. This would provide further information about the students' reaction to self-assessment, their progress in speaking and the difficulties they might have faced. There was a frequency count for the responses to indicate the number of students who made them. The students' comments showed that 15 of them felt that their speaking skills improved and their self-confidence in speaking increased while 6 students also admitted that they enjoyed participating in speaking activities when self-assessment was implemented. Eight students claimed that they improved handling turn-taking, initiating and closing a conversation and they do not hesitate so much when it is their turn to speak. Five students said that they can employ paraphrase thus keeping the conversation going. The majority said they liked self-assessment for four main reasons. 6 students liked self-assessment because they co-operated with their peers, 5 students because it helped them realize their strengths and weaknesses and try to improve by setting a goal to achieve and 4 students because they could assess themselves on their own. The difficulties they faced when self-assessing regarded the ability to elaborate in writing what they could or could not do in each activity (10 students), to set specific goals (6 students) while 4 students mentioned that they needed more support by the teacher.

Observation notes

Observing students in action I noticed that initially they resorted to mother tongue when they did not have the necessary vocabulary. Also when their partners did not cope with them, they used mostly short utterances, simple structures, limited language to express opinion, agreement or disagreement and they did not always use interrogative forms correctly. For example some students sometimes read the information on their role cards using a questioning intonation e.g. "*Name*?" "*Place of birth*?*"*. Most students had problem with Wh-questions. For example one student asked "*Who live with*?". Gradually, however, they were able to use the target language almost exclusively. Towards the end of the study they could also employ longer utterances, more complex structures and a range of formulaic expressions. They were also able to express opinion, agreement or disagreement in a range of ways. Additionally, they employed fillers, false starts and repetition which characterize natural conversation. For example, one student said "*Well in my opinion this is picture one because I can see … I can see two boys … they are walking and they are talking and … they they are near a hotel and .. I think they will climb because they have got mm … I don't know this … when we climb*". Besides, most of them improved their ability to use the interrogative form correctly and they could speak without much hesitation. For example, one student asked "*where do you live*"?

Discussion

The majority of the students appreciated the implementation of self-assessment as an assessment tool despite the fact that they had not had any previous experience in self-assessment either in English or Greek school subjects. The students' overall reaction to the implementation of self-assessment was favourable which does not support Boud's (1988) reservations that inexperienced learners may be resistant to the method of self-assessment in the early stage of its implementation. The fact that most students agreed with the statements concerning the benefits deriving from self-assessment suggests that its use as an

assessment tool was vital for them, while at the same time it confirms its appropriacy for the assessment of young learners. The students found the self-assessment checklists beneficial as they provided students with a useful guide, directing their attention on specific aspects when they were assessing their work. This finding asserts the need for a useful assessment tool to assist students in self-assessment.

The feedback provided either by me or the peers in the end of each activity was also positive for the learners supporting the view that both instructor and peer feedback can highlight points the student missed, discover gaps in the student's analysis, provide other perspectives from which to view performance and raise questions that might lead to further understanding (Loacker, 2003). Both self- and peer-assessment were perceived to have contributed positively in developing critical thinking as students were involved in identifying what, how and why they did something as well as expressing their opinion about their peers' achievement. Noticing the strengths and weaknesses of their peers' work helped them better reflect on their own work, supporting Tudor's (1996: 182) stance that critical reflection on the abilities of other learners with respect to a shared goal helps individuals to assess their own performance more effectively.

The majority of students agreed that they had less stress during self-assessment than during traditional testing. This finding supports the belief that self-assessment as an alternative assessment method can create an anxiety-free environment (Shaaban, 2001). Moreover, direct student involvement in self-assessment activities, setting a goal and trying to achieve it resulted in greater participation in the activities, persistence in their completion and interest in learning English, confirming the view that engagement in self-assessment enhances the students' intrinsic motivation (Oscarsson, 1989).

Self-assessment was beneficial as it helped increase the students' self-esteem and confidence in speaking by making them aware of their strengths and capabilities. Despite the short duration of the study, students developed their speaking skills and increased the quantity and quality of their spoken language. Speech quantity increased through the frequent use of pair- and group-work which allowed each student more time to speak in the target language. Speech quality increased as students gradually produced more extended and complex discourse, used formulaic expressions, spoke without hesitating much, thus developing both fluency and accuracy. Moreover, students became aware of the fact that speaking means not only giving accurate answers to the teacher's questions but most importantly the ability to handle turn-taking and managing the interaction effectively while participating in realistic conversations.

Nevertheless, the findings also indicated that students faced certain difficulties when trying out self-assessment, which might be the reason for their reluctance to use self-assessment on a daily basis. The main difficulties they faced were to reflect and comment on their strengths and weaknesses when they were asked to elaborate more on what they had achieved on each activity, namely to be involved in reflective and critical thinking and to set specific goals. This can be justified as it was the first attempt of the particular students to monitor their progress and assess themselves. Besides, the number of times they were involved in self-assessment in the classroom was relatively small which is taken to mean that their experience was rather limited. The implications of these findings are that the students need time, support and training to get used to this assessment method.

Limitations of the study

The first limitation relates to the short duration of the study and the nature of the sample as this was a small-scale study conducted with only one class of primary school students. As a result the findings may not allow for generalizations to be made. Consequently, further investigation on a larger scale with longer duration is necessary. The second limitation derives from the tools employed, namely questionnaires and observation notes. Regarding questionnaires, there is the possibility of subjectivity, since "when we ask someone a question, we have very often no way of ascertaining the truth of the reply" (Wallace, 2000: 127). Also the students might over-report on the advantages of self-assessment and their progress in speaking due to the fact that the researcher was also the class teacher hence the students may have been cautious not to criticize the implemented assessment method. Observation notes run the risk of being subjective and biased as the focus of observation was pre-determined instead of allowing it to emerge (Bell, 2000) through observation.

Recommendations for further research

Based on the findings of this study, some recommendations for further research can be made:
- Conducting a similar research with a larger sample of participants in order to strengthen the reliability of the results of the present research.
- Integrating self-assessment into the EFL classroom, extending the duration of implementation from the beginning to end of the school year in order to investigate its potential as a learning tool, too.
- Extending self-assessment to lower school grades and compare the findings.
- Training teachers in the effective use of self-assessment and involving them in research, in order to investigate their reactions and beliefs regarding the implementation of self-assessment in their classrooms.

Author's email: chalkiak@yahoo.gr

References

Allwright, R. (1988). *Observation in the language classroom*. London: Longman.

Bell, J. (2000). *Doing your research project*. Buckingham/Philadelphia: Open University Press.

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: 345-359.

Blue, G. M. (1988). 'Self-assessment: the limits of learner independence'. *ELT Documents,* 131: 100-118.

Boston, C. (2002). 'The concept of formative assessment'. *Practical Assessment, Research and Evaluation,* 8/9 (September), accessed http://www.vtaide.com/png/ERIC/Formative-Assessment.htm on 27 Sep 2006.

Boud, D. (1988). *Developing student autonomy in learning*. London: Kogan Page.

Brown, S. & Dove, P. (1993). 'Self and peer assessment: Learning from experience'. *Natfhe Journal.*

Cameron, L. (2004). *Teaching languages to young learners*. Cambridge: Cambridge University Press.

Dickinson, L. (1987). *Self-instruction in language learning*. Cambridge: Cambridge University Press.

Donato, R. (2000). 'Sociocultural contributions to understanding the foreign and second language classrooms'. In Lanfolf, J., P. (Ed), *Sociocultural theory and second language learning*. Oxford: Oxford University Press, 27-50.

Dragemark, A. (2000). 'Self-assessment in language learning – self-assessment within the TENTEC Project'. In *Teaching English for technical purposes, Leonardo da Vinci project*, accessed at http://www.pedc.se/tentec/didactics/assessment.html on 24 Nov 2006.

Ellis, G. & Sinclair, B. (1989). *Learning to learn English*. Cambridge: Cambridge University Press.

Gardner, D. (2000). 'Self-assessment for autonomous learners'. *Links & Letters*, 7: 49-60.

Genesee, F. & Hamayan, E. (1994). 'Classroom-based assessment'. In Genesee, F. (Ed), *Educating second language children*. Cambridge: Cambridge University Press, 212-239.

Gronlund, E. & Cameron, I., J. (2004). *Assessment of student achievement*. Toronto: Pearson.

Harris, M. (1997). 'Self-assessment of language learning in formal settings'. *ELT Journal,* 51/1: 12-20.

Kane, M., B., Khattri, N., Reeve, A., L. & Adamson R., J. (1997). *Assessment of student performance.* Washington D.C.: Studies of Education Reform, Office of Educational Research and Improvement, U.S. Department of Education.

Loacker, G. (2003). 'Taking self assessment seriously'. *Essays on Teaching Excellence*, 15/2.

Lynch, B., K. (2001). 'The ethical potential of alternative language assessment'. In C. Elder (Ed.), *Experimenting with uncertainty: essays in honor of Alan Davies*. Cambridge: Cambridge University Press, 228-239.

Ministry of Education and Religious affairs and the Pedagogical Institute (2002). *Cross Thematic Curriculum Framework for Compulsory Education Diathematikon Programma.* Athens: OEΔB.

Montgomery, K. (2001). *Authentic assessment: a guide for elementary teachers*. New York: Longman.

Mowl, G. (1996): 'Innovative assessment'. *Deliberations*, accessed at http://www.city.londonmet.ac.uk/deliberations/assessment/mowl.cfm on 28 Sep 2006.

Nunan, D. (1988). *The Learner-centred curriculum*. Cambridge: Cambridge University Press.

O'Malley, M. & Valdez-Pierce, L. (1996). *Authentic assessment for English language learners*. New York: Addison-Wesley.

Oscarsson, M. (1989). 'Self-assessment of language proficiency: rationale and applications'. *Language Testing*, 6/1: 1-13.

Pedagogical Institute (2000): *Fun Way English (3) Teacher's book*. Athens: OEΔB.

Pedagogical Institute (2001): *Fun Way English (3) Student's book*. Athens: OEΔB.

Pedagogical Institute (2001): *Fun Way English (3) Workbook*. Athens: OEΔB.

Puhl, C., A. (1997). 'Develop, not judge: continuous assessment in the ESL classroom'. *English Teaching Forum,* 35/2: 2-9.

Rolheiser, C. & Ross, J., A. (2000). 'Student self-evaluation: what do we know?'. *Orbit,* 30/4: 33-36.

Shaaban, K. (2001). 'Assessment of young learners'. *English Teaching Forum,* 39/4 (April): 16-22.

Shepard, L., A. (1994). 'The challenges of assessing young children appropriately', accessed t http://www.cse.ucla.edu/products/parents/cresst_challengesshepard.pdf on 27 Sep 2006.

Shepard, L., A. (2000). 'The role of classroom assessment in teaching and learning'. *CSE Technical Report 517*. CRESST, University of Colorado, Boulder, accessed at http://datause.cse.ucla.edu/DOCS/las_rol_2000.pdf on 13 Oct 2006.

Smith, M., L. & Rottenberg, C. (1991). 'Unintended consequences of external testing in elementary schools'. *Educational Measurement: Issues and Practice,* 10/4: 7-11.

Tsagari, C. (2004). 'Alternative methods of assessment'. In West R. & Tsagari, C., *Testing and assessment in language learning.* Patras: Hellenic Open University, 119-334.

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

Underhill, N. (1987). *Testing spoken language: a handbook of oral testing techniques*. Cambridge: Cambridge University Press.

Wallace, M., J. (2000). *Action research for language teachers*. Cambridge: Cambridge University Press.

Williams, M. (1991). 'A framework for teaching English to young learners'. In C. Brumfit, J. Moon & R. Tongue (Eds.), *Teaching English to children: from practice to principle*. London: Collins, 203-212.

APPENDIX 1

THE QUESTIONNAIRE

<div style="text-align:center">

STUDENTS' QUESTIONNAIRE

</div>
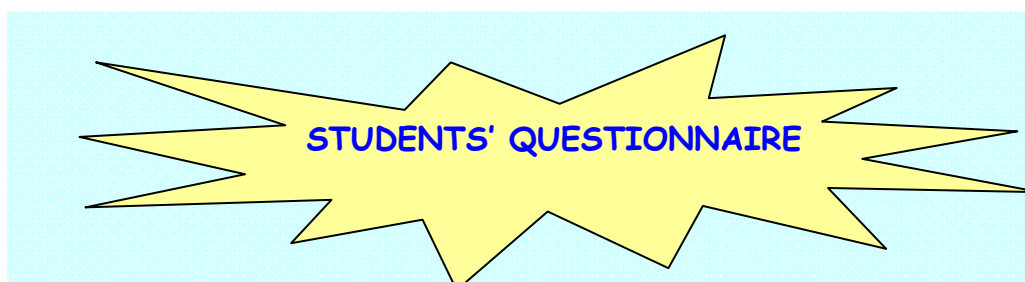
Dear student,
This year you participated for three months in a self-assessment programme in which you had the opportunity to use a self-assessment checklist with a list of specific criteria in order to think about your strengths and weaknesses in various speaking activities during the English lesson. Now that the program has finished, I would like you to answer carefully to the following questionnaire which concerns your opinion on self-assessment. Try to be honest. Thank you very much for your help.
Your teacher.

Read the following statements carefully and tick ☑ the box with which you agree the most.

☺ = I totally agree.   😐 = I don't completely agree.   ☹ = I disagree completely.

	☺	😐	☹
1) Self-assessment was a positive experience for me.			
2) I would like to have self-assessment in every lesson in the English class.			
3) The self-assessment checklist was useful because it helped me focus on what was important in order to improve my speaking skills.			
4) The discussion I had with the teacher after each lesson concerning what I had stated in the self-assessment checklist was useful.			
5) The fact that in the end of each lesson I set a goal to achieve in the next lesson helped me try more every time to achieve this goal and in this way I participated more in every speaking activity.			
6) Self-assessment is more enjoyable than tests.			
7) I had less stress during the self-assessment procedure than I had during a test.			
8) Peer-assessment helped me learn from my classmates' strengths and weaknesses.			
9) Peer-assessment helped me think and understand better the assessment criteria which were important in order to improve my speaking skills.			

236

10) Self-assessment helped me realize my strengths and weaknesses in speaking.			
11) After my participation in the self-assessment program, I feel more confident in speaking.			
12) Being involved in self-assessment I improved my speaking skills.			
13) Now I know that speaking English means not only answering the teacher's questions.			
14) I can participate in a conversation.			
15) I can talk with someone and understand when it is my turn to speak.			
16) I can initiate and close a conversation.			
17) I can keep a conversation on even when I do not have the necessary vocabulary at my disposal.			
18) I can employ paraphrase.			
19) It was easy for me to self-assess.			
20) It was easy for me to understand what I did correct when speaking.			

21) Please answer the following questions: a) do you think you improved your speaking skills? b) what aspects of your speaking skills do you think you improved? c) did you like self-assessment and why? d) what difficulties did you faced during self-assessment?

a)…………………………………………………………………………………………
………………………………………………………………………

b)…………………………………………………………………………………………
………………………………………………………………………

c)…………………………………………………………………………………………
…..……………………………………………………………………

d)…………………………………………………………………………………………
…………………………………………………………………………

ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ ΜΑΘΗΤΩΝ

Αγαπητέ μαθητή/Αγαπητή μαθήτρια

Φέτος συμμετείχες για τρεις μήνες σε ένα πρόγραμμα αυτοαξιολόγησης στο οποίο είχες την ευκαιρία να χρησιμοποιείς λίστες για να σκεφτείς ποιες ήταν οι δυνατότητές σου και οι αδυναμίες σου σε διάφορες προφορικές ασκήσεις στο μάθημα των Αγγλικών. Τώρα που το πρόγραμμα τελείωσε, θα ήθελα να απαντήσεις προσεκτικά στο παρακάτω ερωτηματολόγιο που αφορά στην γνώμη σου για την αυτοαξιολόγηση. Προσπάθησε να είσαι ειλικρινής. Σε ευχαριστώ πολύ για την βοήθειά σου.
Η δασκάλα σου.

Διάβασε προσεκτικά τις ακόλουθες προτάσεις και τσέκαρε ☑ το κουτάκι με το οποίο συμφωνείς περισσότερο.

☺ = Συμφωνώ απόλυτα   😐 = Δεν συμφωνώ εντελώς.   ☹ = Διαφωνώ απόλυτα.

	☺	😐	☹
1) Η αυτοαξιολόγηση ήταν μια θετική εμπειρία για μένα.			
2) Θα ήθελα να κάνω αυτοαξιολόγηση σε κάθε μάθημα στα Αγγλικά.			
3) Η φόρμα με τα κριτήρια αυτοαξιολόγησης ήταν χρήσιμη γιατί με βοηθούσε να επικεντρωθώ στα σημεία που ήταν σημαντικά για να βελτιώσω τις προφορικές μου ικανότητες.			
4) Η συζήτηση που έκανα με την κυρία στο τέλος κάθε μαθήματος σχετικά με όσα είχα αναφέρει στη φόρμα αυτοαξιολόγησης ήταν χρήσιμη.			
5) Το γεγονός ότι στο τέλος του μαθήματος έβαζα ένα στόχο για το επόμενο μάθημα με βοήθησε στο να προσπαθώ περισσότερο κάθε φορά για τον πετύχω, συμμετέχοντας περισσότερο σε όλες τις ασκήσεις.			
6) Η αυτοαξιολόγηση είναι πιο ευχάριστη από τα τεστ.			
7) Είχα λιγότερο άγχος όταν συμμετείχα στην αυτοαξιολόγηση από ότι έχω όταν γράφω τεστ.			
8) Η αξιολόγηση των συμμαθητών μου με βοήθησε να μάθω από τις δυνατότητες και τις αδυναμίες τους.			
9) Η αξιολόγηση των συμμαθητών μου με βοήθησε να σκεφτώ και να καταλάβω καλύτερα τα κριτήρια αξιολόγησης που ήταν σημαντικά για να βελτιώσω τις προφορικές μου ικανότητες.			
10) Η αυτοαξιολόγηση με βοήθησε να καταλάβω τις δυνατότητές μου και τις αδυναμίες μου στα προφορικά.			
11) Μετά τη συμμετοχή μου στο πρόγραμμα αυτοαξιολόγησης αισθάνομαι μεγαλύτερη αυτοπεποίθηση στα προφορικά.			
12) Με την αυτοαξιολόγηση βελτίωσα τις προφορικές μου ικανότητες.			
13) Τώρα ξέρω ότι το να μιλώ Αγγλικά δεν είναι μόνο να απαντώ στις ερωτήσεις της κυρίας.			
14) Μπορώ να πάρω μέρος σε μία συζήτηση.			
15) Μπορώ να συνομιλώ με κάποιον και να καταλαβαίνω πότε είναι η σειρά μου να μιλήσω.			
16) Μπορώ να αρχίζω και να τελειώνω μια συζήτηση.			

17) Μπορώ να συνεχίζω μια συζήτηση ακόμη κι αν δεν έχω το απαραίτητο λεξιλόγιο.			
18) Μπορώ να λέω κάτι με άλλα λόγια.			
19) Ήταν εύκολο να αξιολογήσω τον εαυτό μου.			
20) Ήταν εύκολο να καταλάβω τι έκανα σωστά.			

21) Σε παρακαλώ απάντησε στις ακόλουθες ερωτήσεις: α) νομίζεις ότι βελτιώθηκες στα προφορικά; β) τι δεν μπορούσες να κάνεις πριν και βελτίωσες τώρα; γ) σου άρεσε η αυτοαξιολόγηση και γιατί; δ) τι δυσκολίες είχες στο να αξιολογήσεις τον εαυτό σου.

α)……………………………………………………………………………….
………………………………………………………………………
β)…………………………………………………………………………………
………………………………………………………………………
γ)…………………………………………………………………………………
………………………………………………………………………
δ)…………………………………………………………………………………
………………………………………………………………………

The impact of training adolescent EFL learners on their perceptions of peer assessment of writing

[Η επίδραση της εκπαίδευσης των εφήβων σπουδαστών αγγλικής γλώσσας στις αντιλήψεις τους σχετικά με την έτερο-αξιολόγηση του γραπτού λόγου]

Elena Meletiadou

Peer assessment (PA) has been increasingly used as an alternative method of engaging learners in the development of their own learning. However, very little research has been conducted in the Cypriot and Greek educational systems. This paper describes part of a research project conducted in a Cypriot State EFL Institute. Forty adolescent 'English as a Foreign Language (EFL)' students were involved in PA of writing in an attempt to improve their writing performance and attitudes towards the assessment of writing. Learners received training since they had no PA experience prior to the study. The students' attitudes were canvassed both prior to the PA training and at the end of it by means of a PA questionnaire. The findings showed that students' response to PA was: (a) negative before the training, and b) positive after the training. This study concludes that PA is an innovative method and students have to be given time, training and support to adapt to it, in order to perform to the best of their ability and exploit its full potential.

಄

Η ετερο-αξιολόγηση χρησιμοποιείται όλο και περισσότερο σαν εναλλακτική μέθοδος της εμπλοκής των μαθητών στην ανάπτυξη της μάθησης τους. Πολύ λίγη έρευνα όμως έχει διεξαχθεί στο Κυπριακό και Ελληνικό εκπαιδευτικό σύστημα. Το παρόν άρθρο περιγράφει μέρος ενός ερευνητικού προγράμματος που πραγματοποιήθηκε σε ένα Κυπριακό Κρατικό Ινστιτούτο εκμάθησης της Αγγλικής ως ξένη γλώσσα. Σαράντα έφηβοι σπουδαστές της Αγγλικής ως ξένη γλώσσα πήραν μέρος στην διαδικασία έτερο-αξιολόγησης γραπτού λόγου σε μια προσπάθεια να βελτιωθεί η απόδοση τους και η στάση τους απέναντι στην αξιολόγηση του γραπτού λόγου. Οι μαθητές εκπαιδεύτηκαν εφόσον δεν είχαν σχετική εμπειρία πριν από την εμπλοκή τους στην διαδικασία της έτερο-αξιολόγησης. Διερευνήθηκαν οι στάσεις των μαθητών πριν και στο τέλος της εκπαίδευσης τους στην διαδικασία της έτερο-αξιολόγησης με την χρήση ενός σχετικού ερωτηματολόγιου. Τα αποτελέσματα έδειξαν ότι η ανταπόκριση των μαθητών στην έτερο-αξιολόγηση ήταν: α) αρνητική πριν την εκπαίδευση τους, και β) θετική μετά την εκπαίδευση. Αυτή η μελέτη

καταλήγει στο συμπέρασμα ότι η έτερο-αξιολόγηση είναι μια καινοτόμος μέθοδος και πρέπει να δοθεί χρόνος, εκπαίδευση και υποστήριξη για να προσαρμοστούν οι μαθητές σ' αυτήν, να μπορέσουν να βελτιώσουν ικανοποιητικά την απόδοση τους και να επωφεληθούν από αυτήν στο μέγιστο βαθμό.

Key words: peer assessment, alternative, writing, secondary education, attitudes, training.

Introduction and literature review

Due to the growing focus on learner autonomy in learning and assessment (Falchikov, 1986; Lynch, 1988), peer assessment (PA) has received a lot of attention in recent years. Nevertheless, this method is novel to most English language teachers and students in Cyprus and Greece where traditional assessment is still dominant. According to McDowell (1995), PA is one form of innovation which aims to improve the quality of learning and empower students in contrast to more traditional methods which can leave learners feeling disengaged from the overall assessment process. Moreover, PA: (a) encourages students to take responsibility for their own learning and development, (b) treats assessment as part of learning so that mistakes are seen as opportunities rather than failures, and (c) practises the transferable skills needed for life-long learning particularly related to evaluation skills (Donaldson & Topping, 1996).

Numerous studies evaluate the perceptions and attitudes of the learners towards the PA process. Many researchers report that learners have a positive attitude towards the evaluation of the written work of their peers (Race, Brown & Smith, 2005; Wen & Tsai, 2006). PA increases students' interest in the English language lesson by encouraging self-regulated learning (Ten Berge & Hofstee, 2004). Ballantyne, Hughes, and Mylonas (2002) point out that students commonly report that assessing the work of their peers: (a) can be personally motivating, (b) aid knowledge and understanding of subject content, and (c) help their learning. Peer feedback, a basic feature of PA, has an impact on affect, e.g. increases motivation through the sense of personal responsibility, reduces writing anxiety and improves self-confidence (Topping, 2000). Finally, Brown, Race and Smith (1997) report that resistance by students to informal peer feedback is rare.

Nevertheless, there are studies which indicate the opposite. Student writers may not always trust their peers when the same comment from a teacher will be taken into account (Strijbos, Narciss. & Dunnebier, 2010). In some studies, students became defensive and expressed discomfort (Papinczak, Young & Groves, 2007) and uneasiness about acting like a teacher (Orsmond, Merry & Reiling, 1997). Topping, Smith, Swanson and Elliot (2000) found that most students considered the PA process as time consuming and socially uncomfortable although it was effective in improving their learning. Other researchers actually found increased opposition to PA after student exposure to it (Rushton, Ramsey & Rada, 1993). Finally, students felt uncomfortable in awarding grades seeing it as "risky and unfair" and simply preferred to give feedback (Boud, 2000).

In conclusion, the literature has shown mixed findings regarding learners' attitudes towards PA especially when these received no prior training in PA (Sluijsmans, Moerkerke, Merrienboer & Dochy, 2001). Moreover, most research in PA was conducted with adult learners (Jones & Fletcher, 2002). It would, therefore, be interesting to explore adolescent EFL learners' attitudes towards PA of writing before and after training learners in PA

methods. The aim is to identify whether preparing learners for using PA in their classrooms can have an impact on students' attitudes towards PA.

Rationale for the current study and research questions

According to previous research (Meletiadou, 2011), Cypriot adolescent EFL learners have a negative attitude towards writing and the assessment of writing (Meletiadou, 2011). Peer assessment is one of the most popular tools for 'assessment for learning' currently adopted in education (Falchikov, 2004; Topping, Walker & Rondriguez, 2008). However, very little research has been done in the area of PA performed by adolescent learners (Tsivitanidou, Zacharia & Hovardas 2011). There is also a need for more research into this new method of assessment in relation to the impact of training students in PA methods on their attitudes (Cheng & Warren, 1997). In an attempt to explore the potential of using PA to support adolescent student achievement in EFL writing and to improve students' attitudes towards writing and the assessment of writing, a study was conducted in order to find answers to the following research questions: (1) What are adolescent EFL learners' perceptions of PA of writing? (2) Does training adolescent EFL learners prior to implementing PA of writing improve their attitudes towards PA of writing?

Subjects

The study involved two groups of 20 adolescent learners who had attended EFL classes for the past five years at a local State Institute in Cyprus. All students were provided with training in PA before using it in class. The learners had to write three essays (an informal letter, a descriptive and a narrative essay) in two drafts according to the demands of their curriculum (Ministry of Education, 2010). Both groups received teacher feedback, while Group A (student/assessees) also received peer feedback from group B (student/assessors) using an analytic rating scale. The teacher, who was also the researcher of this study, was a qualified EFL teacher with several years of experience and a postgraduate degree in 'Teaching English as a Foreign Language (TEFL)'.

Instruments

The researcher employed two instruments in order to address the research questions. These were: a) a PA questionnaire and b) a PA form.

The PA questionnaire

Students' attitudes were monitored before and after the PA training by means of a questionnaire (Table 1) adapted by Cheng and Warren (1997) and developed for lower secondary EFL students.

All statements required learners to respond using a five-point Likert type scale ranging from "strongly agree" to "strongly disagree. This questionnaire aimed to elicit students' reaction to PA before and after having experienced it. The questionnaires were administered during class time immediately before and after completing the training so that the students could easily recall and express their opinions. The aim was to detect any differences in students' attitudes towards PA after they received training in this 'alternative' assessment method.

Taking into consideration the students' language level, the statements were presented both in Greek and English. The language and the wording of the statements were also intended to

be simple so as to correspond to students' age and cognitive abilities. Besides that, the questionnaire was anonymous so as to encourage students to respond as sincerely and freely as possible.

	Peer Assessment Questionnaire	SD	D	DK	A	SA
1	Students should take part in assessing their peers.	1	2	3	4	5
2	An adolescent student can provide reliable marks to his/her peer.	1	2	3	4	5
3	I feel comfortable when I assess my peers.	1	2	3	4	5
4	I think I can make a fair and responsible assessment of my peers.	1	2	3	4	5
SD: Strongly disagree, D: Disagree, DK: Don't know, A: Agree, SA: Strongly agree.						

Table 1. Peer assessment questionnaire

The first statement referred to the issue of students' participation in PA. Research has shown that adolescent students often feel that they should take part in PA (Meletiadou, 2011). PA also helps students reflect upon their writing performance (Brown, 1998) and learners benefit both by providing (peer assessors) and/or receiving PA (peer assessees) (Topping, 2010). Students realize their strengths and weaknesses, become better-organized (Brown & Hudson, 1998) and follow up with actions to improve their work (Boud, 1995).

The second statement referred to the issue of the reliability of adolescent student-generated marks. It aimed at validating researchers' claim that PA is only suitable for adult learners because it is very demanding (Brown & Dove, 1991) and that adolescent students cannot provide reliable marks (Chang, Tseng, Chou & Chen, 2011). The third statement was related to students' feelings when they assess their peers. Research has shown that students may feel frustrated or reluctant when they assess their peers (Sluijsmans et al., 2001). Finally, the fourth statement raised the issue of fairness of PA. It also investigated the capacity of peer assessment to make students feel responsible for their own and others' learning since they were asked to assume the role of the teacher (Papinczak et al., 2007).

The PA form

The PA form (Table 2) was devised by the student/assessors (Group B) with the help of their teacher during the training sessions.

The PA form was a rather controlled type of rubric in the form of a checklist. Nevertheless, it provided student/assessees (Group A) with marks and feedback which play an important role to students' educational development (Black & William, 1998). It was intended to make things easy both to student/assessors and to student/assessees who could be asked to provide their own comments instead of simply ticking a list in the future.

The form was used to guide student/assessees in reading and revising their own texts and to guide the peer-response activity. Being simple and 'procedural' in nature, it was expected to provide the learners with basic guidelines for giving feedback to peers' drafts and

consequently approaching their own drafts critically and revising them more effectively (Johns, 1986).

Criteria/Weighting	Excellent-Very Good	Good-Average	Fair-Poor	Very Poor
A. 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?				
B. 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.				

Table 2. Peer assessment form

Taking this form into consideration, student/assessors (Group B) had to read the drafts of the other group (Group A) carefully and tick accordingly. They were then able to quickly view the completed form, assign marks to each one of the criteria and calculate the total score.

The training phase

Supporting learners in using PA is of paramount importance because this is an activity in which learners need guidance and time to grow into. Approaching PA step by step helps reduce student concerns, build their self-confidence and gain the necessary experience. Learners need to build up a shared understanding of the nature, the purposes and the requirements of the PA method (Stewart & Cheung, 1989).

In January 2010, the researcher prepared a PA training session for the students. The main purpose of the session was to make decisions about and establish the assessment criteria. It lasted about three class hours and comprised a number of different stages (Table 3).

Groups	Group A	Group B
Propaganda phase	✓	✓
Revision strategies	✓	✓
Revision with peer assessment form	✓	
Model texts	✓	✓
Mock/rating, commenting		✓
Discussion	✓	✓

Table 3. Training sessions for learners

During the 'propaganda phase', all learners were taught revision strategies with guided tasks and were involved in a brief discussion regarding PA of writing. Student concerns were also discussed and reasons were provided why peers at the same level can give helpful feedback. The teacher explained to student/assessors (Group B) in particular, that in order to assess something, the most crucial steps were to distinguish what they were going to assess and design a set of criteria in order to do the assessment. The researcher chose to employ explicit student owned criteria because, according to research (Falchikov, 1986), these seem to enhance the overall reliability of PA by increasing the correlation between teacher and student marks. These also had to be as simple as possible in order to correspond to students' age and cognitive abilities. As a result, learners created a PA form (Table 2) with the help of the teacher, which was used during this study to provide feedback for their essays. Namely:

- Student/assessors (Group B) completed the PA form for all student/assessees' (Group A) drafts of all essays, and
- The teacher provided a mark and comments to all drafts of all essays keeping in mind the PA form.

Furthermore, student/assessees (Group A) were asked to revise three samples of other students' drafts (an informal letter, a descriptive essay and a narrative) together with the completed PA forms in groups of three. Students' revisions were then discussed in class so as to prepare both groups for revising their own drafts. Student/assessors (Group B) were presented with three samples of students' compositions and were asked to mock rate/comment on them in groups of three using the rating instrument. Their ratings and any significant differences with the teacher's ratings and comments were discussed in class to clarify any misunderstandings. Sample essays that presented errors in all areas of the PA form (i.e. Mechanics) were used.

Findings

To answer the research questions, the teacher administered the PA questionnaire (Table 1) to all students before and after the training session in PA. The findings are now going to be presented in the same order as the statements of the questionnaire.

Firstly, the learners' answers to the first statement of the questionnaire revealed that students believed that they should not take part in assessing their peers (Table 4) before the

training sessions. They were rather reluctant (50%) to get actively involved in this demanding procedure without being adequately prepared for it since they were totally inexperienced in PA. Only 25% of the students seemed to be interested in experimenting with this new method and another 25% were rather uncertain. Training the learners in PA had an enormous impact on learners' attitudes towards PA. It helped them realize that they should take part in PA since this is an exciting procedure that promotes learning. As can be seen in Table 4, the majority of the learners seemed to have a positive attitude towards taking part in PA while only a few students (15%) still remained reluctant. Longer training sessions may have convinced even those hesitant learners to take an active part in this innovative method.

Statement 1: Students should take part in assessing their peers.			
Phase / Attitudes	Positive attitude	Neutral	Negative attitude
Pro-training	25%	25%	50%
Post-training	85%	-	15%

Table 4. Peer assessment questionnaire: Statement 1

Regarding the second statement (Table 5), most of the learners believed that adolescent students cannot provide reliable marks to their peers before the training session. They obviously did not know how this could be done. Only 35% of the students thought that they could evaluate their peers' writing skills reliably. Finally, a small part of the students (25%) were unsure. After receiving training, students seemed to have changed their disposition towards PA. They claimed that adolescent students can provide reliable marks to their peers (Table 5). Careful creation of the instruments employed seemed to enable these relatively young students to take part in this procedure quite successfully as was indicated during the training sessions. Moreover, involvement of the learners in the design of the criteria seemed to enhance their interest in PA.

Statement 2: An adolescent student can provide reliable marks to his/her peer.			
Groups/ Attitudes	Positive attitude	Neutral	Negative attitude
Group A	35%	20%	45%
Group A	85%	5%	10%

Table 5. Pro-training PA questionnaire: Statement 2

As can be seen in Table 6, it is obvious that most of the learners (90%) felt quite uncomfortable about assessing their peers because they had no previous experience. They were reluctant to experiment with this new type of assessment and take up the role of the teacher. After participating actively in the PA process during the training sessions, learners changed their attitudes radically (100% positive). This contradicts previous research which claims that learners are often reluctant when they are asked to assess their peers (Kwan & Leung, 1996). Students were obviously afraid to play the role of the teacher without adequate training. Moreover, they were reluctant to provide negative feedback to them in an attempt to avoid conflict. Training and anonymity of the learners in the present study resolved both problems and enhanced students' positive attitude towards PA.

Statement 3: I feel comfortable when I assess my peers.			
Groups/ Attitudes	Positive attitude	Neutral	Negative attitude
All groups	-	10%	90%
All groups	100%	-	-

Table 6. Pro-training PA questionnaire: Statement 3

Furthermore, most of the students were either negative (Table 7: 35%) or uncertain (45%) regarding their ability to assess fairly and responsibly their peers since they had never been involved in providing PA before. Only one fifth of the students were eager to experiment with this approach although they had no previous experience in using any form of alternative assessment. Training radically changed students' attitudes since learners seemed quite confident (80%) that they could fairly and responsibly assess their peers.

Statement 4: I think I can make a fair and responsible assessment of my peers.			
Groups/ Attitudes	Positive attitude	Neutral	Negative attitude
All groups	20%	45%	35%
All groups	80%	20%	-

Table 7. Pro-training PA questionnaire: Statement 4

To sum up, training in PA methods seemed to affect students' attitudes in a very positive way, improving their self-confidence and motivation. This was also confirmed by previous research (Topping, 1998). The very fact that a high percentage of the students thought of it as beneficial (Table 4) means that its use as an assessment tool was perceived as vital.

Discussion

The current study produced some very interesting findings regarding the effect of training on students' attitudes towards PA.

Students' responses in the pro-training questionnaire clearly indicated their reluctance to employ this innovative method without adequate preparation. There are a number of issues related to this attitude. First of all, most adolescent EFL learners in Cypriot State schools and Institutes are used to: (a) summative assessment, (b) a teacher-centred teaching style and (c) a product-approach to writing. Although the curriculum of the Cypriot Ministry of Education encourages the use of alternative methods, these are not actually employed in the EFL classrooms due to complete lack of teacher and student training (Meletiadou, 2011). Consequently, PA is an unfamiliar method and careful training and guidance are required in order to ensure successful implementation as various practitioners have pointed out (Digiovanni & Nagaswami, 2001; Jacobs et al., 1998; Porto, 2001). The process requires ongoing and repeated practice for students to become competent assessors (Sadler, 1989). Students may have reservations toward PA due to their own perceived inability of assessing their peers (Ellington, Earl & Cowan, 1997). This reservation usually comes from lack of experience or knowledge which might also cause students to feel stressful towards peer evaluation (Topping, 1998). It is unavoidable that students might be anxious when encountering a new assessment method especially when they are relatively young. To solve this, teachers should:

- explain to students that experiencing stress and anxiety is normal;
- make sure that students obtain some degree of satisfaction while their responsibility and power increases;
- become aware of students' past classroom experiences i.e. their total lack of experience in: (a) alternative assessment methods, (b) a process approach to writing, and (c) student-centred teaching styles as was the case in the present case study and their assumptions about language learning and assessment i.e. assessment is the teacher's job, and
- given opportunities for scaffolding by their teachers in the process to build their confidence and ability (Bassano, 1986).

In the present study, students' attitudes towards PA radically changed after the training sessions as was indicated from students' responses in the post training PA questionnaire. Learners felt that due to the training session, they were able to confidently and reliably assess the language proficiency of their peers (Cheng & Waren, 2005). They also had reservations about trusting their peers' comments at first and about their peers' ability to comment on subject areas they did not specialize in. Nevertheless, some time later during the training, they felt: (a) they benefited from peer response, (b) they acquired a reasonable grounding in PA procedures, (c) they were eager to take on the responsibility for their learning, (d) they turned into active participants who perceived that learning and assessing is a shared experience, and (e) they were favourably disposed to participating in PA in the future.

Given the results of the present study, it is reasonable to conclude that PA should be introduced into the curriculum gradually and in a consistent way while involving students in the design and development of the assessment criteria used for PA (Williams, 1992). These also need to be clarified and exemplified. Careful adaptation of the instruments to meet students' level can also allow the learners to participate actively in the assessment procedure.

The findings should also encourage teachers to utilize PA in their own EFL writing classrooms. EFL teachers are obviously concerned with improving their assessment methods, but appear to lack the opportunity, time and means to revise and update their assessment approaches. Therefore, sufficient and relevant assessment in-service training has to be planned and provided by the state.

In summary, the current study was fairly limited due to the small number of participants. Future research should attempt to: (a) train learners for longer periods to improve students' attitudes towards PA even more, (b) provide an in-depth analysis of learners' opinions, experiences and attitudes towards more theoretical issues relevant to PA, and (c) to investigate teachers' attitudes towards PA.

This study concludes that, with careful planning and training, PA is a viable alternative assessment in secondary education.

Conclusion

PA has great potential and is becoming a prominent tool in various subject areas at the secondary level including the field of EFL. It seems as a viable alternative to involve students in the assessment process and promote independence in secondary education. With careful preparation, monitoring and implementation, PA can "yield gains in cognitive, social, affective and transferable skill domains" that are at least as those from teacher assessment (Topping, 1998, p. 269). This study has indicated that providing learners with training in PA methods is a pre-requisite in order to ensure students' active participation and the successful implementation of PA in adolescent EFL writing classes. Finally, Ministries of Education should embrace this practice, include it in their curricula and provide adolescent EFL learners with more opportunities to develop their EFL skills after ensuring students' familiarization with this promising alternative method of learning and assessment.

Author's email: eleni.meletiadou@gmail.com

References

Ballantyne, R., Hughes, K. & Mylonas, A. (2002). 'Developing procedures for implementing peer assessment in large classes using an action research process.' *Assessment & Evaluation in Higher Education*, 27/5: 427-441.

Bassano, S. (1986). 'Helping children to adapt to unfamiliar methods.' *ELT Journal*, 40/1: 13-19.

Berg, 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. & William, D. (1998). 'Assessment and classroom learning.' *Assessment in Education: Principles, Policy and Practice*, 5/1: 7-74.

Boud, D. (2000). 'Sustainable assessment: rethinking assessment for the learning society.' *Studies in Continuing Education*, 22/1: 151-167.

Boud, D. & Holmes, H. (1995). 'Peer and self-marking in a large technical subject.' In D. Boud (Ed.), *Enhancing learning through self-assessment*. London: Kogan page, 63-78.

Brown, J. D. (1998). *New ways of classroom assessment*. Alexandria VA: Teachers of English to Speakers of Other Languages.

Brown, J. D. & Hudson, T. (1998). 'The alternatives in language assessment.' *TESOL Quarterly*, 32/4: 653-675.

Brown, S. & Dove, P. (1991). *Self and peer assessment*. Birmingham, England: Standing Conference on Educational Development.

Brown, S., Race, P. & Smith, B. (1997). *500 tips on assessment*. London: Kogan Page.

Chang, C-C., Tseng, K-H., Chou, P-N. & Chen, Y-H. (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,* 57/1: 1306-1316.

Cheng, W. & Warren, M. (1997). 'Having second thoughts: student perceptions before and after a peer assessment exercise.' *Studies in Higher Education*, 22: 233–239.

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

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.

Digiovanni, E. & Nagaswami, G. (2001). 'Online peer review: an alternative to face-to-face.' *English Language Teaching Journal*, 55/3: 263-271.

Donaldson, A. J. M. & Topping, K. J. (1996) 'Promoting peer assisted learning among students in further & higher education.' *SEDA Paper 96*.

Ellington, H., Earl, S., & Cowan, J. (1997). 'Making effective use of peer and self-assessment.' Accessed at http://apu.gcal.ac.uk/ciced/Ch26.html#1 on accessed 8 Oct 2010.

Falchikov, N. (1986). 'Product comparisons and process benefits of collaborative peer group and self assessments.' *Assessment & Evaluation in Higher Education*, 11: 146-165.

Falchikov, N. (2004). Involving students in assessment. *Psychology Learning and Teaching*, 3, 102-108.

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

Johns, A. M. (1986). 'Coherence and academic writing: some definitions and suggestions for teaching'. *TESOL Quarterly*, 20: 247-265.

Jones, L. & Fletcher, C. (2002). 'Self-assessment in a selective situation: an evaluation of different measurement approaches.' *Journal of Occupational and Organizational Psychology*, 75, 145-161.

Kwan, K. & Leung, R. (1996). 'Tutor versus peer group assessment of student performance in a simulation training exercise.' *Assessment and Evaluation in Higher Education,* 21/3: 205–214.

Lynch, T. (1988). 'Peer evaluation in practice.' In A. Brookes, & P. Grundy (Eds.), *Individualization and anatomy in language learning*. London, UK: Modern English Publications, 119-125.

McDowell, L. (1995). 'The impact of innovative assessment on student learning.' *Innovations in Education and Training International*, 32: 302-313.

Meletiadou, E. (2011). 'Peer assessment of writing in secondary education: its impact on learners' performance and attitudes.' Unpublished M.A. thesis, University of Cyprus.

Mendoca, C. & Johnson, K. (1994). 'Peer review negotiations: revision activities in ESL writing instruction.' *TESOL Quarterly*, 28/4: 745-768.

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.

Orsmond, P., Merry, S. & Reiling, K. (1997). 'A study in self-assessment: tutor and students' perceptions of performance criteria.' *Assessment & Evaluation in Higher Education*, 22(4): 357-369.

Papinczak, T., Young, L. & Groves, M. (2007). 'Peer assessment in problem-based learning: a qualitative study'. *Advances in Health Sciences Education*, 12/2: 169-186.

Porto, M. (2001). 'Cooperative writing response groups and self-evaluation.' *English Language Teaching Journal*, 55/1: 38-46.

Race, P., Brown, S., & Smith, B. (2005). *500 tips on assessment*. London and New York: Routledge Falmer.

Rothschild, D. & Klingenberg, F. (1990). 'Self and peer evaluation of writing in the interactive ESL classroom: an exploratory study.' *TESL Canada Journal,* 8/1: 52-65.

Rushton, C., Ramsey, P. & Rada, R. (1993). 'Peer assessment in a collaborative hypermedia environment: a case study.' *Journal of Computer-Based Instruction*, 20/3: 75.

Sadler, P. M. & Good, E. (2006). 'The impact of self and peer-grading on student learning.' *Educational Assessment*, 11/1: 1-31.

Sivan, A. (2000). 'The implementation of peer assessment: an action research approach.' *Assessment in Education*, 7/2: 193-213.

Sluijsmans, D., Moerkerke, G., Merrienboer, J. J. G. V. & Dochy, F. J. R. C. (2001). 'Peer assessment in problem based learning.' *Studies in Educational Evaluation,* 27/2: 153-173.

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

Stewart, M. & Cheung, M. (1989). 'Introducing a process approach in the teaching of writing in Hong Kong.' *Institute of Language in Education Journal*, 6/1: 41-48.

Strijbos, J. W., Narciss. & Dunnebier, K. (2010). 'Peer feedback content and sender's competence level in academic writing revision tasks: are they critical for feedback perceptions and efficiency?' *Learning and Instruction*, 20/4: 291-303.

Ten Berge, J. M. F. & Hofstee, W. K. B. (2004). 'Personality in proportion: a bipolar proportional scale for personality assessments and its consequences for trait structure.' *Journal of Personality Assessment*, 83/2: 120-140.

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

Topping, K. J. (2000). *Peer assisted learning: a practical guide for teachers*. Cambridge, MA: Brookline Books.

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

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

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

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

Wen, M. L. & Tsai, C. C. (2006). 'University students' perceptions of and attitudes toward (online) peer assessment.' *Higher Education*, 51: 27-44.

West, R. & Tsagari, D. (2004*). Testing and assessment in language learning (Vol. 3).* Patras, Greece: Hellenic Open University.

Employing Computer Assisted Assessment (CAA) to facilitate formative assessment in the State Secondary School: a case study

[Χρησιμοποιώντας την υποβοηθούμενη από ηλεκτρονικούς υπολογιστές αξιολόγηση για τη διευκόλυνση της διαμορφωτικής αξιολόγησης στην Δημόσια Δευτεροβάθμια Εκπαίδευση: Μία μελέτη περίπτωσης]

Effimia Karagianni

Based on theories of assessment as well as on the pedagogical and administrative advantages Computer Assisted Assessment (CAA) has to offer in foreign language learning, the study presented in this paper examines how computers can facilitate the formative assessment of EFL learners and enhance their feeling of responsibility towards monitoring their progress. The subjects of the study were twenty five 14-year-old students attending the third class of a State Gymnasium in Greece. The instruments utilized were questionnaires on motivation and learning styles, three quizzes designed with the software Hot Potatoes, a self–assessment questionnaire and an evaluation questionnaire showing the subjects' attitudes towards the experience of using computers for assessing purposes. After reviewing formative assessment, CAA and how these two can be combined, the paper focuses on the description of the three class quizzes used in the study. Information from the questionnaires filled in by students combined with the results of the quizzes, shows how computers can be used to provide continuous ongoing measurement of students' progress needed for formative assessment. The results are also used to show how students and teachers can benefit from formative CAA and the extent to which such kind of assessment could be applicable in the Greek state school reality.

೫

Η μελέτη που παρουσιάζεται σε αυτό το άρθρο, επιχειρεί να εξετάσει με ποιο τρόπο οι υπολογιστές μπορούν να διευκολύνουν τη διαμορφωτική αξιολόγηση των μαθητών των Αγγλικών, καθώς και την ανάπτυξη του αισθήματος υπευθυνότητας τους προς την παρατήρηση της προόδου τους. Η μελέτη βασίζεται σε θεωρίες σχετικά με την αξιολόγηση καθώς και στα παιδαγωγικά και χρηστικά πλεονεκτήματα που η υποβοηθούμενη από υπολογιστές αξιολόγηση μπορεί να έχει για την εκμάθηση μιας ξένης γλώσσας. Τα άτομα που συμμετείχαν στην έρευνα ήταν είκοσι πέντε δεκατετράχρονοι μαθητές που φοιτούν

στην τρίτη τάξη ενός δημόσιου Γυμνασίου στην Ελλάδα. Τα εργαλεία που χρησιμοποιήθηκαν για τη μελέτη ήταν ερωτηματολόγια σχετικά με τα κίνητρα και τον μαθησιακό τύπο των μαθητών, τρία quiz σχεδιασμένα με το πρόγραμμα Hot Potatoes, ένα ερωτηματολόγιο αυτοαξιολόγησης καθώς και ένα ερωτηματολόγιο που δείχνει τη συμπεριφορά των μαθητών προς την αξιολόγηση αυτού του είδους. Το άρθρο μετά από αναφορά στη συνεχή αξιολόγηση, στην υποβοηθούμενη από υπολογιστές αξιολόγηση καθώς και στο πως αυτά τα δύο μπορούν να συνδυαστούν, επικεντρώνεται στην περιγραφή των τριών quiz που χρησιμοποιήθηκαν στην έρευνα. Οι πληροφορίες από τα ερωτηματολόγια σε συνδυασμό με τις επιδόσεις των μαθητών στα quiz υποδεικνύουν τον τρόπο με τον οποίο οι υπολογιστές μπορούν να χρησιμοποιηθούν για να καταγράψουν την συνεχή μέτρηση της προόδου των μαθητών που απαιτείται για τη διαμορφωτική αξιολόγηση. Επίσης τα αποτελέσματα υποδεικνύουν το αν οι μαθητές και οι καθηγητές μπορούν να επωφεληθούν καθώς και το κατά πόσο ένα τέτοιο είδος αξιολόγησης θα μπορούσε να εφαρμοστεί στην Ελληνική σχολική πραγματικότητα.

Key Words: Computer Assisted Assessment (CAA), English as a Foreign Language (EFL), formative assessment, Information Technology (IT)

Introduction

There has been an increasing interest in the interaction between learning and assessment in recent years. This is partly due to the change of focus from teacher-centered classroom environments to learner-centered ones as well as due to the dissatisfaction from final examinations which in some cases may not be adequately linked to the students' and learning's goals (Maughan, Peet & Willmott, 2001). It is widely agreed that the desirable interconnection between learning and assessment can be achieved through formative assessment which provides information to be exploited by both teachers and learners for the improvement of teaching and learning situations through observation and the provision of feedback (Black & William 1998b, Brown & Knight, 1994). As such, assessment is integrated with instruction and facilitates learning by encouraging learners "to 'think evaluatively' and "self-assess" (Lynch, 2003, p. 9) leading thus to important learning gains.

Since the 1960's, when The PLATO project at the University of Illinois pioneered the use of computers for language practice and assessment (Godwin-Jones, 2001), and especially after the arrival of the World Wide Web in 1993, computers are being used more and more to facilitate assessment processes. The term CAA (Computer-Assisted Assessment) is now used to refer to the use of computers to deliver, mark and analyze students' assignments and examinations, to keep a record of assessment and to provide feedback (McKenna & Bull, 2000). CAA can facilitate the way assessment is carried out and presents various pedagogical and administrative advantages.

However, despite being a promising field of Information Technology (IT) with applications in Computer Assisted Language Learning (CALL), little implementation in CAA has taken place. So far instances of CAA implementation have been restricted in institutions of Higher Education.

With these in mind, the main aim of the present paper is to make use of CAA in order to facilitate formative assessment in the State Secondary Schools. For this reason all the instruments utilized in the study – questionnaires and class quizzes- are delivered via

computers. Data analysis aims to establish whether computers can be used effectively to facilitate formative assessment and whether formative CAA can be beneficial to students and teachers and applied in State Schools.

Literature Review

According to the Greek Cross-Thematic Curriculum for compulsory Education (Pedagogical Institute, 2001, p. 29) "the main purpose of student assessment is to provide feedback about both pupil progress and teacher success and also to identify learner strengths and weaknesses". More specifically, assessment according to the Curriculum aims to find out whether and to what degree learning and teaching aims have been achieved as well as to provide feedback on learner performance either in individual or class level. It furthermore aims to identify learning difficulties and plan future teaching accordingly.

However, the purpose of assessment may vary greatly given the various contexts in which it is applied. Assessment can be used to measure proficiency, to diagnose needs or simply to determine achievement in relation to syllabus objectives (Lynch, 2003, p. 1). These purposes as well as the stage of the course of studies, at which assessment occurs, are usually the factors that determine the kind of assessment to be used as seen in Figure 1.
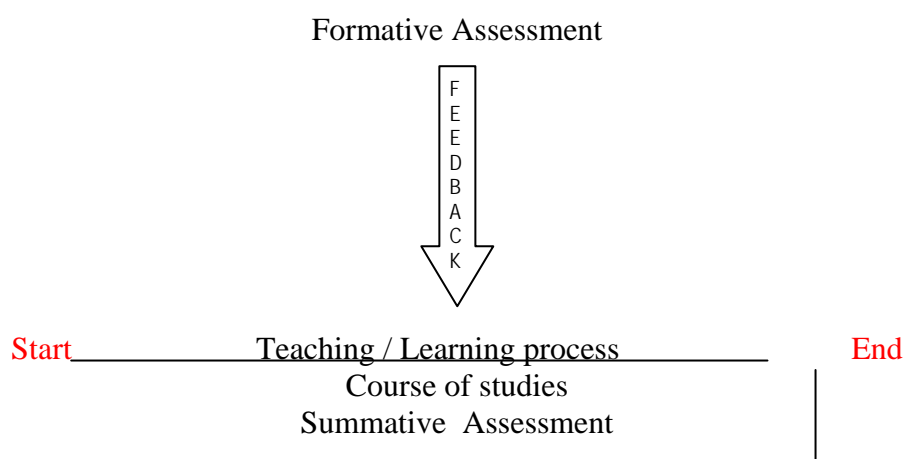
Formative Assessment

Start—————————Teaching / Learning process————— End

Course of studies

Summative Assessment

Figure 1. Linear representation of the two forms of assessment and their relationship to the teaching and learning process

Formative assessment is realized as a kind of assessment which is ongoing, can happen anywhere during the teaching and learning process (Tunstall and Gipps, 1996) and is used primarily to assist rather than judge the learner (Lynch, 2003,West, 2004). Formative assessment has much to offer to the students as such by placing them in the centre of the assessment process. Furthermore, formative assessment can help students become skillful judges of their own strengths and weaknesses and set realistic goals becoming thus self-directed and autonomous learners (Brindley, 2001) and develop lifelong learning skills (Lemos, 1999).

On the other hand, summative assessment is of a larger scale and happens usually at the end of a course of studies to judge the learner and the results of teaching. To sum up as Carden (2000, p. 2) puts it, formative assessment is assessment *for* learning whereas summative assessment refers to assessment *of* learning.

However, for formative assessment to be properly used certain criteria need to be met. First of all it needs to provide positive and immediate feedback (Freeman and Lewis, 1998), and

actively involve students and teachers in the assessment procedure (Brown & Knight, 1994; Maughan et al, 2001). What is more, formative assessment needs to enable students to assess themselves and their peers and understand how to improve (Grabe and Kaplan, 1996). These attributes, alongside standard test qualities such as validity, reliability, practicality, utility and washback formed a set of criteria which were kept in mind during the design of the materials employed in the present study.

Computers and EFL

As mentioned earlier, computers have been used in language assessment as early as the 1960's. However, this kind of assessment did not become popular until the late seventies and early eighties when personal computers appeared (Godwin-Jones, 2001). What is more, the introduction of the Internet in 1993 provided language teachers with unlimited options and possibilities for either language learning or language testing.

Computer Assisted Assessment (CAA) refers to the use of computers to deliver, mark and analyze student assignments or examinations (Bull & McKenna, 2000; McKenna, 2001; Seale, n.d.) and can cater for various types of assessment such as diagnostic, formative or summative assessment. So far CAA is usually associated with testing skills such as reading and listening, whereas limited attempts have been made towards testing productive skills such as writing and speaking. However, attempts are made towards improving possibilities. Continuous research in the area of CAA attempts to create systems that will measure language proficiency as accurately as traditional means of foreign language assessment (Giouroglou and Economides, 2004).

CAA offers both pedagogical and administrative advantages. Among others, CAA presents students with the opportunity to adjust assessment and learning to their own needs (McKenna, 2001) promoting thus learner autonomy (Giouroglou and Economides, 2004, p.748). Furthermore, it allows for quick and objective marking (Oliver, 2000).

The Formative CAA Chain

The Formative CAA Chain (see Figure 2) is a representation of the way computers can facilitate formative assessment. It serves as a melting pot where key notions of the two areas are combined so that CAA can facilitate formative assessment procedures. At the same time this procedure functioned as a rationale which was kept in mind during the design of the materials.

Computers can help students become aware of their motives and learning habits in an attempt to involve them more actively into the learning process. Through appropriate software the teacher designs the assessment tasks. Student feedback needs to be immediate and relevant if it is to help the learning process (Freeman and Lewis, 1998). At the same time information has to be provided to the teacher so as to keep records of students' progress, of task scores and of time taken to complete each task. This information will enable the teacher to decide if any remedial work is needed or not.

All these principles are presented in Figure 2 below which presents all the actions needed to be taken either by students or teachers during a typical CAA task. It also shows the order in which actions need to occur.
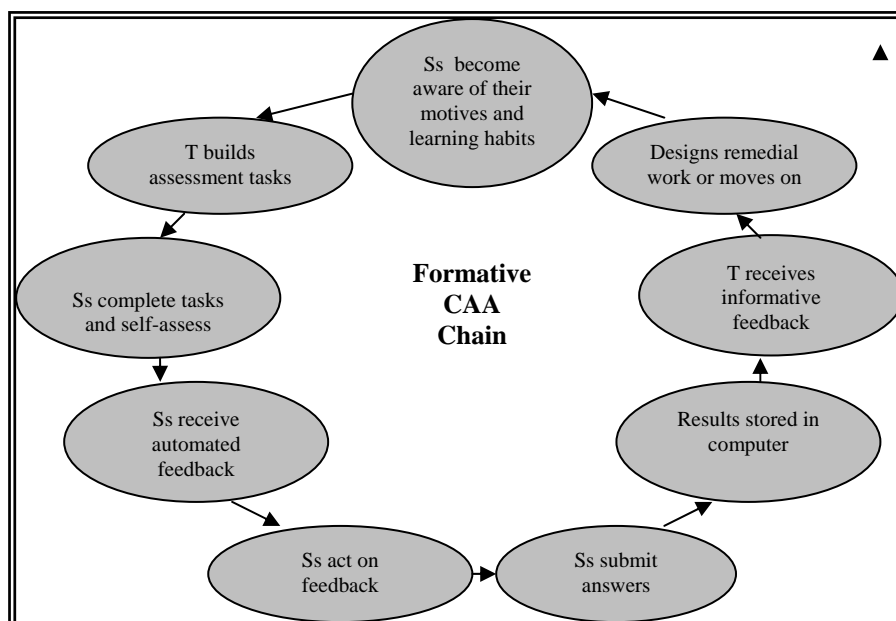
Figure 2. The Formative CAA Chain
(Ss=Students, T=teacher)

Research context

The class participating in the study was the third class of the 4th State Gymnasium of Arta, a town in the northwest of Greece. It consisted of twenty- five students-10 boys and 15 girls-aged 14-15 years old. According to the classification provided by the Common European Framework of Reference for Languages (CEFR; Council of Europe 2001) students of this age belong to B1 + Level (Threshold level). However, the majority of the students attended classes at frontisteria i.e private language institutions. As an effect, their level is increased. In reality, most of the students in this class fell somewhere between the Threshold (basic language users) and the Vantage level (Independent language users) with a small number of students belonging at the Waystage level comprising thus a mixed- ability class.

According to the Formative CAA Chain, students have to become aware of their motives and ways of learning as a first step towards learning autonomy. To achieve this, students had to complete two online questionnaires concerning their motivation (Appendix I) i.e., the reason for learning English, and their learning styles (Appendix II). Based on students' answers (see Karagianni, 2007) most of them were instrumentally motivated (Gardner and Lambert, 1972, Stern, 1983, in Papaefthymiou-Lytra, 1993, p. 92), i.e. they learned English for quite practical reasons such as to obtain an English certificate.

The questionnaires on learning styles – adapted from Victori (1992) – aimed at helping learners understand how they learn in order to involve them more actively in the learning and assessment process and pave their way towards autonomous and personalized learning. According to the results, most of the students (11) presented a combination of all three styles, whereas fewer students were categorized as distinctly visual (6), auditory (5) or haptic (3) learners.

In this context, various assessment techniques are used. Formative assessment is realized in the form of homework and oral assessment in class whereas the progress tests written twice or three times a year are used to judge rather than assist the learner. Each June students take an achievement test designed and administered by the class teacher which tests their

understanding and use of the structures, functions and lexis taught in the syllabus. Such techniques though, need reconsideration as most of the times do not seem to provide the teacher and the students with all the information needed for a successful teaching and learning process. This dissatisfaction from the assessment techniques along with the lack of technology in the teaching of English in the State School, led to the idea of employing CAA in order to facilitate formative assessment in this context.

As such, the decision to carry out formative assessment tasks via computers was based on a desire to make assessment more attractive and more informative not only to the students but also to the teachers who need to take into account their students' special strengths and weaknesses. It also reflects an attempt to incorporate new technologies in the classroom.

Methodology

The authoring tool used for creating the computerized format of the quizzes was the software *Hot Potatoes*, a user-friendly program which allows teachers to create a variety of tasks quickly and easily. Such software "gives the teacher the potential to localize activities, using learners names, texts they have produced, authentic material from newspapers, brochures or leaflets" (Slaouti, 2005, p. 184). Hardisty & Windeatt (1989, p. 8, in Slaouti, 2005, p. 185) claim that: "the main effect […] on methodology is that students can work through some exercises and have them marked automatically by the computer'. Furthermore, such programs have the potential to promote independent learning within the classroom and allow both the teacher and the student to have some control over the classroom curriculum (Slaouti, 2005, p. 185). However, the implementation of such changes is not always easy. Various parameters needed to be taken into account such as availability of laboratory and time, reliability of the computer network as well as students' positive reaction towards computer use.

The three quizzes designed aimed to assess students' mastery of ideas, lexis and structures taught in Units 16-19 of the book used in class. To achieve this aim the quizzes were designed to cover areas such as reading, grammar and vocabulary. However, skills such as speaking and writing were not covered, as Hot Potatoes would not cater for the subjective marking needed for the productive skills. The core aim of the three class quizzes was to involve students into the assessment process through carefully chosen task types, provision of necessary feedback and self assessment. The term *quiz*, i.e. small-scale quick informal tests focusing on one or two language points, is used here in order to differentiate these tasks from large-scale formal tests which are more summative in nature and range over a large number of language items covering a longer period of study. These quizzes were meant to be used not only as a means of detecting the progress of the target group but also as a means of diagnosing strengths and weaknesses of the students.

All three quizzes included reading, grammar and vocabulary components. In the first two quizzes all parts were shorter whereas the third quiz consisted of a larger text and longer grammatical and vocabulary exercises. For each task item the students were provided with feedback guiding them to the correct answer if needed. Objective scoring techniques were employed in all three quizzes. As the focus was on specific areas of language and the questions need to generate specific answers from the students, the item types used were multiple choice questions (MCQ) and cloze as well as multiple matching and text ordering.

More specifically, the reading section of the quizzes aimed at peering into the student's ability to understand the coherence of a text as well as assessing if students applied skimming and scanning strategies correctly (see Figure 3).
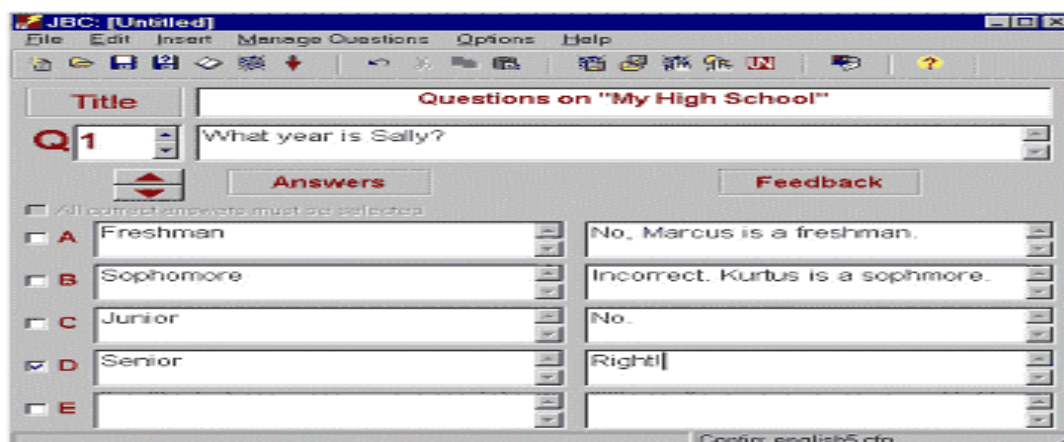
Figure 3. Reading quiz

The texts presented were authentic and came from literature books and magazines. Students had to apply both skimming and scanning strategies as both fragmentary and detailed understanding was required for the successful completion of the tasks.

As the students were still acquiring the linguistic code of the English language, both grammar and vocabulary components were included in the quizzes (see Figure 4). These tasks aimed to explore whether the students had acquired newly taught lexis and structures (see Karayianni, 2007).

Figure 4. Vocabulary quiz

After the completion of the quizzes students were asked to fill in an online self-assessment questionnaire (Appendix III) in order to conclude the assessment process and provide the students with a more direct self-assessment experience.

Considerations on design and effectiveness
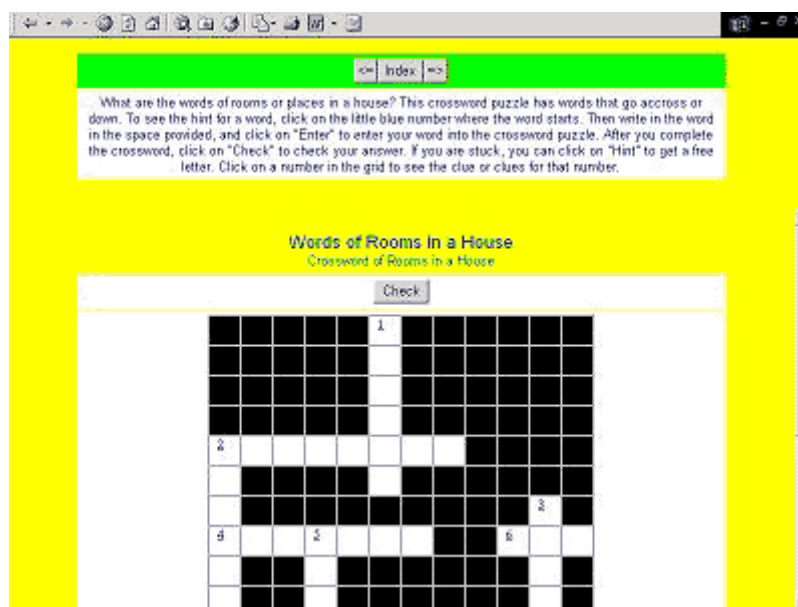
The suitability of the tasks was judged with reference to the basic principles of formative assessment and the Formative CAA Chain presented earlier. To begin with the feedback provided to the students during the tasks was quick and relevant to the task and seemed to be of great support, as it guided the students successfully to the completion of the tasks. The students were actively involved into the assessment process by acting on the feedback. What is more, they were really keen on commenting on the nature of the tasks and self-assessing.

Results were stored successfully in hotpotatoes.net and then downloaded at each student's computer for easier access. Each student then received a printed page informing them about their score on each task. However, detailed information about exact mistakes was not provided by the programme, an aspect which may prove problematic in some cases where more details are needed for each student.

Results

The computer allocated various marks for each task depending on the effort invested in each task. The more effort put in a task, the lower the mark awarded for the correct answer. Students were given marks for each task separately, with the top mark being 100. Then the score of each student was calculated by the teacher so as to give a total mark for each quiz. For ease of reference, the scores were divided into three broad categories. Based on the rating scale of one hundred, scores over 80 indicated quite good performance, scores between 60 and 80 showed good performance but in need of improvement, whereas scores below 60 showed weak performance.

After the students took each quiz, the results pinpointed areas of weakness and whether any remedial work was needed. Thus Class Quiz III covering Units 16-19 took place after completion of remedial work and served as an indicator of whether the students had improved after the two first quizzes or not.

A careful look at the comparative results of the three class quizzes showed that such a kind of assessment managed to actively involve the students in the assessment process and made them improve as far as grammar is concerned. As shown in Graph 1, in the first and second quiz high scores were received by 8 (32%) and 9 (36%) students respectively whereas in the last grammar quiz 10 (40%) students scored over 80.
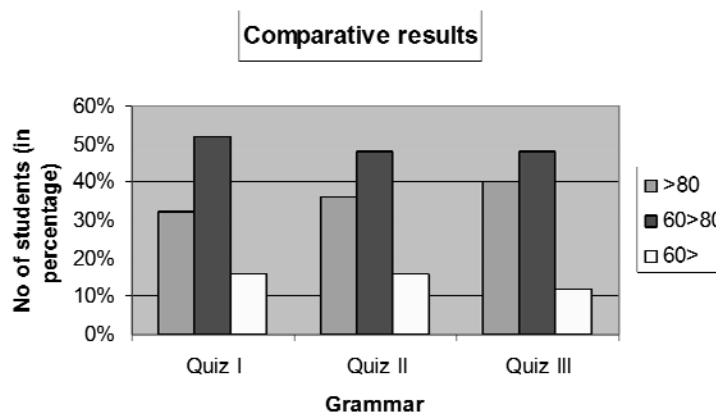
Graph 1. Comparative results in grammar

The students improved in reading too. 7 students (28%) scored less than 60 in the third quiz in comparison with Quizzes I & II where students with less than 60 were 12 (48%) and 10 (40%) respectively(see Graph 2).
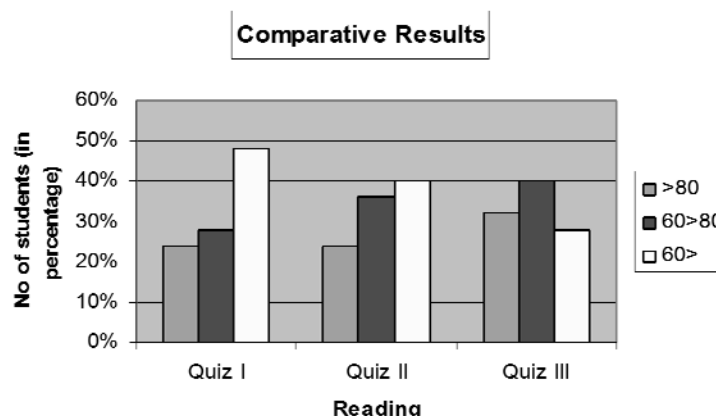
Graph 2. Comparative results in reading

In terms of time taken to complete each quiz, it seems that the students devoted more time in completing the third quiz (Table 1). It should be stressed though that the tasks included in the third quiz were more demanding and as such needed more time to be completed.

	Time in minutes
Quiz I	15 -16min
Quiz II	20min
Quiz III	30-35min

Table 1. Comparative table showing completion time for the three quizzes

In general, students' scores seemed to improve over the three quizzes, given the practice in class and the revision exercises done after each quiz. However, it is possible that the scores could have improved due to familiarisation of students with the format of the quizzes. Thus, students need to be exposed more to similar types of assessment before its exact impact on student performance can be specified.

Students' perceptions

The online evaluation questionnaires (Appendix IV) completed by students after the quizzes, showed their varying attitudes towards the quizzes and identified benefits as well as problems. Almost all the students (24) seemed to actually enjoy the whole process and found the assessment entertaining and less stressful than other kinds of assessment. Opinions on the difficulty of the tasks varied. 11 students found them quite easy, whereas others pinpointed areas of difficulty especially in the reading (9 students) and the grammar components (13 students) of the third quiz.

When the students were asked about their attitude towards computers 16 of them admitted that they were anxious about being assessed through computers; the rest though – especially computer literate students – highly valued the fact that they could do the tasks even from home and as many times they wanted. In fact, some of them asked for more tasks for the rest of the units of their book to be uploaded to the Internet. In general, despite some negative comments the students appreciated this kind of assessment.

Conclusion

This paper set out to describe and analyze a way of facilitating formative assessment in Greek State Secondary Schools by implementing CAA. A study was conducted towards this end, based on a theoretical framework established after examining the nature of formative assessment and CAA. The study, which focused on a specific group of students, attempted to identify whether the particular students and the teaching and learning process itself would benefit from the implementation of CAA for formative assessment purposes. In addition, it aimed at examining whether such an approach to assessment would be applicable to the specific context. The main question of this study was whether assessment via computers could be effective as a whole to students and class: if it contributed to the learning process itself and if it managed to comply with the theoretical framework underlying the implementation of computers for formative assessment purposes.

Each stage of the research study undertaken, from identifying the students' motives and learning styles to storage of assessment results, was carried out solely by computers. The procedure followed was shown in the Formative CAA Chain. According to this procedure, online questionnaires were used so as to trace the motivation and learning styles of the specific group of students. This was considered necessary as it was a good starting point for involving students into the learning and assessment process. Then students were asked to complete online quizzes which would serve as a means of detecting their progress as well as a tool to help them and the teacher identify possible weaknesses. Online self –assessment questionnaires were also put forth in order to involve the students more into the assessment process. Lastly, the students were asked to evaluate this experience by completing an online evaluation questionnaire.

The analysis suggests that, on the whole, computers were successfully used for formative assessment purposes, as they managed to provide the necessary feedback when needed and succeeded in involving the students in the assessment tasks in a meaningful and productive way.

From a pedagogical viewpoint, the study showed that computers can turn assessment into a worthy experience for students by motivating them and engaging them actively in the assessment process, thus changing them from passive recipients to active participants. Furthermore, computers can greatly facilitate the provision of feedback and detection of problematic areas. What is more, involving computers into the assessment process provides students with more practice in the area of IT, helping them to acquire skills needed for their future life. With regard to administration, computers can save a lot of time when it comes to marking assessment tasks and storing as well as retrieving of marks.

Despite some negative comments, students seemed to have gained valuable experience not only participating in formative assessment but also in the area of technology.

According to this study, there is strong evidence that CAA can influence assessment in a positive way. Thus educational institutions should consider the pedagogical and administrative advantages which can be readily obtained by its implementation. However, the exact impact on learners' future performance cannot be specified unless longer exposure to and further research on such a kind of assessment takes place.

Author's email: effieka@hotmail.gr

References

Black, P. J., & Wiliam, D. (1998b). 'Inside the black box : raising standards through classroom assessment.' *Phi Delta Kappan,* 80/2: 139-149.

Brindley, G. (2001) 'Outcomes-based assessment in practice: some examples and emerging insights.' *Language Testing*, 18/4: 393-407.

Brown, S & Knight, P. (1994). *Assessing learners in higher education.* London. Kogan Page.

Bull, J. & McKenna, C. (2000). *Proceedings of the 4th international CAA conference*, Loughborough: Loughborough University.

Council of Europe (2001). *Common European framework of references for languages: learning, teaching, assessment.* Cambridge: Cambridge University Press.

Carden, T. (2000). 'The Concepts of Assessment', accessed at http://www.tcarden.com/coursework/testassess.htm on 22 Sep 2006.

Freeman, R. & Lewis, R. (1998). *Planning and implementing assessment*. London: Kogan Page.

Gardner, R.D. & Lambert, W.E. (1972). *Attitudes and motivation in second language learning*. Rowley, Mass: Newbury House.

Giouroglou, H. & Economides, A. (2004). 'State-of-the-art and adaptive open-closed items in adaptive foreign language assessment.' *Proceedings of the 4th Hellenic Conference with International Participation: Informational and Communication Technologies in Education,* Athens, pp.747-756. Accessed at http://www.etpe.gr on 5 Dec 2006.

Godwin-Jones, B. (2001). 'Emerging Technologies: Language Testing Tools and Technologies.' *Language Learning & Technology,* 5/2: 8-12, accesed at http://llt.msu.edu/vol5num2/emerging on17 Mar 2007.

Grabe, W. & Kaplan, R. (1996). *Theory and practice of writing*. London: Longman.

Green, J. M. (1998). 'Constructing the way forward for all students.' A speech delivered at "Innovations for Effective Schools" OECD/New Zealand joint follow-up conference, Christchurch, New Zealand.

Hardisty,D. & Windeatt, S. (1989). *CALL.* Oxford: OUP.

Karagianni, E. (2007). 'Employing Computer Assisted Assessment (CAA) to facilitate formative assessment in the State Secondary School: A Case Study.' Unpublished M.Ed. dissertation. Patras: Hellenic Open University.

Lemos, M.S. (1999). 'Students' goals and self-regulation in the classroom.' *International Journal of Educational Research*, 31/6: 471-485.

Lynch, B. (2003). *Language assessment and programme evaluation.* Edinburgh: Edinburgh University Press.

Maughan, S., Peet, D. & Willmott, A. (2001). 'On-line Formative Assessment, Item Banking and Learning Support', accessed at http://www.caaconference.com/pastConferences/2001/proceeding/01.pdf on 18 Dec 2006.

McKenna, C. (2001). 'Introducing computers into the assessment process: what is the impact upon academic practice?' Paper presented at the *Higher Education Close-Up Conference 2, Lancaster University,* 16-18 July 2001*,* accessed at http://www.leeds.ac.uk/educol/documents/00001805.htm on 4 May 2007.

Mc Kenna, C. & Bull, J. (2000). 'Quality Assurance of computer-assisted assessment: practical and strategic issues.' *Quality Assurance in Education,* 8/1: 24-31.

Oliver, A. (2000). 'Computer-assisted assessment'. *LTDU Magazine*, 1, accessed at http://www.herts.ac.uk/lis/ltdu/journal/april2000/caa.html on 10 Dec 2006.

Papaefthymiou-Lytra, S. (1993). *Language, language awareness and foreign language learning.* Athens: University of Athens Press.

Pedagogical Institute. (2001**).** *Diathematikon programma: cross-thematic curriculum framework for compulsory education*. Athens: Pedagogical Institute.

Seale, J. (n.d.). 'Using CAA to support student learning', accessed at http://www.alt.ac.uk/docs/eln004.pdf on 20 Sep 2006.

Slaouti, D. (2005). 'Computers and writing: the role of the word processor, e-mail and the WWW.' In D. Slaouti & P. Bouboureka, *Educational Technology in ELT (Vol. II)*. Patras: Hellenic Open University, 77-134.

Slaouti, D. (2005). 'Exploring text through technology'. In D. Slaouti & P. Bouboureka, *Educational Technology in ELT (Vol. II)*. Patras: The Hellenic Open University, 165-216.

Stern, H.H. (1983). *Fundamental concepts of language teaching.* London: Oxford University Press.

Tunstall, P., & Gipps, C. (1996) 'Teacher feedback to young children in formative assessment: A typology.' *British Educational Research Journal*, 22/4: 389-404.

Victori, M. (1992). 'Investigating the metacognitive knowledge of students of English as a second language.', Unpublished thesis. Los Angeles: University of California.

West, R. (2004) 'Introduction and principles of language testing.' In R. West, *Testing and assessment in language learning (Vol. 1).* Patras: Hellenic Open University, 21-80.

West, R. (2004) 'Diagnostic and progress testing.' In C. Tsagari & R. West, *Testing and assessment in language learning (Vol. 3)*. Patras: Hellenic Open University, 13-73.

Appendix I

Questionnaire on motivation

Dear students,
The following questionnaire refers to your motivation on learning English. The answers will also help me and you understand better why you learn English. Thanking you in advance. Your English teacher.

Read each statement below and choose the answer that indicates your opinion to the statement.					
	Strongly Disagree	Disagree	Uncertain	Agree	Strongly agree
I began to study English because my parents required me to learn it. *	☐	☐	☐	☐	☐
The main aim of my English learning is to obtain a certificate in the English language. *	☐	☐	☐	☐	☐
Learning English is important for me, because English is very useful in our society. *	☐	☐	☐	☐	☐
Only with good English can I find a good job in the future. *	☐	☐	☐	☐	☐
I learn English because I am interested in English speaking people and their cultures. *	☐	☐	☐	☐	☐
I like language learning in general. *	☐	☐	☐	☐	☐
I want to learn English because of my love for English songs/movies. *	☐	☐	☐	☐	☐
I want to learn English because I like this language. *	☐	☐	☐	☐	☐
I learn English because I want study or work abroad. *	☐	☐	☐	☐	☐
Knowledge of good English shows good education. *	☐	☐	☐	☐	☐
Learning English is important for one's success in life. *	☐	☐	☐	☐	☐
Thank you for filling in this questionnaire. I hope you enjoyed it.					

Appendix II

Questionnaire on Learning Styles

Dear students,
The following questionnaire is about your learning preferences when learning English. The answers will help me and you understand better the way you learn English. Thanking you in advance.
Your English teacher.

Tick the sentences which apply to your situation.

□ 1. I can remember something better if I write it down.

□ 2. When reading something I read it aloud.

□ 3. I need to discuss things to understand them better.

□ 4. I prefer doing sth directly rather than listening to directions.

□ 5. I can study better with the radio on.

□ 6. I need frequent breaks when studying.

□ 7. I don't like studying at a desk.

□ 8. When I read or listen to the teacher I take lots of notes.

□ 9. I prefer hearing a lecture rather than reading a textbook.

□ 10. When I don't remember a word I use another word.

□ 11. I can easily follow a speaker even though I am not looking at him/her.

□ 12. I study better when it is quiet.

□ 13. It's easy for me to understand maps.

□ 14. I remember what people look like.

□ 15. I remember what people say.

□ 16. I take notes but I don't look at them again.

□ 17. It's hard for me to picture things in my head.

□ 18. I Talk to myself when I do my homework.

□ 19. When taking a test, I can "see" the textbook page and the correct answer in it.

□ 20. I like to complete one task before starting another.

□ 21. I move my lips when I read.

□ 22. When I am trying to remember something new it helps me to picture it in my head.

□ 23. When I learn something new I prefer to listen to information on it, then read about it, then do it.

Thank you for filling in this questionnaire. I hope you enjoyed it.

Appendix III

Self- assessment questionnaire

Self-assessment sheet
Read the following statements and choose the one that is true for you.

1.	I can find general information on a text. *	☐	Not very well	☐	Normally	☐	Well	☐	Very well
2.	I can find specific information on a text. *	☐	Not very well	☐	Normally	☐	Well	☐	Very Well
3.	I can put a text in sequence *	☐	Not very well	☐	Normally	☐	Well	☐	Very Well
4.	I can use conditionals,modals and question tags *	☐	Not very well	☐	Normally	☐	Well	☐	Very Well
5.	I can discriminate between Past Simple, Past Continuous and Used to. *	☐	Not very well	☐	Normally	☐	Well	☐	Very Well
6.	I can understand and use target vocabulary. *	☐	Not very well	☐	Normally	☐	Well	☐	Very Well

Appendix IV

Evaluation Questionnaire

Evaluation Questionnaire

Dear students,
By filling this questionnaire you will help me reflect on the practicality of the kind of assessment you went through this year. Thank you!
Your English teacher.

1. Which part of the process did you like best? Why?

2. Did you feel more responsibility for your work?

3. How is this kind of assessment better than the usual assessment?

4. What was difficult for you?

5. Would you like to change some items? Which? Why?

6.Would you like to take such a kind of assessment again? Why / Why not?

7.   How did you feel about being assessed by computers?

8.   Add your suggestions and ideas to help me in the future.

Thank you for filling in this questionnaire.

Note: All the above forms were created at http://www.formdesk.com/

Developing and assessing EFL students' writing skills via a class-blog

[Αναπτύσσοντας και αξιολογώντας τις γραπτές δεξιότητες των μαθητών στην αγγλική γλώσσα μέσα από ένα ιστολόγιο τάξης]

Eleni Daskalogiannaki

This paper presents the implementation and the positive findings of a study that merges blog use and portfolio development for teaching and assessing writing. More specifically, it investigates whether a class blog can be integrated into the Greek EFL teaching context as an effective means to engage learners in process writing and as a form of e-portfolio, where they can keep track of their writing development. It also examines blog use for enhancing students' motivation, interaction, participation and learning. The study followed a project-based approach and was conducted in a state Junior High School in Greece. Data was collected over a 4-month period via a questionnaire as well as from analyzing students' writing samples and teacher's observations of whole-class behavior during blogging. The findings reveal that the blog encouraged students to approach writing as a cognitive process of constant modification, motivated them to write more and better in various writing genres, and helped them become competent, autonomous and critical writers.

ଔ

Η μελέτη παρουσιάζει την εφαρμογή και τα θετικά αποτελέσματα μιας έρευνας η οποία συγχωνεύει τη χρήση ιστολογίου και την ανάπτυξη ατομικού φακέλου αξιολόγησης για τη διδασκαλία και την αξιολόγηση του γραπτού λόγου. Πιο συγκεκριμένα, ερευνά κατά πόσο ένα ιστολόγιο τάξης μπορεί να ενσωματωθεί στο ελληνικό πλαίσιο διδασκαλίας της αγγλικής ως ξένης γλώσσας σαν ένα αποτελεσματικό μέσο για να εμπλακούν οι μαθητές στα διάφορα στάδια της διαδικασίας συγγραφής κειμένου και σαν μια μορφή ηλεκτρονικού φακέλου, όπου θα μπορούν να παρακολουθούν την πρόοδο τους στην παραγωγή γραπτού λόγου. Εξετάζει επίσης κάτα πόσο η χρήση ιστολογίου μπορεί να ενισχύσει το κίνητρο των μαθητών, την αλληλεπίδραση, τη συμμετοχή και τη μάθησή τους. Η μελέτη ακολούθησε τη μέθοδο project και διεξήχθη σε ένα δημόσιο Γύμνασιο σχολείο στην Ελλάδα. Τα στοιχεία συγκεντρώθηκαν σε διάστημα τεσσάρων μηνών μέσα από ένα ερωτηματολόγιο και από την ανάλυση αντιπροσωπευτικών γραπτών κειμένων των μαθητών, καθώς και παρατηρήσεων της καθηγήτριας αναφορικά με τη συμπεριφορά των μαθητών καθόλη τη διάρκεια της συγγραφής στο ιστολόγιο. Τα ευρήματα αποκαλύπτουν ότι το ιστολόγιο ενθάρρυνε τους μαθητές να προσεγγίσουν το γράψιμο ως

μια γνωστική διαδικασία συνεχούς τροποποίησης, τους παρότρυνε να γράψουν περισσότερο και καλύτερα σε διάφορα είδη και τύπους γραπτών κειμένων, και τους βοήθησε να γίνουν ικανοί, αυτόνομοι και στοχαστικοί συγγραφείς.

Introduction

Traditional approaches to writing see it as an extension of pre-taught grammar and vocabulary or as an additional opportunity to practice other language skills, focusing on the accuracy and correctness of the final product (Hyland, 2002). However, such approaches are currently challenged since they can neither develop nor measure the writing skills learners will actually need to communicate effectively in the target language (Song and Wang, 2005). In search for new ways to assist authentic learning performance in the Greek ELT writing class, the present study examines the potential of one of the most recent Web 2.0 tools, the blog, to promote the approach of writing as a process, intrigue the learners and form a web-portfolio where they can systematically reflect on their writing progress.

Literature review

The process approach to writing

In the 1970's dissatisfaction with traditional approaches to writing led to the development of the process approach. According to it, writing is acknowledged as a non-linear, problem-solving cognitive process whereby writers discover and reformulate ideas as they attempt to approximate meaning (Silva, 1990). It is an on-going process of modification in which audience expectations, writer's purpose and context of writing are simultaneously taken into consideration. Flower and Hayes (1980) divide the writing process into three recursive stages: 1) planning (generating ideas from the long-term memory, organizing these ideas and using them to set goals), 2) translating (putting the retrieved ideas on paper and producing multiple drafts of one's work), and 3) reviewing (reflecting on the text, and editing it to produce the final draft). In actual teaching, emphasis has been placed on setting realistic and purposeful tasks, stimulating thinking, helping students organize their ideas, and supporting revision. The process approach has also changed the way feedback is provided. The teacher is not the only person responsible for evaluating students' work but rather one of the people responding to student writing, the others being the students themselves and their peers.

Towards an alternative way to develop and assess process writing

With the advancement of technology, blogging has emerged offering a rich, interactive, user–friendly platform that facilitates the sharing of ideas among users on the web. A blog is a type of diary that is kept online (Allford and Pachler, 2007). It consists of short postings about the reflections of the blogger as well as comments. Although blog use for pedagogical purposes appears to be in its infancy, many linguists have stressed its potential as an effective tool to enhance collaborative and autonomous learning (Constantino & De Lorenzo, 2006) and as a way to assess written communication skills

(Wu, 2005). Blogs bridge the gap between the outside world and the classroom, offering an authentic purpose and an instant audience to communicate (Warschauer, 2004). In addition, they require learners to actively construct meaning, organize, synthesize and publish their thoughts in blog-entries which can in turn be archived for retrieval and review (Du & Wagner, 2007). Nellen (2000) stresses that blogs provide students with a tangible product to show for their efforts at the end of the semester and take pride in their work, by writing for a real audience and not just for the instructor to read and grade. Finally, Yuen and Yang (2008) consider blogs an ideal instrument for portfolio development due to their flexibility and interactive structure.

Portfolio assessment is strongly associated with the efforts to pursue a more authentic and realistic means of assessing a student's level of learning (Testerman and Hall, 2001) as it can map the process students go through when they draft, write and review, showing traces of their achievements and difficulties while they progress (Bishop, 1993). Portfolios offer various opportunities (O'Malley and Valdez Pierce, 1996; Tsagari, 2004) to education:
- comprising an on-going record of language development and a tangible evidence of overall performance in different skills,
- linking assessment with instruction, and
- promoting collaborative peer-assessment, constant feedback, responsibility for self-assessment, autonomy in learning, as well as creativity and excitement about learning.

The originality of the present study lies in that it attempts to merge the benefits of blogs and portfolios, creating a class-blog as a venue for authentic written communication among secondary students and as a platform where their writing web-portfolios (blogfolios) can be exhibited, shared, updated and evaluated.

Methodology

Research questions

A small-scale, project-based classroom research was carried out in order to draw conclusions regarding the effect of the class-blog on students' writing performance. It addressed the following research questions:

a) How can blogging engage students in process writing?
b) Can blog-writing increase learners' motivation?
c) Can students' blogfolios gradually transform them into competent, autonomous and self-reflective writers?

Context of study

The present study was conducted in a Greek State Junior High School situated in a rural part of Crete. The participants were 12 fourteen-year-old students attending English as a foreign language twice a week. Their linguistic background in English ranged from Threshold (B1) to Vantage (B2) level (Council of Europe, 2001) based on a level test administered to them at the beginning of the school year. Being their English teacher for two years, the researcher has regularly noticed the unwillingness and difficulties her students experienced when expressing themselves in writing as well as their enthusiasm every time new technologies were used in class.

Implementation of blogfolios in class

The project followed the stages for classroom-based portfolios suggested in Tsagari (2004, pp. 197-208). The first step was to determine the teaching goal to assess through students' writing blogfolios. The next step was to decide on the format of the blog and specify the blogfolio content (Appendix I). Then, the teacher introduced the idea to the participants and notified the School Head Master as well as students' parents. Afterwards, she set up the blog, she created email accounts and log-in names for the students and embedded them as authors, retaining administrator's privileges. Before setting the project into action, the teacher gave clear instructions for the procedure; she also devoted class-time to familiarise learners with the various blog functions and practiced process-writing, self- and peer-assessment, posting and commenting.

The writing tasks

Four 120-minute blog sessions were designed, published on the blog (www.daskalogi.edublogs.org), and administered covering four distinct writing genres and text types: 1) an email of advice to a friend with anorexia, 2) a descriptive article on famous landmarks, 3) a balanced essay on using technology in education and 4) a narrative on natural disasters. By the end of the project, students produced nine posts and fifteen comments each.

To prevent students from seeing blogging as an extra burden, there was task continuity between the blog-sessions in the computer lab and the in-class coursebook activities. More specifically, the writing tasks served as a post-writing activity of the material being taught to assess learners' mastery of the linguistic and functional forms dictated by the relevant unit of their textbook (Appendix II). Although the topics and genres practiced remained the same as in students' coursebook, the tasks were adapted, using authentic input from the World Wide Web, so as to meet students' needs and add a flavor of authenticity to their writing. Regarding task design, eight criteria had been taken into account: a) comprehensibility of instructions, b) clear task environment, c) variety of topics and writing genres, d) authenticity of input, e) variety of writing strategies required, f) equal opportunities for participation, g) enough time given for task completion and h) skills integration (listening to video input and to classmates in the planning stage, speaking while reporting ideas to class or during pair work, reading class and peer posts, and writing comments and posts on the blog). What is more, the tasks displayed *product authenticity,* replicating a real-world text, and *process authenticity* since learners worked through the full writing process, from planning to reviewing.

The blog-writing lessons

The blog-lessons were divided into stages (Flower and Hayes, 1980) and had the same format so that students could develop a work routine:

1. *Planning stage:* Students responded to different blog-entries uploaded on the blog by leaving a comment.

2. *Translating stage:* Students used ideas from the previous tasks and -with the aid of a writing plan (Appendix III) – they produced the first draft of their written work, which was then saved on their personal blogfolio.

3. *Reviewing stage:* Students evaluated their draft based on a self-evaluation checklist (Appendix IV), implemented changes and posted it. In rotation, they commented on each other's posts using a peer-assessment checklist and an evaluation form for each writing genre (Appendix V). The checklists were presented in English to familiarise students with important English structures and language that they would use during peer-commenting. Then, the teacher provided focused and indirectly evaluative feedback. After taking the received comments into account, students revised and produced their final posts.

Data collection

In order to increase the validity and acceptability of the findings, both qualitative and quantitative data was collected using multiple sources:

Classroom observation

Classroom observation was employed in the form of an open-ended diary with anecdotal records of events. According to Tsagari (2004, p. 284), "anecdotal records are factual, non-judgmental notes of students' activity or behavior." While students were blogging, the teacher kept notes of all the things that were taking place in the computer lab: students' motivation and co-operation, their contributions to peer feedback, their progress and perceived difficulties as well as unexpected events and delays. An observation grid was also designed to check students' use of writing strategies.

Collection of writing samples and peer feedback

As soon as each blog session ended, the teacher collected and analysed students' comments on their peers' writing and on the class posts, as well as the writing samples on their blogfolios, to gain insight into their writing and editing skills. Students' final posts were also marked by the teacher with reference to an analytic rating scale for assessing writing. Each scale contained descriptors[1] in note form, which outlined the student's expected writing performance in terms of task achievement, language use, organisation, and mechanics (Appendix VI, a). Assessment was further enriched by teacher's comments on process writing issues (Appendix VI, b) as evidence of their overall writing performance. The results were passed on to students in individual conferences.

Overall Evaluation Questionnaire

At the end of the project, an evaluation questionnaire -embedded in the blog- was answered on-line by the learners, regarding the impact of the blogfolio experience on their writing performance, feedback provision, participation and motivation. The questionnaire consisted of eight structured, close-ended questions for reasons of practicality and reliability (Richards, 2001). A five-point Likert scale was adopted with 5 representing a strong agreement and 1 representing a strong disagreement for each item. The questionnaire was administered in English and explained in Greek to ensure that all students had the same understanding of the questions.

---

[1] The rating scale was structured according to the guidelines provided by the Council of Europe (1996, pp. 118-120) as well as in accordance with the rating scale for writing proposed by Cohen (1994, pp. 328-329).

Evaluative blog post

The questionnaire was supplemented by a final post named "To blog or not to blog" in the form of a free writing task, encouraging learners to reflect on their difficulties and share their blogging experience.

Data analysis

The quantitative data obtained from students' responses in the questionnaire was used to identify salient variables with regard to the goals of the study. Six analytical categories were developed: 1) planning, 2) translating, 3) reviewing and feedback provision, 4) motivation, 5) writing development and 6) writing autonomy. A qualitative analysis of teacher's observation notes, students' writing samples as well as their comments on peer posts and on the evaluative blog-post was subsequently carried out, in an attempt to examine how the sorting categories from the questionnaire data were reflected and further elaborated in students' writings.

Results

Based on the processing of the aforementioned qualitative and quantitative data, certain conclusions can be drawn providing answers to the research questions:

Students' engagement in process writing

Regarding the process of planning and translating, Table 1 summarises learners' responses to the questionnaire. Students unanimously reported that blog entries and videos triggered their background knowledge on a specific writing topic and encouraged them to come up with relevant ideas and information (Table 1, No. 1). Moreover, nearly everyone agreed that the reflection comments under each class-post helped them plan their writing (Table 1, No. 2). Responses also show that the writing guides and the self-evaluation checklists - retrieved from the blog - helped everyone structure and elaborate the information and ideas gathered at the planning stage, so as to meet the goal of the writing task (Table 1, No. 3). Therefore, blogging facilitated the interpretation of the task environment, triggered brainstorming and enabled instant retrieval of useful ideas and information which are essential in order to set goals, generate ideas and translate them into words.

Part 1: While I was planning what to write….	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1. Videos and class posts were a great source of information and ideas				7	5
2. The comments on every class post helped me plan my writing.			2	6	4
Part 2: Using different documents on the class blog					
3. The writing guides and self-evaluation checklists helped me organise my writing				4	8

Table 1. Planning and translating

Qualitative data from teacher's observation notes further shows the extent to which blog-functions facilitated, and engaged students in, planning and drafting:

"Nick and Maria click on Google search and on-line dictionary to find more information and look up words. Emma clicks on the Google translator to check if what she is writing is correct. Paul and Kostas try to remember what we discussed in the previous session and click the comments button to read again what their classmates had written…..All of them try to organise the ideas they had gathered following the relevant model plan". (Teacher's notes, 15 December 2010)

Regarding reviewing, qualitative data analysis of students' writing samples reveals that blog-writing encouraged learners to revise their drafts, after taking into consideration the comments and alternative suggestions from teacher and peers. Reference to the exact location or to the kind of the expected change was necessary for learners to take advantage of peer or teacher comments and make changes. Consequently, the longer and more constructive the comments became as time went by, the more students identified and corrected their postings for content, language and mechanical errors. To illustrate the above, the quotes below show an extract from a student's first draft, then the feedback by the student's peer and finally the student's revised draft following peer feedback:

"On the other hand, some of the advantages are: first it will be harm for children. Moreover, technology harms the environment. Finally children spend a lot of time in front of screams and they don't study their lessons." (Student first draft, from the essay 'Technology in education', 10 March 2011)

"You could explain some things in more detail. It is not right to say 'some disadvantages are' because it is like you list things. Use fuller sentences and explain why you believe this. Also, the first 'harm' you use is not correct because you need an adjective there. Finally, you have spelling mistakes (screams) and you must put commas after linking words. Keep trying!" (Peer feedback, 16 March 2011)

"On the other hand, technology has got many disadvantages. First of all, it can be harmful for children because many hours on the internet may isolate them from their friends. Moreover, technology can harm the environment because computers need electricity and consume a lot of energy. Finally, students spend all their free time in front of the screens watching videos or playing games and as a result they don't study their lessons". (Student's revised draft, 18 March 2011)

The content of revisions was addressed by Questions 11-14 in the questionnaire (Table 2). Students seemed to revise both content and form, with grammatical changes being the commonest (Table 2, No11). Nine students answered that they rarely or sometimes changed their words (Table 2, No 12). This could be attributed to their access to lexical resources while planning and web-searching via the blog. The surprisingly high number of students reporting that they seldom or rarely rewrote and reordered whole paragraphs (students' responses, No 13) shows that with the guidance of the model plans and writing tips they had at their disposal, learners managed to organise their ideas and information appropriately. Therefore, guided instruction and scaffolding are necessary when writing. Moreover, eight of them regularly added new ideas (Table 2, No 14). This shows their eagerness to elaborate and justify their initial thoughts in order to produce a text that would satisfy their peers.

Part 3: While I was revising and editing my final post	Never	Rarely	Sometimes	Often	Always
11. I corrected grammar mistakes.		2	3	4	3
12. I improved vocabulary.		4	5	2	1
13. I rewrote and reordered whole paragraphs.	3	7	2		
14. I added new ideas.	3	1	5	3	

Table 2. The content of revisions

The summary of responses in Table 3 reveals that students experienced the positive impact of feedback on reviewing their work. All of them found peer comments more helpful than traditional teacher's corrections (Table 3, No. 10) since they were encouraged to look back at their texts, find and correct mistakes on their own (Table 3, No 9), and consequently improve their writing (Table 3, No. 8).

Part 4: My teacher's and classmates' comments …	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
8. helped me improve my writing.				5	7
9. helped me find my mistakes and correct them on my own.				6	
					6
10. were useless. I prefer when the teacher corrects my mistakes.	9	3			

Table 3. Reviewing drafts

In terms of feedback provision, ten students viewed commenting on their partner's post as an interesting and useful procedure because not only did they help their friend revise his/her work (Table 4, No. 4), but they also spotted mistakes that they made as well (Table 4, No. 6). It is interesting that only three students had difficulties in finding mistakes (Table 4, No. 5). This low number is probably due to the fact that the peer-edit sheet guided them while reviewing their classmate's writing as well as while structuring their comments (Table 4, No.7).

Part 5: Commenting on my partner's post	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
4. was interesting because I could help my friend improve his/her work.		2		7	3
5. was difficult. I couldn't find mistakes.	3	1	5	3	
6. was very helpful. Spotting my friend's mistakes helped me improve my mistakes.		2		5	5
7. The peer-edit sheet helped me spot mistakes on my partner's post and advise him/her on how to improve it.				7	5

Table 4. Feedback provision

Data from students' responses in the evaluative blog-post further shows the importance of instant feedback provision via the blog in learning to write. For instance:

> "Comments had a good result on me. I learnt to correct my mistakes like grammar, spelling and syntax and improve my writing. Through the blog you can learn to be more careful with some mistakes, you have seen to do your

classmates or they have told you to correct." (Students' comments, March 2011)

Student's motivation towards blog writing

Qualitative data from teacher's observation notes makes evident that the attitude of students towards writing, which until that point was a burden for them and was usually restricted in handing their compositions to the teacher, changed for the better, adding an element of fun, authenticity, purpose and real communication to the writing lessons. Here is an example:

"By the moment students log in the blog, they are so intrigued with the new class post. They are sensitised by Sara's email and keep asking me if Sara is a real person. I can't reveal them the truth and let them believe that Sara actually saw the video on anorexia in the blog and decided to write to them. They are excited that our blog gave them the opportunity to help someone real". (Teacher's notes, 14 February 2011).

Blog-writing also boosted students' participation and co-operation. For instance:

"When the bell rings, George is still writing. Since he is a bit slow in typing, Manolis offers his help. Amazing! During the break, George is reading his corrected essay and Manolis is typing it!" (Teacher's notes, 17 March 2011)

Part 6: Using the blog….	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
21. made me pay more attention to the lesson				4	8
22. was like communicating with others in real life	-	2	4	4	2
23. made me enjoy writing more	-	-	1	5	6
24. I felt excited that others could read my posts.	-	1	3	3	5
25. I was ashamed that my classmates would laugh at my mistakes	6	3	1	2	-
Part 7: I prefer...					
26. writing traditional compositions with pen and paper	6	1	5	-	-
27. writing on the blog	-	-	4	2	6
28. I don't mind. Both ways of writing are boring.	10	1	1	-	-
29. I don't mind. Both ways of writing are exciting.	-	-	5	3	4

Table 5. Motivation

Data derived from the questionnaire regarding students' motivation and blog-writing preferences showed that half of the students considered blogging a way of communicating with others in real life (Table 5, No. 22). What is more, eight students felt excited rather than ashamed that other people could read their posts (Table 5, No. 24 and 25). In addition, all of them agreed that blogging raised their attention and made writing more enjoyable (Table 5,

No. 21 and 23). All the students expressed positive feelings towards writing with eight preferring blogging to conventional pen-and-paper writing (Table 5, No. 26-29). All in all, blogging enhanced participation and interest in the lesson, and made students see writing as a worth-while process.

Writing development

Table 6 presents the participants' responses regarding the effect of blogging on their writing skills. It is evident that according to the students, blogging increased the quantity, quality and ease of their writing (Table 6, No.15-17). Furthermore, all the students agreed that the blog served as an on-line archive where they could organise and showcase their written achievements to classmates (Table 6, No.18 &19). What is more, blog-writing familiarised them with the use and functions of computers, leading to their typing skills being improved (Table 6, No. 20).

Part 8: Writing on the blog helped me….	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
15. write more than on paper	-	-	1	3	8
16. write better	-	-	1	4	7
17. write more quickly and more easily	-	2	1	4	5
18. plan and organize information and ideas	-	-	1	6	5
19. archive my work and showcase my progress in writing	-	-	1	6	5
20. improve my computer skills	-	-	2	1	9

Table 6. Skills development

Writing development was also verified by comparing students' qualitative accounts. The samples below and in Appendix VII depict learners' progress from simpler and shorter texts to more complex and longer ones, with fewer mistakes.

> *"The Parthenon is an amazing sight for all people.It is located in the Akropolis of Athens.Many tourists visit Parthenon every day.Parthenon was made by Kalikratis and Iktinos.If you go on holiday you must visit Parthenon.It is a brilliant sight for holiday. Parthenon is an ancient greek building and also many people visit this building for the his beauty.I reccomend you to visit this amazing sight".* (Student first draft, from the essay Famous landmarks, February 9, 2011)

> *"The Parthenon is an amazing sight for all people .It is located in the Akropolis of Athens. Many tourists visit the Parthenon every day to admire this classical and impressive building.*
> *The Parthenon was originally made by Kalikratis and Iktinos as a temple for the Greek goddess Athena, who she protected the city of Athens in the past. The construction began in 447 BC and finished in 438 BC. Fidias made the sculptures of the temple which are wonderful. The Parthenon was and is the symbol of Democracy and was built during the Gold Century of Pericles. Unfortunately, many years ago, a man called Elgin stole the marbles of the Parthenon which are now in the British Museum in London. Greece is trying to bring back the Elgin marbles but without success. Today, opposite from the Parthenon, there*

is the new museum of Acropolis and inside it the visitors can admire the marbles and statues of the Parthenon.
All in all, the Parthenon is an ancient Greek building and many people visit this building for its classical beauty. I recommend you to visit this amazing sight, if you go on holiday to Greece. It is a brilliant sight to explore for holiday."
(Student revised draft, February 13, 2011)

Students' commenting skills were improved as well. In the initial blog sessions, peer-comments were short and general referring to the overall impact of the post. Some students even used their mother tongue to express what they wanted to say:

"Of course you have done a very good post!!!!!!!! Κάτι ακόμα που θα ήθελα να αναφέρεις είναι για ποιον λόγο χτίστηκε ο Πύργος της Πίζας…….also you follow the photocopy that Mrs gave us." (Peer feedback on "Famous landmarks", December 2010).

By the end of the project their comments became supportive and constructive. The readers/editors not only pinpointed the exact error observed but they also provided tangible help for the writer in the reviewing stage; in particular, they suggested alternative ways of expressing words, phrases, sentences or even whole paragraphs so as to help their peers accomplish the goals set by the task:

"Dear Irene, your essay is pretty good. Some of your sentences made me admire your essay but I also noticed some mistakes which made it confusing. First of all, in your introduction there are some grammar and spelling mistakes like 'a easy'. You have also repeated many times the words 'technology' and 'students' in every part of your essay. I think you should replace these words with others that have the same meaning like 'computers' and 'learners'. In the second paragraph you should put 'on the one hand' in the beginning… and use linking words. Don't be disappointed with my notices. I thought that in this way you'll improve your essay and make it perfect." (Peer feedback on "Technology in Education", February 2011)

Teacher's observation notes concerning students' use of writing strategies were quantitatively analysed and summarised in Table 7 below:

Criterion	Number of Students using writing strategies	
	1st blog session	4th blog session
Use pre-writing techniques (Brainstorming, planning organizing ideas)	5	12
Identify context and style of writing	5	11
Revise writing before posting first draft	1	11
Use conventions of grammar, mechanics and usage correctly in the first draft	4	Improvement in all posts depending on student's level
Proofread and comment using the peer edit checklist	9 (surface feedback)	12 (constructive feedback)
Revise final post effectively based on received feedback	8	11

Elaborate ideas	4	12
Write effectively in different genres	3	10
Complete writing tasks on time	9	1

Table 7. Teacher's observation grids

The increasing number of students demonstrates their improvement in process-writing skills from the first to the final blog-session. Learners gradually developed cognitive skills (generating ideas while planning, decision making when drafting and reviewing), metacognitive skills (editing, peer evaluation, self-reflection) and became strategic in order to meet readers' expectations and task requirements.

Students' writing autonomy

Autonomy is defined by a capacity for critical reflection, decision making, independent action and responsibility for one's own choices (Crabbe, 1999) and it was addressed by Questions 31 and 32 (Table 8). The results confirm that blog-writing allowed learners to work and learn by themselves, to reflect on their mistakes, writing progress and difficulties and consequently to take pride in their achievements.

Part 9: *Autonomous learning*	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
31. Writing on the blog allowed me to work at my own pace and learn on my own.	-	-	1	6	5
32. Through my blogfolio I could reflect on my strengths and difficulties and be proud of my progress.	-	-	2	1	9

Table 8. Writing autonomy

Data from participants' self-evaluative comments shows that while blogging students started to see their work critically, reflecting on what they have managed as well as acknowledging what they did wrong or what they should have done. For example:

> *"When I entered my blogfolio I saw the huge difference between my first writing and my second. I cannot accept that the first was mine. I had many grammar mistakes, I hadn't used rich vocabulary and my sentences were incomplete. When I noticed my second post, I felt proud of myself. I completed my paragraphs, used better vocabulary, corrected grammar and I explained my ideas better. However, I could have divided the main body into 2 paragraphs."* (Students' self-evaluative comment, March 2011)

Nonetheless, learning does not take place in a vacuum, and independence does not necessarily imply learning on one's own. Interaction, collaboration and negotiation are important factors in promoting writing autonomy (Bhattacharya and Chauhan, 2010). Consequently, by encouraging peer-feedback, students did not only learn to evaluate each other's attempts throughout the writing process, but they were also trained to look at their own written work critically, at a later stage in the course. For example:

> *'Commenting on my classmate's post helped to find my mistakes easier and correct them alone'*. (Students' comments on the evaluative blog-post, March 2011)

Discussion

In terms of the first research question, it has been shown that the blog familiarises learners with process-writing as it involves reading each other's posts, interacting and commenting on these posts as well as writing new ones in response to received feedback. According to the findings, authentic blog input offers a real purpose for writing, activating students' background knowledge and content–related schemata to plan their drafts. When drafting, learners exploit already generated material and organise it based on the writing plans and checklists retrieved from the blog. This is in accordance with the process-writing principles according to which teachers should set purposeful tasks that stimulate thinking and brainstorming, in order to develop students' strategic competence for planning, while they should provide modelled instruction to guide learners when drafting (Hyland, 2002). Data further reveals that, when writing on a blog, the cognitive process of reviewing develops significantly and feedback plays a major role in making revisions. Blog makes effective peer feedback available, engaging students in extensive reviewing in order to meet their classmates' expectations. Similarly, an action-research project, carried out by Quintero (2008) in a Colombian university, showed that feedback is beneficial in the EFL writing process, enabling students to progress from simpler to more complex texts.

As far as the second research question is concerned, the student's active participation and accomplishment of all the tasks were the greatest evidence of their enhanced motivation. Data indicates that the blog motivates students to write, as it is not common for them to stay in the classroom during the break in order to finish their writing, or to volunteer to type their friend's post. This comes in agreement with Campbell's (2003) view that blogging improves learners' attitude towards writing. Moreover, it has been shown that blogging simulates a way of real-life communication, extending learning beyond classroom walls. This finding was also supported by Fellner and Apple (2006). What is more, in agreement with Huffaker (2005), the prospect of helping their classmates improve their work as well as of sharing their thoughts and ideas on line increases their interest in and attention to the lesson, facilitating collaboration.

Regarding the third research question, it can be argued that students gradually improved their writing skills while they further acquired cognitive and metacognitive skills. The blog emerged as a platform for showcasing written achievements, as was also supported by Du and Wagner (2007). The idea that their work can be viewed enhances students' pride and self-esteem, as well as their eagerness to write more and better. These findings come to verify what was stated by Nellen (2000) and Campbell (2003), regarding the role of blogs in encouraging learners to demonstrate their progress to a mass audience and improve the quantity and quality of their writing. Finally, through their blogfolios students have gradually become autonomous, since they developed self-reflective skills, they gained responsibility of their work and that of others, and they started to see their writing development as centred in themselves, rather in the teacher (Hedge, 2000).

Pedagogical implications

The findings of this small-scale research clearly demonstrate the potential of integrating a class blog in an EFL course in order to promote the approach of writing as a process, to

foster learning growth by developing students' writing, cognitive, metacognitive and communication skills, to increase motivation in the classroom and provide a venue for self-reflection and autonomy where learners can showcase their progress to everyone interested. The amount and quality of students' posts showed that feedback plays a scaffolding role in their writing progress. Moreover, the blog increased the amount of attention paid to content and accuracy due to the existence of a real audience who could see and read their work. Additionally, it facilitated effective peer and self assessment which are undoubtedly useful skills for students to have as they embark on their journey of life-long learning. Finally, the class-blog created a relaxing, interactive, student-centred learning environment. Nevertheless, learning with this tool can only be effective if students receive clear instructions and guidelines for writing posts as well as feedback and work routines in order to embrace the full potential of blogs in education. Information Technology support in every school and teacher training are also indispensable.

Limitations

The research presents a number of limitations. First of all, the sample was small, rendering the results rather tentative. Moreover, due to time constraints, the duration of the project and the number of blog-sessions was minimal. Issues of generalizability were also raised as the study was based on a specific situation, measuring the writing development of 2nd grade learners in a small, regional, state school. The particular students were not well-acquainted with computers, not to mention with blogs, whereas due to their socioeconomic background some of them did not have computers at home. Last but not least, because of the problematic internet access at school and the poor quality of the computers in the lab, various technical problems occurred, making the lessons rather time-consuming and resulting in useful data being lost.

Concluding remarks

The study has succeeded in adding findings to the future and very promising field of using blogs in the teaching and assessment of EFL writing. It would be interesting to investigate, whether blogs could be employed to develop and measure other skills in English courses or whether they could be gradually implemented in all Greek schools, complementing existing teaching materials and changing traditional, summative testing practices in every school subject. The integration of weblogs in pedagogies can expand and flourish only if teachers receive appropriate training, support innovation in their classrooms, transfer the ownership of learning (even of blogs) to the students as well as ensure dialogue and partnerships between schools, universities, teachers, community and learners about new approaches to learning that involve collaboration and student-centeredness. There are "*now richer and more engaging pathways to learn than ever before, but this calls for us to engage with the new tools and gain a deeper understanding of their potential for enabling choice, creativity and self-direction for learners*" (McLoughlin and Lee, 2010, p. 672).

Authors' email: daskalogi@gmail.com

References

Allford, D. & Pachler, N. (2007). *Language, autonomy and the new learning environments.* Oxford: Peter Lang.

Bhattacharaya, A. and Chauhan, K. (2010). 'Augmenting learner autonomy through blogging'. *ELT Journal*, 64/4: 376-384.

Bishop, W. (1993). 'Not only assessment: teachers talk about writing portfolios.' *Journal Of Teaching Writing*, 12/1: 33-55.

Campbell, A. P. (2003). 'Weblogs for use with ESL classes'. *The Internet TESL Journal*, 9/2.

Cohen, A. D. (1994). *Assessing language ability in the classroom (2nd ed).* Boston, Mass.: Heinle and Heinle.

Constantino, P.M., & De Lorenzo, M. N. (2006). *Developing a professional teaching portfolio: A guide for success (2nd ed).* Boston, MA: Allyn & Bacon.

Council of Europe (1996). *Common European Framework of Reference for language learning and teaching.* Strasbourg: Council of Europe.

Council of Europe (2001). *Common European Framework of Reference for languages: learning, teaching, assessment.* Cambridge: Cambridge University Press.

Crabbe, D. (1999). 'Defining the field: introduction'. In S. Cotterall and D. Crabbe (Eds.), *Learner autonomy in language learning*: *defining the field and effecting change*. Frankfurt: Bayreuth.

Du, H.S., & Wagner, C. (2007). 'Learning with weblogs: Enhancing cognitive and social knowledge construction'. *IEEE Transactions on Professional Communication*, 50/1: 1-16.

Fellner, T. & Apple M. (2006). 'Developing writing fluency and lexical complexity with blogs'. The *JALT CALL Journal*, 2/ 1:15-26.

Flower, L.S. & Hayes J.R. (1980). 'The dynamics of composing: making plans and juggling constraints'. In L.W. Gregg & E.R. Steinberg (Eds.), *Cognitive processes in writing*. Hillsdale, N.J.: Erlbaum.

Hedge, T. (2000). *Teaching and learning in the language classroom*. Oxford: Oxford University Press.

Huffaker, D. (2005). 'The educated blogger: using weblogs to promote literacy in the classroom'. *AACE Journal*, 13/2, 91-98.

Hyland, K. (2002). *Teaching and researching writing*. London: Pearson Education.

McLoughlin, C. & Lee, M. J. W. (2010). 'Personalised and self regulated learning in the web 2.0. era: international exemplars of innovative pedagogy using social software.' AJET journal, 26/1: 28-43 (March), accessed at http://ascilite.org.au/ajet/ajet26/mcloughlin.html accessed 4 November 2010.

Nellen, T. (2000). *The webfolio: why write to the Internet*, accessed at www.tnellen.net/cyberenglish/webfolio.html on 19 Aug 2010.

O'Malley, J. M. & Valdez Pierce, L. (1996). *Authentic assessment for English language learners.* New York: Addison-Wesley.

Quintero, L.M. (2008). 'Blogging: a way to foster EFL writing'. *Colombian Applied Linguistic Journal*, 10: 7-49.

Richards, J. C. (2001). *Curriculum development in language teaching*. New York: Cambridge University Press.

Silva, T. (1990). 'Second language composition instruction: developments, issues and directions in ESL'. In B. Kroll (Ed), *Second language writing: research insights for the classroom.* Cambridge: Cambridge University Press, 11-23.

Song, H. & Wang, S. (2005). 'Assessing EFL's language use via electronic portfolios: a pedagogical framework', accessed at http://www.editlib.org/p/18981 on 10 Sep 2010.

Testerman, J.K. & Hall, H. D. (2001). 'The electronic portfolio': A means of preparing leaders for application of technology in education.' *Journal of Educational Technology Systems,* 29/3: 199-206.

Tsagari**,** C. (2004). 'Alternative methods of assessment'. In C. Tsagari & R. West, *Testing and assessment in language learning: assessing students without tests, vol. 3.* Patras: Hellenic Open University, 119-341.

Warschauer, M. (2004). 'Technological change and the future of CALL'. In S. Fotos & C. M. Browne (Eds.), *New perspectives on CALL for second language classrooms,* New Jersey: Lawrence Erlbaum Associates, 15-24.

Wu, W.S. (2005). 'Using blogs in an EFL writing class'. Unpublished paper, Chung Hua University, accessed at www.chu.edu.tw/~wswu/publications/papers/book_chapters/01.pdf on 13 Oct 2010.

Yuen, S.C.Y. & Yang, H. (2008). 'The Blogfolio approach for enhancing students' interaction and learning'. Paper presented at the World Conference on E-learning, accessed at http://www.editlib.org/p/30140 on 10 Nov 2010.

APPENDIX I

Blog implementation

Scoping	Notes
Purpose of the blog	✓ To develop and assess writing growth ✓ To create authentic and motivating learning opportunities ✓ To foster self-reflection, peer feedback and collaboration
Blog content	✓ reflection prompts (class posts) for Ss to brainstorm ideas, and write their first drafts ✓ Blog entries to set the writing task ✓ individual blogfolios set up by T where Ss archive and assess their work
Objectives	By the end of this project Ss should be able to ✓ practice process and collaborative writing ✓ use the blog as a means of expression and communication of opinions and ideas ✓ write for different purposes and real audiences using blog functions ✓ develop a critical eye to their work and that of others
Learning outcomes	To develop competent and autonomous writers
Meeting curriculum objectives	By designing purposeful tasks and writing lesson plans based on the Ss' course books.
Providing scaffolded learning	Via instructions for tasks and work routines, peer/self-evaluation task-sheets, guidelines for good commenting, and model plans for each writing genre.
Evaluating the effectiveness of blog writing	✓ By analyzing students' posts, comments and drafts, ✓ by observing their participation and work, ✓ via a questionnaire

The blog *"iCanWrite"* can be accessed at http://daskalogi.edublogs.org.

APPENDIX II

Matching blog and textbook writing tasks

Blog Entry Unit in "Think Teen 2 "	Written assignment title	Language function	Grammar form/Vocabulary
1. Devastation and horror during the Hurricane Katrina in New Orleans Unit 2, Lesson 5 "Across the ages"	Write a story about a natural disaster you have experienced	Describing natural disasters Talking about past events Talking about past actions in progress Expressing feelings and reactions	Simple Past Past continuous Vocabulary related to natural disasters
2. Famous Landmarks Unit 1, Lesson 3 "Different places, different buildings"	Write an article for the school newspaper describing a landmark	Article writing Describing a place or building Organising factual info Recommending	Present Simple Active and passive voice Relative clauses Descriptive adjectives Types of material
3. We are what we eat/ Could anyone help me? Unit 6, Lesson 16 "You are what you eat"	Write an email of advice to a friend with anorexia	Responding to an email Giving advice Making suggestions And Justifying them	Expressions of advice Modals Conditionals greetings and salutations Food/ health/eating habits
4. Educational change is only a click away Unit 4, Lesson 11 "Change through times"	Write an essay discussing the advantages and disadvantages of technology for learning	Discussing for and against a topic Expressing personal opinion Giving reasons	Conditionals, linking words of sequencing ideas, concluding, expressing opinion, reason, and contrast School issues and technological equipment

Think Teen – 2nd Grade of Junior High School (Προχωρημένοι) is the course book applied to meet the needs of the specific class and corresponds to Threshold B1 level (Council of Europe, 2001).

APPENDIX III

Sample of writing guide

Writing an essay

<u>INTRODUCTION</u> 1ST PARAGRAPH (not more than 5 lines)	✓ **You can use a general statement about the topic** e.g. *Nowadays, the Ministry of education believes that technology…/ encourages the use of technology…* ✓ **And then, restate the question using your own words:** *e.g. But are they right?*
<u>MAIN BODY</u> 2ND PARAGRAPH 3RD PARAGRAPH	✓ **Present the side of the issue you do not support. State at least two arguments in favor of this view.** e.g. On the one hand, using technology in the classroom can be time-consuming. ✓ **Present the side of the issue you do support. State at least two arguments in favor of this view and give examples to support it.** e.g. On the other hand, technology is an essential part of everyone's life nowadays and has various benefits for learning.
<u>CONCLUSION</u> 4TH PARAGRAPH (not more than 5 lines)	✓ **Sum up what you have said so far and state your opinion** e.g.In conclusion, I think that……It seems to me that…

APPENDIX IV

Self-evaluation checklist

Before answering the checklist below, try to read your post as an external reader

Think about the following points …..	Yes	No	Revise if necessary…
Do you have a clear picture of the topic?			
Have you included all the points stated in the task rubrics and model plan? Is there something missing?			
Are all the sentences complete (subject-verb-object)?			
Have you used conjunctions (but, and, however, also, because, etc) to join your sentences and make your text coherent?			
Have you used correct and rich vocabulary or do you repeat the same words again and again?			
Are there any spelling mistakes or grammar mistakes (check tenses, plural/singular, etc)			
Are there any points you are not sure about and where you need help from teacher or classmates?			

Are you ready to PUBLISH your post?

APPENDIX V

Sample of peer-edit checklists and evaluation forms

A.   WRITING AN EMAIL OF ADVICE

Put a √ (to indicate a satisfactory remark) and an X (to indicate a problem) in the boxes below to help you write your comment on your partner's post.

▪ <u>**Spelling, capital letters and punctuation**</u>

_____Check for spelling mistakes. Indicate any words you think might be wrong.
_____Check that a capital letter is used in the first letter of each sentence and in the personal pronoun "I".
_____Check that commas (,), full stops (.), question marks (?), exclamation (!) and quotation marks ("") are used correctly.

▪ <u>**Grammar and vocabulary**</u>

_____All the sentences are complete: they have a subject, a verb and an object.
_____All the sentences make sense and are not confusing.
_____Your classmate has used rich vocabulary (linking words, expressions for giving advice)
_____Your classmate doesn't repeat the same words again and again.

▪ <u>**Structure**</u>
_____The email has an appropriate greeting and ending.
_____The email has all key parts (introduction, main body, and conclusion).
_____The paragraphs are clearly organized.
_____The main paragraph contains at least 3 pieces of advice.
_____Your classmate clearly explains how each piece of advice will help.

▪ <u>**Impact**</u>
_____The advice given is convincing and helpful.
_____The email shows understanding and friendliness.

B. EVALUATION FORM			
Check point	**Wow, You did a great job! Well done!**	**Hey, pretty good. You should be proud.**	**Oops! You need to do some more work!**
Format	Your email has all key parts (greeting, introduction main body, conclusion)	Your email has all the key parts but they are not equally structured or fully developed.	Your email is missing some important parts.
Content: theme & expression	You have given convincing and helpful advice. You use rich and clear language.	Your email meets the task but some parts are rather weak or make no sense. Your language is good but could be richer or clearer.	Your email is too simple and your sentences are disconnected. Your advice won't help. Your language is poor and confusing
Grammar and Vocabulary	You have no or minimal errors!	You have some spelling and sentence mistakes. Use more expressions to give advice!	You have many sentence and vocabulary mistakes!

APPENDIX VI

Scoring procedures

A. ANALYTIC RATING SCALE

Scale	Task Achievement	Language use (Accuracy)	Organization	Mechanics
5 Excellent	The task is carried out successfully and very easily. All required information is given in an effective way	Mastery of grammar with only 1 or 2 minor mistakes Full control of complex structures Very effective choice of words Wide range of linking words	Very well organised and perfectly coherent	Mastery of spelling and punctuation
4 Very good	The task is carried out successfully and with relative ease. Good manipulation of information and knowledge of the topic	A few minor mistakes only (prepositions, articles, etc.) Good control of structure Effective choice of words and word forms with only a few errors	Fairly well organised and generally coherent	Few errors in spelling and punctuation
3 Good	The task is adequately developed and shows basic knowledge of the topic. It lacks detail. Some parts are incomplete or unclear	Some major mistakes in the use of tenses and a few minor ones without hindering communication Adequate choice of words but some misuse of words and idioms	Loosely organised but main ideas clear. Basic cohesion. Logical but incomplete sequencing	Fair number of spelling and punctuation errors
2 Poor	The task is inadequately developed. Limited knowledge of the topic. Some parts are missing	Major mistakes leading to difficulty in understanding. Poor control of sentence structure. Limited range of vocabulary (often repetitive and confusing)	It needs effort to understand. Ideas disconnected; lacks logical sequencing	Frequent errors in spelling and punctuation that make understanding hard.

1 Very poor	The task does not prove knowledge of the topic. There is not enough to evaluate!	Numerous serious mistakes. No control of sentence structure. Almost unintelligible. Very limited range and very poor knowledge of words.	Very hard to understand. No organisation; incoherent Communication is hindered.	Numerous errors in spelling and punctuation. Several words are hard to recognize.

B. TEACHER COMMENTS ON PROCESS WRITING ISSUES

FIRST DRAFT 1. Has the student identified the specific audience and the purpose for writing? 2. Has the student use ideas from the comments shared with his/her classmates on the class post? 3. Has the student organised ideas based on the writing guide? 4. Has the student place a greater emphasis on content than on form?	
FINAL DRAFT 5. Has the student taken into consideration teacher's previous comments? 6. Has the student taken into consideration peer comments? 7. Based on received comments what changes has he/she made? ▪ Corrected grammar ▪ Improved vocabulary ▪ Added ideas ▪ Added linking words ▪ Corrected linking words ▪ Reorganised/reordered sentences and paragraphs ▪ Improved title ▪ Improved impact 8.Has the student made other changes than those pinpointed by teacher and peers?	
BETWEEN FIRST AND FINAL DRAFT 9.Has the student made substantive or minor changes? 10. Has the student identified and corrected mechanical errors? 11. Has the student identified content errors?	

APPENDIX VII

Comparing students' first and final writing samples

<u>Student first draft, from the essay Technology in education, March 14, 2011</u>

Nowadays, the Ministry of Education believes that technology is necessary for people and children. Technology can doing your life more interesting!

On the one hand, using technology in the classroom can students encourage, entertain and motivate for reading and they success your goals in the future. Students can work at their own pace and they can learn whatever interests them. Also, students find information on their own.

On the other hand, technology has got many disadvantages like: a lot of children spend their free time in front of the computer screen downloading favourite songs and movies from the Internet or playing computer games instead of studying their homework. Finally, technology spends a big amount of energy too.

To conclude, I think that tecnology is very useful for every student and all schools must use it.

<u>Student revised draft, March 17, 2011</u>

Nowadays, the Ministry of Education believes that technology is necessary for a student at the school. They say that technology can do the lessons more interesting! But are they right?

On the one hand, technology has got many disadvantages. Firstly, a lot of children spend their free time in front of the computer screen downloading favourite songs and movies from the Internet or playing computer games instead of studying their homework. Secondly, technology spends a big amount of energy too that is expensive and harms the environment.

On the other hand, when students use technology in the classroom , they are encouraged, entertained and motivated for reading. In that way, they will manage their goals in the future more easily. Moreover, students can work at their own pace and they can learn whatever interests them. At last, students find information in the Internet on their own and they become independent learners.

To conclude, technology has positive and negative results. In my opinion, if we use tecnology with measure it will help education. I also think that every school must change ,and for that, technology is necessary!!!

Online Diagnostic Assessment: Potential and Limitations (The case of DIALANG in the Greek Junior High School Context)

[Διαγνωστική Αξιολόγηση online: Δυνατότητες και Περιορισμοί (Η εφαρμογή του DIALANG σε Ελληνικό Γυμνάσιο)]

Vassiliki Baglantzi

This study explored the potential of DIALANG to serve placement purposes in the 1st grade of Greek state junior high school. The DIALANG system of language assessment is an application of the Common European Framework of Reference (Council of Europe, 2001), which assesses the user's level in five skills (listening, writing, reading, vocabulary and structures) and places them on one of the six (A1-C2) CEFR levels. It is a valid and reliable, online diagnostic test, which makes extensive use of self-assessment and provides several types of feedback. So far, DIALANG has been mainly used by adult learners in 14 languages in self-access, low-stakes contexts around Europe. However, the research focus in this study is whether DIALANG can replace teacher made placement tests for younger learners of English. Teacher-made tests are often criticized on the grounds of validity, reliability and content, therefore the use of a standardized alternative would facilitate placement procedures in a valid and reliable way. 20 students took DIALANG and then reported on their experience through a questionnaire. The students were also involved in self-assessing their reading, writing and listening skills using the DIALANG self-assessment statements. Their teacher also used the DIALANG scales to assess the students' proficiency level in these three skills. The research findings showed that DIALANG can be a highly practical and useful placement tool provided that it is limited to the assessment of three skills (listening, writing and reading) out of the five it can assess. Results also showed that neither the students nor the teacher are adequately familiar with the DIALANG statements and consequently the CEFR skills level descriptors, although these levels are extensively used in the description of textbooks, language courses and exams in the ELT field in Greece.

℣

Η παρούσα μελέτη ερευνά τις δυνατότητες της αξιολόγησης DIALANG να εξυπηρετήσει κατατακτήριους σκοπούς στην Α' τάξη του δημόσιου Γυμνασίου στην Ελλάδα. Το σύστημα γλωσσικής αξιολόγησης DIALANG είναι μια εφαρμογή του Κοινού Ευρωπαϊκού Πλαισίου Αναφοράς των Γλωσσών (Council of Europe, 2001), που αξιολογεί το επίπεδο του χρήστη σε πέντε δεξιότητες (κατανόηση προφορικού λόγου, παραγωγή γραπτού λόγου, κατανόηση γραπτού λόγου, λεξιλόγιο και δομές) και τον τοποθετεί σε ένα από τα έξι επίπεδα (Α1-Γ2)

του Πλαισίου. Είναι ένα έγκυρο και αξιόπιστο διαγνωστικό τεστ (Council of Europe, 2001), διαθέσιμο από το Διαδίκτυο, το οποίο κάνει εκτενή χρήση της αυτοαξιολόγησης και παρέχει διαφορετικά είδη ανατροφοδότησης. Ωστόσο, μέχρι τώρα έχει χρησιμοποιηθεί κυρίως από ενήλικους μαθητές 14 γλωσσών σε περιβάλλοντα αυτοπρόσβασης και χαμηλού διακυβεύματος σε όλη την Ευρώπη. Ο ερευνητικός στόχος αυτής της μελέτης είναι αν η αξιολόγηση DIALANG μπορεί να αντικαταστήσει τα κατατακτήρια τεστ που φτιάχνονται από τους καθηγητές και απευθύνονται σε μαθητές μικρότερης ηλικίας. Τα τεστ που φτιάχνονται από τους καθηγητές έχουν συχνά δεχτεί κριτική ως προς την εγκυρότητα, την αξιοπιστία και το περιεχόμενό τους, επομένως, εναλλακτικά η χρήση ενός σταθμισμένου τεστ, θα διευκόλυνε την κατατακτήρια διαδικασία με έγκυρο και αξιόπιστο τρόπο. 20 μαθητές έκαναν το τεστ DIALANG και συμπλήρωσαν ένα ερωτηματολόγιο σχετικά με την εμπειρία τους. Οι μαθητές επίσης αξιολόγησαν τις δεξιότητές τους στην κατανόηση και παραγωγή γραπτού λόγου και την κατανόηση προφορικού λόγου χρησιμοποιώντας τις προτάσεις αυτοαξιολόγησης του DIALANG. Επιπλέον, η καθηγήτριά τους χρησιμοποίησε τις κλίμακες DIALANG για να αξιολογήσει το επίπεδο των μαθητών της σε αυτές τις τρεις δεξιότητες. Τα ευρήματα της έρευνας έδειξαν ότι η αξιολόγηση DIALANG μπορεί να αποτελέσει ένα άκρως πρακτικό και χρήσιμο κατατακτήριο εργαλείο με την προϋπόθεση ότι θα περιοριστεί στην αξιολόγηση τριών δεξιοτήτων (κατανόηση και παραγωγή γραπτού λόγου και κατανόηση προφορικού λόγου) από τις πέντε που μπορεί να αξιολογήσει. Τα αποτελέσματα επίσης έδειξαν ότι ούτε οι μαθητές ούτε η καθηγήτρια είναι αρκετά εξοικειωμένοι με τις προτάσεις του DIALANG και κατ' επέκταση με την περιγραφή των επιπέδων δεξιοτήτων του Κοινού Ευρωπαϊκού Πλαισίου Αναφοράς των Γλωσσών, παρά το ότι τα επίπεδα αυτά χρησιμοποιούνται ευρέως στην περιγραφή βιβλίων, προγραμμάτων γλωσσών, και εξετάσεων στον τομέα της Διδασκαλίας της Αγγλικής Γλώσσας στην Ελλάδα.

Key words: DIALANG, diagnostic assessment, placement tests, self-assessment.

Introduction

Every year, in September, teachers of English in Greek state junior high schools give students of the 1st grade a placement test in order to stream them into two levels: less and more advanced learners (Ministry of Education, 1996). There are no ready-made, commercially available or specially-designed placement tests which can work for every institution and every language programme, since placement tests are constructed for particular situations and are based on the key features of different levels in each situation. As Hughes (2003, p. 17) states: "[Successful placement tests] *are tailor-made rather than bought off the peg.*" Thus, the design of a valid and reliable placement test is undertaken by the teachers.

Since placement is based on the exclusive responsibility of diagnosing what a pupil can or cannot do, using a diagnostic test could be an alternative. The problem in this case lies in the fact that diagnostic tests are scarce and very often other types of tests, such as proficiency or progress tests are used for diagnostic purposes (Alderson, 2005; Hughes, 2003).

The DIALANG language assessment system is an application of the Common European Framework of Reference (CEFR; Council of Europe, 2001). It is a free, online, diagnostic test which places test takers at the six CEFR levels (A1-C2) in five skills: listening, writing, reading, vocabulary and structures. It involves self- assessment and provides a variety of feedback types. Its validity and reliability have been established by groups of testing experts in Universities around Europe and it has been widely used as it can assess fourteen European

languages. Studies that report on the DIALANG system use (Floropoulou, 2002; Haahr et al, 2004; Huhta et al, 2002; Yang, 2003) refer to adult test takers who self-access DIALANG.

The present study researches the possibility of offering teachers of English in Greek state junior high schools an alternative placement procedure, which can replace teacher-made placement tests, whose validity and reliability are often criticized. Another aim is to assess students' familiarity with self-assessment procedures and shed light on the students' perceptions of the CEFR skill level descriptors. A third aim is to assess the extent to which the teacher can accurately match students' performance to the CEFR skill level descriptors and how familiarity with the CEFR can enhance the accuracy of their assessment.

The DIALANG test: components and feedback

When entering the DIALANG site (http://www.lancs.ac.uk/researchenterprise/ dialang/about), users first choose the language and the skill to be tested. They can also choose the language (mother tongue or other) of the test rubrics, self-assessment statements and feedback.

Having chosen the above, DIALANG users are presented with the Vocabulary Size Placement Test (VSPT). Their task is to study a list of verbs and decide which are real and which are not. Because of the assumption that "*the size of someone's vocabulary can be used as a rough guide to other language skills*"(Meara, 1992, p. 5) this test is used as a quick way of profiling the user's vocabulary size, and determines the level of difficulty (easy, medium, difficult) of the test that follows.

Another optional placement procedure is a set of self-assessment statements (18 for each skill) that are available in 3 skills (Listening, Writing and Reading). Vocabulary and Grammar are not included since they are more language specific and, therefore, difficult to develop in relation to all languages (Alderson, 2005). Test takers are presented with "Can do" statements, which are arranged in order of increasing difficulty, and are asked to click on the 'Yes' or 'No' answer. The aim of this part of the test is two-fold: On the one hand, combined with the VSPT it provides an estimate of the user's level, according to which DIALANG will be adjusted. More importantly, though, it is an opportunity for the users to reflect on their language ability and be involved in their own assessment. The system calculates the user's level and, when the test is over, it compares this level with the one decided according to the test results. It can provide possible explanations of any discrepancies between them in the feedback provided. Self-assessment is an optional component and the users can skip it, but, as with the VSPT, they are advised to take it and are warned about the consequences if they do not. If a user decides to skip both the VSPT and the self-assessment component, they receive a test of "medium" difficulty.

After users complete or skip the two introductory procedures, they receive the test of the skill and language they chose originally. They are offered the possibility to have immediate feedback on their answers item by item, which they can switch on and off throughout the test. DIALANG test items can have four different formats: multiple-choice, drop-down menus, text-entry and short-answer questions. Thus, users often need to type their answers. This requires basic typing skills.

DIALANG aims to provide informative and useful feedback (Alderson, 2005, p. 209), which takes a number of forms:

(1) *Placement Test Feedback:* It is offered as soon as the VSPT is completed and consists of a numerical score (out of 6 bands ranging from 1-1000).
(2) *Immediate Feedback (IF):* As soon as a test item is answered, a window pops up and informs the user if the answer is correct, and presents the correct option if it is not.
(3) *CEFR Level.*
(4) *Item Review*: Users are also given the chance to have an overview of their right and wrong answers and go back to a particular test item and review their answer.
(5) *Self-assessment and explanatory feedback:* Through this type of feedback users are informed whether their self-assessment agrees with DIALANG results.
(6) *Advisory feedback:* This section of the DIALANG system uses more detailed CEFR scales at each of the levels, for example, by including descriptions of text types. The user can also see what needs to be done so that they can reach a higher level.

Research studies on DIALANG

Since the completion of the DIALANG project and the availability of DIALANG to users, research has focused mainly on adult users. Floropoulou (2002) researched Greek and Chinese foreign language learners' attitudes to self-assessment and DIALANG. She found that even learners who were untrained in self-assessment managed to accurately self-assess their language levels making use of the self-assessment questionnaire of DIALANG. She also found that her Greek learners could make better use of DIALANG because it is based on the CEFR and because they could choose Greek as the language of instructions and feedback. Moreover, DIALANG proved to be very helpful for all her learners to diagnose their strengths and weaknesses while the majority agreed that the Advisory Feedback was the most useful part of DIALANG.

Yang (2003) investigated how test-takers use the DIALANG feedback. Her sample consisted of 12 postgraduate students at Lancaster University. Similarly to Floropoulou (2002), she found that Item Review and Advisory Feedback are the two most helpful forms of feedback. She also found that the explanatory feedback prompted her learners to reflect on their language learning processes which helped them set appropriate learning goals. Her study also indicated that the elaborated feedback DIALANG provides can facilitate learning more than a score would. DIALANG feedback helped learners realize the cognitive, affective and social factors involved in learning and provides suggestions for improvement. The researcher concluded that the availability of elaborated and comprehensive feedback by DIALANG was possible because it is a computer-based test.

Haahr et al (2004) proposed DIALANG as a tool for the assessment of language skills for adult speakers of European languages in search of employment. The report acknowledged the utility, validity and quality and relevance of skills definitions, but raised questions in relation to time requirements and the delivery platform. The final proposal was a shortened version of DIALANG, limited to the assessment of listening, reading and grammar.

Methodology

The context of the study

English Language instruction in Greek state schools has been regulated by the Cross-thematic Curriculum Framework for the English Language (Ministry of Education, 2001; DEPPS stands for the initials in Greek) since 2002. This curriculum was the basis of the textbooks that are currently used. According to the Pedagogic Institute, which is responsible

for the design of the curriculum, the students' level in this stage of education ranges from A2+ to B1+ and placement tests are recommended in case students were not taught English in primary school. In practice, all first grade students are given a placement test at the beginning of the first grade. English teachers in each school are responsible for the design, administration and marking of the placement test.

The need for diagnostic testing is also apparent in the DEPPS (2001, p. 29), and discussion of levels is done in relation to the CEFR and its skill level descriptors. However, it is unclear whether Greek teachers of English are familiar with the CEFR and how their students' performance is reflected in its descriptors. Therefore, teachers might need an assessment tool linked to the CEFR, to help them assess their students' levels and needs. Moreover, because the DEPPS promotes self-assessment, the question that arises is whether Greek students are experienced in self-assessment and familiar with the CEFR, so that they can benefit by being involved in the self-assessment opportunity DIALANG offers them.

Research Questions

Previous research studies on DIALANG all focused on adult learners. The present study aims to answer the following research questions:
1. Could DIALANG be used for placement purposes in Junior High School? How can it be administered and how can its 'practicality', 'utility' and 'discrimination' qualities be beneficial in the particular context?
2. Are Greek students experienced in self-assessment and familiar enough with the CEFR to benefit from DIALANG procedures and feedback?
3. Do Greek teachers of English need DIALANG to help them in the assessment of their students' level?

The case study and the sample

To research the questions above, a 'case study' of a group of students was conducted, which combines several types of data. It was considered to be an appropriate method because according to Cohen et al (2007, p. 254) "*case studies strive to portray 'what it is like' to be in a particular situation, to catch the close up reality and 'thick description' of participants' lived experiences of, thoughts about and feelings for a situation,[…] in its real-life context.*" Critics of the case study approach draw attention to the fact that it is not always possible to make generalizations out of it, but the researcher can hope that her case study can achieve 'relatability', that is, help decision-making of researchers and teachers in similar contexts (Bassey, 1981, in Bell, 2005, p. 11).

The sample on which this case study was conducted is a typical state school class, representative of the characteristics of populations in any state school around Greece. Therefore, research findings may not be generalizable, due to the small size of the sample, but will hopefully relate to other classes in similar situations.

20 students, 13 boys and 7 girls, aged 14-15 took part in the present study. Seven of them are of Albanian origin and one is bilingual Greek-Dutch. They were all born in Greece and have received Greek schooling since kindergarten. They all have been taught English since the 3rd grade of primary school, that is for seven years, and 18 have also had private tutoring or attended evening classes in private language schools.

 This class was chosen for a number of reasons. First, it had to be a 3ʳᵈ grade junior high school class since the students in grades 1 and 2 have been given placement tests and it was anticipated that DIALANG would yield a narrower range of levels. Moreover, it can be considered to be a typical mixed ability, multilingual, multicultural class, representative of the population in Greek state schools in many geographic areas in the last decade, due to the demographic changes in schools, as 10% of the student population in Greek public schools in 2006-07 consisted of repatriated and foreign students (Zachos, 2009, p. 142). Finally, it had to be a class which the teacher-researcher had taught before and knew well in order to be able to make more valid assessment decisions about what the students can and cannot do with the language.

Research procedure and tools

The study was conducted in three parts. First, the students took DIALANG, despite problems with installing and running the platform on their personal computers or the school computer lab. Then, they were presented with the DIALANG 'Can do' statements in Greek and asked to self-assess their level. These are the same statements that appear in the optional, self-assessment introduction to the reading, listening and writing components of DIALANG. The teacher also assessed her students' levels (research question 3) using the DIALANG scales (Council of Europe, 2001, pp. 231-237). Apart from obtaining the students' scores on DIALANG to address the first research question, the study also involved a questionnaire to address the second research question. Through the questionnaire the participants could report on their experience of DIALANG and other types of assessment and feedback, which could shed light on their thoughts and feelings about the experience.

Presentation of research findings

Analysis of the data

SPSS (version 17.0) was used to analyze frequencies of the students' answers to the questionnaire. The next step was to compare the results of the three assessments: (a) DIALANG test results in listening, reading, writing, vocabulary and structures, (b) teacher assessment in three language skills (listening, reading, writing), and, finally, (c) students' self-assessment of the same three skills. To check the strength of the association between the assessments, *Pearson's correlation* coefficient (r) was also calculated (Norusis, 1984).

DIALANG as a diagnostic assessment tool

As shown in Table 1, DIALANG placed most of the students at levels A1 and A2 for listening, writing and reading. The majority of the students (13) were placed at A1 and A2 as regards their vocabulary. Their results are different in the grammar component, where 10 students were placed at A1 and A2 and 10 at B1 and B2.

The comparison of the students' self-assessment (Table 2) and the teacher's assessment (Table 3) with DIALANG results (Table 1) shows that there is agreement among the three in the assessment of the listening skill. As far as reading is concerned, correlations in Table 4 show that the teacher agrees with the DIALANG assessment more than the students, whereas students agree more with the DIALANG when writing is assessed. Correlations for listening were very similar.

	Levels			
DIALANG test	A1	A2	B1	B2
1. Listening	12	4	3	1
2. Writing	8	10	1	1
3. Reading	8	7	4	1
4. Vocabulary	6	7	6	1
5. Structures	6	4	6	4

Table 1. Frequency of students across levels according to DIALANG test results

	Levels			
DIALANG test	A1	A2	B1	B2
1. Listening	6	6	8	-
2. Writing	1	10	4	5
3. Reading	-	9	10	1

Table 2. Frequency of students across levels according to students' self -assessment

	Levels			
DIALANG test	A1	A2	B1	B2
1. Listening	5	8	7	-
2. Writing	7	7	6	-
3. Reading	4	8	7	1

Table 3. Frequency of students across levels according to teacher assessment

	DIALANG test assessment		
	Listening	Writing	Reading
1. Teacher assessment	0.55*	0.38	0.58**
2. Students' self -assessment	0.51*	0.59**	0.31

*= p<0.05; **= p<0.01.

Table 4. Correlation of DIALANG assessment with teacher's and students' assessment

The next section reports on the questionnaire, where students had the chance to report on their experience of assessment and feedback in general and their experience of DIALANG. They were also able to express their agreement or disagreement with the DIALANG assessment and present the grounds on which their comparisons are based.

The questionnaire

General assessment/tests experience

The first part of the questionnaire revealed that students take non-diagnostic tests more often than diagnostic ones (Table 5).

	Never	Rarely	Sometimes	Often	Very Often
Placement tests at school.	1	17	-	1	1
Placement tests for evening classes.	10	6	-	2	-
Progress tests at school.	-	1	5	8	6
Progress tests for evening classes	1	2	3	4	8
Achievement tests at school.	1	4	8	5	2
Achievement tests for evening classes.	6	3	5	3	1
Diagnostic tests at school.	13	3	1	3	
Diagnostic tests for evening classes	10	3	3	1	1
Language exams (Cambridge/Michigan/ KPG etc).	14	2	1	-	3
Other (what):	-	-	-	-	-

Table 5. Tests students take

However, students acknowledge that the main function of tests is for their teachers to spot their weaknesses (diagnosis!) and help them by providing remedial instruction (Table 6).

	Not at all	A little	Quite	A lot	Very much
To give you marks.	1	2	7	9	1
To check your progress.	1	-	6	8	5
To spot your weaknesses and help you.	1	-	1	12	6
They have to.	-	8	7	4	1
To report to parents.	1	5	9	3	2
Other (what):	-	-	-	-	-

Table 6. Usefulness of tests for teachers

They also admit that tests can be a way to spot their strengths and weaknesses themselves (Table 7).

	Not at all	A little	Quite	A lot	Very much
They are a reason to revise.	1	3	10	3	3
I try to apply new knowledge.	1	2	12	4	1
They help me spot my strengths and weaknesses.	1	2	2	12	3
The mark is important.	-	4	6	4	6
None. They are useless.	1	-	1	3	15
Other (what):	-	-	-	-	-

Table 7. Usefulness of tests for students

The affective factor that most affects their performance in tests is anxiety caused mainly by the time limit and the level of difficulty of tests (Table 8).

	Never	Rarely	Sometimes	Often	Very Often
The time.	-	5	5	4	6
The level of difficulty.	-	-	6	11	3
I'm afraid I won't do well.	1	1	5	6	7
Other (what):	-	-	-	-	-

Table 8. Sources of anxiety during a test

The type of feedback students value the most is the mark they get as well as the indication of the mistakes they have made (Table 9).

	Not at all	*A little*	*Quite*	*A lot*	*Very much*
A mark.	-	3	4	5	8
A mark and a comment (Excellent, Very good etc).	-	3	6	8	3
A mark and indication of your mistakes.	-	1	9	7	3
Comments on your strengths.	2	1	6	6	5
Advice on how you can work on your weaknesses.	2	1	5	6	6
Other (what):	-	-	-	-	-

Table 9. Students' opinion regarding feedback

DIALANG test experience

The second part of the questionnaire showed that the majority of the students liked the computer-based DIALANG test more than traditional paper-and-pencil tests: 12 students answered positively and 8 gave a negative answer. However, they had difficulty coping with the background noise in the classroom where they took the test (Table 10).

	Not at all	*A little*	*Quite*	*A lot*	*Very much*
Time.	5	7	3	4	1
Classroom noise.	2	4	2	6	6
Typing the answers.	8	7	3	2	-
Navigating round the test.	6	6	3	4	1
Other (what):	-	-	-	-	-

Table 10. Problems students had during the test

Only a few made use of the feedback options they were offered by DIALANG: 5 students turned the 'Immediate Feedback' option on, and 8 reviewed their answers, making use of the 'Item Review' option. This is not surprising, if we take into account the fact that they were more interested in the final score, as it was shown in the previous section. Students also answered that the test was 'Too long' (4 students) or 'Long' (11 students). Only 5 students thought that the length of the test was moderate. Students reported that the listening and writing components of the test were the most difficult and considered the grammar component to be the easiest (Table 11).

	Too difficult	*Difficult*	*OK*	*Easy*	*Very easy*
Listening	1	4	13	1	1
Reading	-	4	11	4	1
Writing	1	7	8	3	1
Vocabulary	-	4	10	5	1
Structures	-	4	5	8	3

Table 11. Students' opinion regarding difficulty

Students did not agree very much with their assessment by DIALANG except for their grammar skill, where they achieved the highest score (Table 12).

	Not at all	A little	Quite	A lot	Very much
Listening	5	4	6	4	1
Reading	4	3	6	6	1
Writing	4	6	3	5	2
Vocabulary	3	3	8	5	1
Structures	-	3	7	8	2

Table 12. Students' agreement with DIALANG levels

Students compared DIALANG results with the level of the evening classes they attended, the level of the textbooks used there and the exams they were preparing for (Table 13).

	Not at all	A little	Quite	A lot	Very much
The level of the class I attend at school.	3	9	5	1	2
The level of the books we use at school.	4	10	4	1	1
The level of the evening class I attend.	-	4	5	6	3
The level of the books we use in the evening classes.	1	2	8	3	4
The level of the language exam I'm preparing for.	1	4	4	7	4
Other (what):	-	-	-	-	-

Table 13. Sources of comparison with DIALANG test levels

Finally, students reported that evening classes is where they practice their skills most and that those skills that are tested in high-stakes exams are practiced more than others (Tables 14, 15, 16).

	Never	Rarely	Sometimes	Often	Very Often
Intensive practice at school.	1	6	9	4	-
Intensive practice in evening classes.	-	-	5	6	7
It is one of the skills in language exams.	1	3	3	10	3
I listen to music in English.	2	3	1	3	11
I watch films in English.	1	3	6	3	7
I watch videos on the Internet.	1	4	4	4	7
I have friends with whom I speak English.	8	3	4	4	1

Table 14. Listening practice

Self-assessment

The final part of the questionnaire revealed that students are not often engaged in self- or peer-assessment and their assessment is usually done by their teachers, tests and exams (Table 17).

Whenever students were engaged in self assessment, they preferred to think of real situations and wonder whether they could cope. They were also often able to spot their strengths and weaknesses. Less often they compared themselves with their classmates or

native speakers and even less often did they compare themselves with their teachers (Table 18).

	Never	Rarely	Sometimes	Often	Very Often
Intensive practice at school.	1	6	8	5	-
Intensive practice in evening classes.	-	1	5	3	9
It is one of the skills in language exams.	1	4	4	7	4
I read when I surf the Net.	4	2	9	3	2
I try to read books in English.	6	6	7	-	1
My favourite magazine is in English.	13	5	2	-	-
Other (what):	-	-	-	-	-

Table 15. Reading practice

	Never	Rarely	Sometimes	Often	Very Often
Intensive practice at school.	5	8	5	2	-
Intensive practice in evening classes.	-	-	4	7	7
It is one of the skills in language exams.	1	3	6	5	5
I keep a diary in English.	17	2	-	1	-
I send emails in English.	11	1	5	1	2
I chat in English.	8	3	6	1	2
Other (what):	-	-	-	-	-

Table 16. Writing practice

	Never	Rarely	Sometimes	Often	Very Often
My teachers.	-	-	1	8	11
Tests.	-	1	3	7	9
Exams.	-	3	4	7	6
My classmates.	4	8	7	1	-
Other (what):	-	-	-	-	-

Table 17. Agents of students' assessment

	Never	Rarely	Sometimes	Often	Very Often
I compare myself with my classmates.	4	3	8	3	2
I compare myself with my teachers.	12	4	2	1	1
I compare myself with what a native speaker would do.	5	5	5	4	1
I can spot my strengths and weaknesses.	-	2	6	7	5
I think of real situations and wonder if I could cope.	-	2	4	10	4
Other (what):	-	-	-	-	-

Table 18. Self-assessment practices used by students

Finally, students reported that they are not always aware of the criteria used for their assessment regardless of the agent (Table 19).

Never	Rarely	Sometimes	Often	Very Often
2	2	10	6	-

Table 19. Students' awareness of assessment criteria

Discussion and Implications

DIALANG as a placement tool

The first research question explores the use of DIALANG as a placement procedure and the issues of interest are its administration and its potential regarding practicality, utility and discrimination.

Administration

To be considered for placement purposes DIALANG must fulfill two conditions: (a) be available in the school computer lab, so that up to 15 students can take it at one time, and (b) last approximately a teaching hour, 45 – 60 minutes, as test takers cannot try some of its parts, stop, save their progress and come back another time for the rest.

Condition (a) can be fulfilled provided that the necessary software application is downloaded from the internet and installed onto the computers where the test is to be taken. Condition (b), however, is more intricate, as a shorter version of the test should be considered. In order to shorten the administration time and enhance DIALANG feasibility as a placement tool two steps need to be taken: (a) The introductory VSPT component, which is optional, should be compulsory if test takers are to skip the vocabulary skill, and (b) students do only three of the skills, i.e listening, reading, and writing.

According to DIALANG results, grammar is the skill where the students achieved their highest scores. Vocabulary, on the other hand, is part of the initial placement procedures (VSPT) and even before the test begins the system already has an estimate of the test taker's vocabulary size. Moreover, the test takers' vocabulary range can be assessed in an integrated manner through their performance in the reading skill and their grammatical accuracy through the writing skill.

Practicality

DIALANG is a highly practical test due to the automaticity of test scores and feedback provision. The only extra material the teacher needs to prepare is the form on which students report their scores and feedback (Appendix). The form can yield invaluable information about students, especially new ones. For example, the students' accuracy of their self-assessment (i.e the extent to which they understand the DIALANG descriptors), their use of feedback (IF or Item Review or both) and whether feedback reflects what is true for them (Explanatory feedback). This information cannot be obtained through teacher-made placement tests, which, in most cases, provide just marks: the teacher(s) decide(s) the cut score and the students are streamed depending on whether their mark was higher or lower than the cut score. Therefore, DIALANG has much more to offer as a placement procedure than teacher-made placement tests.

Utility

The DIALANG system provides a plethora of information other than the test takers' level in the skills it assesses. Combined with the questionnaire that followed, it revealed a number of issues concerning the students' experience of testing and assessment and the types of feedback they value. It also yielded information on the sources of students' anxiety while they are taking a test and the factors that affect their performance. Finally, it provided insights into the students' perceptions of their language abilities and the sources these perceptions are based on.

The first finding regarding the students' experience of assessment is that students do not take diagnostic tests as often as they take other types of tests although they acknowledge that the reason why their teachers give them tests is to spot their weaknesses and apply remedial teaching.

What is quite surprising is the importance of the mark for the test takers as the type of feedback they value most. The result, the level in this case, was what mattered most and this was the reason why they treated the feedback options they were given by DIALANG the way they did: only 6 of them switched the IF option on, and only 8 reviewed the right and wrong items at the end of each test part. Students are accustomed to assessment practices which put them on numerical scales (1-20) without being aware of what a place on the scale reflects about their performance. This is indicative of the assessment practices their teachers use, which, in turn, is indicative of the teachers' inability to define what good performance is and what should be done to achieve it. But this is what DIALANG can do through the provision of advisory feedback.

The sources for the students' anxiety while they are taking a test are the time, the level of test difficulty and their fear that they may not do well. Students also report that having been well prepared for a test lessens their anxiety. The fact that there is no time limit for the completion of DIALANG and test level adaptivity are two benefits that DIALANG offers test takers, which can eliminate the anxiety factors that impinge on their test performance. In this sense DIALANG can be regarded as a test taker friendly test.

The fact that students performed better in the 'structures' component of DIALANG is not surprising as they have probably worked hard on their grammar accuracy from an early stage in language learning. The prominent role of grammar in the students' language learning practices is also evident in the relevant question in the questionnaire they answered. The perception of language as a concrete whole with grammar at its centre is also evident in the final part of the statistical analysis of the research findings, where an attempt was made to identify factors that can affect the students' performance in DIALANG. The students' attitude towards grammar can be interpreted in two ways: First, because of the emphasis placed on grammar by the teaching instruction they may have received, they tend to think that language is grammar. On the other hand, they may not be aware of the fact that language can be analyzed into skills, which are further analyzed into levels, at each of which they should be able to perform certain tasks. This would mean unfamiliarity with skills descriptors and the need for systematic training on them.

The students' performance in the listening part of DIALANG is the one that was most affected by the practical problems in the administration of the test. The fact that some of them had to tolerate considerable background noise while taking the listening test has made it more difficult for them to perform well. On the other hand, reporting difficulty in the

listening skill can be indicative of the lack of adequate practice. This is true, for at least those who do not attend evening language classes, since listening practice at school has been problematic in the last two years, because of deficiencies in the materials which accompany textbooks.

The sources on which students base their perceptions of their language abilities and how they usually practice their skills are also possibilities in the DIALANG explanatory feedback. In the questionnaire the students reported that their basic sources of comparison are the level of their evening classes, the textbooks they use in those classes and the language proficiency exams they are preparing for. The issues that arise from these findings are: (a) whether these sources are all linked to the CEFR, so that they can be compared to the DIALANG results, and (b) what placement procedures have been followed to place students in those levels. On the other hand, although the students do not seem to value them enough, the level of their English class at school and the textbooks used seem to be more suitable for their level, which is lower than they think, according to the DIALANG results. It turns out, therefore, that DIALANG is well suited to the Greek state school context because the CEFR is the link between them.

Discrimination

According to DIALANG results the majority of the students belong to levels A1 and A2 for most of the skills. However, this comes as a contradiction to the fact that they are C class students, who, according to the Pedagogic Institute, should have reached B1+ level by now. The textbook they are using also corresponds to B1 level. This raises questions as to the suitability of the teaching materials.

If DIALANG were to be used to place students into different classes, based on their language proficiency, schools should allow teachers flexibility in two areas. First, the maximum number of students in each class should be defined by the placement test results, not the need to have groups of equal numbers. Second, there should be a variety of teaching materials so that they are suitable for the students' proficiency level.

Self-assessment

As far as the second research question is concerned, the results show that the students have fairly overestimated their language abilities, probably because they view their language skills as concrete entities: it is not uncommon to hear a student say that they are good at grammar but not so good at listening, for example. Taking DIALANG was the first time they were asked to compare their language abilities against specific criteria which describe specific real-life language instances. However, students also reported that when they were engaged in self-assessment 'unofficially', they tried to estimate how they would cope in real-life situations (DIALANG explanatory feedback) and also spot their strengths and weaknesses (DIALANG advisory feedback). This is where DIALANG and its explanatory and advisory feedback can be invaluable: students can find out what over- or underestimating their language skills involves, and what they need to be able to do in order to reach a higher level.

Teacher assessment

As regards the third research question, the accuracy of teacher assessment, compared to DIALANG, is indicative of her experience of the students' skills and, consequently, her teaching practices. The teacher manages to assess her students' reading skills fairly

accurately, is closer to DIALANG for the listening skill, and misjudges her students' writing skills.

What is common in both the teacher's assessment and the students' self-assessment is the use of CEFR descriptors. If there was absolute agreement between the DIALANG assessment and the other two assessments, that would mean that CEFR descriptors are clearly interpretable and all assessors are familiar with them. But this is not the case.

Weir (2005) argues that there are transparency problems in using the CEFR regarding the wording of the descriptors, which may be worse when translation into other languages is involved. Papageorgiou (2009) reports that expert judges, who tried to relate two Trinity College London examinations to the CEFR, faced problems with, among others, the wording of the CEFR descriptors and the unsuitability of the CEFR scales for judging young learners. In the DIALANG case, however, which is a test already linked to the CEFR, the aforementioned shortcomings have already been taken care of. From the CEFR descriptors the most concrete, clear and simple were chosen, changed from 'Can do' statements into 'I can' statements to account for self-assessment use, and translated by two or three translation experts per language. Therefore, despite the inherent problems of the CEFR descriptors, the DIALANG statements should be regarded as clearly interpretable.

This would mean that the users are not familiar with the DIALANG statements and, consequently, the CEFR scales. It was the first time the particular students had ever thought of their language abilities in terms of concrete criteria, which reflected real-life language use. This is indicative not only of their inexperience of and unfamiliarity with the criteria, but also of the types of tasks they are used to performing, which are far from authentic. The students themselves acknowledge their inexperience of self-assessment and limited awareness of the criteria used by other agents (teacher, tests, exams) for their assessment in the final part of the questionnaire. The teacher, on the other hand, who is supposed to have a better grasp of the criteria, fails to agree with the DIALANG assessment of her students' writing skill. This means that either the criteria are not clear to the teacher or she does not have enough evidence of her students' performance on the particular skill on which to base her judgments. The latter is not surprising, since writing is not tested at the end-of-year final achievement test, under exam specifications dictated by the Ministry of Education, thus quite often writing instruction and practice are neglected.

Limitations of the study and suggestions for further research

The major limitation of the present study was the size of the sample (20 students, 1 teacher). It is not enough to draw generalizations but can be considered to be indicative of a number of issues such as the need for valid and reliable diagnostic tools, the need to train students to use the CEFR descriptors when they assess themselves and teachers when they assess their students, and the need for materials that are suitable for the students' levels, after reliable placement, and reflect real-life language use. The technical and practical problems with the administration of DIALANG constitute another limitation. Nevertheless, the study managed to show that DIALANG can serve as a reliable placement tool on certain conditions.

It would be interesting to check the findings of this study against a larger number of high school students in Greece, who would take DIALANG and deliver concurrent or retrospective verbal reports, so that no information goes amiss. This could be conclusive as to whether DIALANG is suitable for this particular age group of test takers and whether self-assessment

deficiencies are widely present in the Greek high school context. More research on the perceptions of teachers of the DIALANG 'I can' statements and the CEFR 'Can do' statements would also be revealing of the Greek teachers' of English awareness of and familiarity with the CEFR scales.

Author's email: vas_bag@hotmail.com

References

Alderson, J. C. (2005). *Diagnosing foreign language proficiency*: *the interface between learning and assessment*. New York: Continuum.

Bassey, M. (1981). 'Pedagogic research; on the relative merits of the search for generalization and study of single events.' *Oxford Review of Education*, 7/1: 73-93.

Bell, J. (2005). *Doing your research project*. Maidenhead: Open University Press.

Cohen, L., Manion, L. & Morrison, K. (2007). *Research methods in education (6th ed)*. London: Routledge.

Council of Europe (2001). *Common European Framework of Reference for languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.

Figueras, N. & Noijons, J. (2009). *Linking to the CEFR levels: research perspectives*. Arnhem: Cito, EALTA.

Floropoulou, C. (2002). 'Foreign Language learners' attitudes to self-assessment and DIALANG: a comparison between Greek and Chinese learners of English'. Unpublished MA dissertation, Lancaster University.

Haahr, J., Shapiro, H., Sorensen, S., Stasz, C., Frinking, E., van't Hof, C., Green, F., Mayhew, K., & Fernandez, R. (2004). 'Defining a strategy for the direct assessment of skills: final report'. Danish Technological Institute, accessed at http://ec.europa.eu/education/pdf/doc288_en.pdf on 14 Apr 2011.

Hughes, A (2003). *Testing for language teachers (2nd ed)*. Cambridge: Cambridge University Press.

Huhta, A., Luoma, S., Oscarson, M., Sajavaara, K., Takala, S. & Teasdale, A. (2002). 'A diagnostic language assessment system for adult learners.' In C. Alderson (Ed). *Common European Framework of Reference for languages: learning, teaching, assessment. case studies*. Strasbourg: Council of Europe, 130-145.

Jang, E. (2009). 'Demystifying a Q-matrix for making diagnostic inferences about l2 reading skills.' *Language Assessment Quarterly*, 6/3: 210-238.

Kaftandjieva, F. (2010). *Methods for setting cut scores in criterion-referenced achievement tests*. Arnhem: EALTA.

Meara, P. (1992). 'EFL vocabulary tests', accessed at http://www.lognostics.co.uk/vlibrary/meara1992z.pdf on 7 Jan 2012.

Ministry of Education (1996). 'Teaching English in levels. Circular C2/4230/12.7.1996', accessed at http://www.fa3.gr/nomothesia_2/nomoth_education/6_xenes_glosses_epipeda.htm on 7 Jan 2012.

Ministry of Education (2001). 'DEPPS: a cross-thematic curriculum framework', accessed at http://www.pi-schools.gr/programs/depps/ and http://www.pi-schools.gr/lessons/english/aps_agg.php on 16 Aug 2011.

Norusis, M. J. (1984). *SPSS/PC for the IMBPC/XT*. Chicago-Illinois: SPSS Inc.

Papageorgiou, S. (2009). 'Analyzing the decision-making process of standard setting participants.' In N. Figueras & J. Noijons (Eds.), *Linking to the CEFR levels: research perspectives*. Arnhem: Cito, EALTA, 75-79.

Weir, C. (2005). 'Limitations of the Common European Framework for developing comparable examinations and tests.' *Language Testing*, 22/3: 281-300.

West, R. (2004). 'Selection and placement testing.' In C. Tsagari & R. West, *Testing and assessment in language learning (vol. 3)*: *assessing students without tests*. Patra: Hellenic Open University, 343-410.

Yang, R. (2003). 'Investigating how test-takers use the DIALANG feedback.' Unpublished MA dissertation, Lancaster University.

Zachos, D. (2009). 'Citizenship, ethnicity, and education in modern Greece.' *Journal of Modern Greek Studies,* 27/1: 131-155.

Appendix

DIALANG test scores and feedback

Name: _____ **Class:** _____
Date: _____

VSPT score: _____

LISTENING		
I/F	ON	OFF
LEVEL:		
Item review:	Yes	No
Self-assessment and test score	Agrees	Disagrees
Advisory feedback		
Explanatory feedback		

READING		
I/F	ON	OFF
LEVEL:		
Item review:	Yes	No
Self-assessment and test score	Agrees	Disagrees
Advisory feedback		
Explanatory feedback		

WRITING		
I/F	ON	OFF
LEVEL:		
Item review:	Yes	No
Self-assessment and test score	Agrees	Disagrees
Advisory feedback		
Explanatory feedback		

Book review

Learning the Arts of Linguistic Survival: Languaging, Tourism, Life
Alison Phipps (2007). Clevedon: Channel View Publications, 205 pp., £19.95 (paperback), ISBN 1-84541-053-X

The number of people learning languages for tourist reasons in the West is a most significant one and yet literature has ignored this student group, focusing instead on language learning either in a functionalist manner or at an academic level. Moreover, neither literature on linguistics nor on tourism has addressed how aspects of their fields are linked to social relationships. The above gap is being filled by the latest book of Alison Phipps, whose main purpose is to address the relationship between tourism and languages through exploring tourist language learning.

Phipps brings a fresh look at language learning and language interaction through tourism by encouraging us to pay attention to what she defines as the 'quick' of human relationship, that is 'anything characterised by the presence of life'. Tourist language learners or 'languagers' – as Phipps has termed people who interact with the world through putting a language they are learning into action – step outside their habitual ways of speaking, letting go of their linguistic power. To do that signifies an important social phenomenon and it is the aim of this book to explore why people bother to invest time and energy in learning to speak a tourist language and what happens when people attempt to shift their language norms, not as a necessity, but as part of their tourist experience.

The distinctiveness of this book is that it emphasises the everydayness of human action and human relationships, overcoming the separation between traditional liberal education and practical knowledge, between mind and body and between the act of learning and the act of using the acquired knowledge. Most importantly for educationalists and linguists, it challenges the established hierarchies of what counts as worthwhile learning/teaching by showing how apparently simple linguistic abilities and practices, like ordering a coffee in another language, are more than functional competencies. They can be practices through which our perception of the Other and ourselves are transformed.

Phipps, coming from a humanist approach and drawing on a wide range of disciplines – such as anthropology, postcolonial theory, linguistics and cultural studies – perceives language learning as a way of dwelling in another world. She, thus, redirects our attention to small level – and yet very common, significant and so far ignored – interaction with one another and with the Other. This book looks at different aspects of a tourist language class, such as practising oral speech or using games, but instead of focusing on the linguistic elements of these practices, it explores their social and psychological implications. In other words, Phipps

does not seek to establish a language teaching framework, but to enter the social world of the learning experience.

For this reason, the author does not write as a pedagogue, but as a tourist language learner. This is linked to the methodology of the study undertaken and it is reflected in the writing style of the book. Taking a phenomenological perspective, she engages in an empirical ethnographic study of tourist languages classrooms through acting herself as participant observer in a variety of courses. She uses a reflective journal with field notes in the form of language learning notes or reactions to the educational experience. These narratives and images are inserted into the text, giving life to the book. In this manner, not only is the book is written from a tourist language learner's point of view, but it is also read in this way. Readers are, therefore, rethinking the experience of learning a language and putting it into action, which is immensely important for a language educationalist.

Although links can be made, the book does not repeat what would apply to any language class. Language learning for tourism entails a quest for recreation and for a time that differs from habituated daily action. The learning involves touristic imaginings as memories or hopes for future encounters. Accordingly, learners bring with them in the class different needs, aims and yearnings. Imagining and enjoying comes together with struggling as part of the educational experience. Pleasure is accompanied by dealing with risk and transforming oneself when putting the acquired language skills in practice. The above feelings are explored in the book in the seeking to answer the question 'why bother' to learn a language for tourist purposes when the gains appear so futile and temporary.

Through examining lessons on way finding and pronunciation and by looking into language learning activities, such as games and oral practice, Phipps illustrates the uniqueness of tourist language learning curriculum and classroom interaction. As the reader becomes absorbed in this world, it is revealed how language classes, which remain in the margin of the educational system, encourage a heightened awareness of place and environment. They develop the ability to transform a travel destination into an inhabited place and enable learners to make meaning and relate to the unfamiliar. Phipps contends that learning to converse in the destination's language represents a charitable act of stepping out of one's own comfort zone.

This act is non-comprehensible and unnoticed by the functionalist approach that dominates language learning policies and literature. And yet this oversight allows tourist language classes to develop in a different manner involving a sense of play – although a play where the stakes can be too high – both in the classroom and when the learned language is put into action. In this way, the class acts as a rehearsal for being a tourist. It gives a feel for using the language and enables play and social bonding to flourish both in the classroom and in the tourist destination. Accordingly, Phipps argues that, against the prevailing insulation of the West, this ignored student group, by breaking with the dominant tongue and its perceived power, shows a willingness to open up themselves in ways that impact on their self-perception and relating to the world. Unwittingly political, switching from English is a courtesy that attempts to add to simple interaction a sense of social relatedness and commitment.

In conclusion, this is a very engaging and stimulating book, full of dense meaning and not for someone who wants an easy read. It generates an excitement as so many new ideas are explored in a way that the reader can refer to personally – as a language educationalist, as a language learner and even as a tourist. Most important of all, this book exerts a feeling of hopefulness in a time where functionalism on the one hand and postmodernism on the

other have denied us such feelings. This is an optimism that derives from people who are not content to operate according to the dominant modes suggested by tourism, but who choose to dwell in different worlds and engage in vertical travel through which relationships with places, people and life are deepened. It is to these people and this phenomenon that this book draws our attention to. Through exploring tourist language learning, Alison Phipps manages to reshape our worldview too.

Ira Papageorgiou
Hellenic Open University
ira_papageorgiou@tutors.eap.gr

Authors' biographical information

Vassiliki BAGLANTZI has been a state high school teacher of English since 2000. She holds a B.A in English Studies from the University of Athens and an M.Ed in TESOL form the Hellenic Open University. She has conducted research in the area of diagnostic assessment and her field of interest is assessment practices in the state school context.

Melpomeni BARABOUTI has been an EFL teacher in State Primary Education since 1993 and deputy head teacher at the school she has been appointed since September 2011. She holds a Master's Degree in TEFL from the Hellenic Open University. She is particularly interested in the improvement of the quality of education in the Greek State School. She participates as an examiner of oral speech in the State Language Examinations (KPG), as well as in the student mentoring program run by the University of Athens. She was the coordinator of a Comenius program of European school partnerships (2004-2007). Since 2010 she has been the co-designer of the material used for the teaching of English to the young learners of B class in the PEAP project undertaken by the Faculty of English Studies of the University of Athens.

Irene BOMPOLOU holds a B.A. in English Language and Literature from the University of Athens and a M.Ed. in Teaching English as a Foreign Language from the Hellenic Open University. She works as a teacher of English in secondary education. She has been an active P.E.K.A.D.E. Board Member for six years. She has also been an oral examiner and a script rater for the K.P.G. examinations since 2003. Her research interests concern the methodology of teaching languages and include assessment, ICT and English Language Teaching, task-based language learning and CLIL.

Ekaterini CHALKIA was born in Arta, Greece. She is married and she has got two sons. She studied English Language and Philology at the Aristotle University of Thessaloniki and graduated in 1991. Since then she has been teaching English in the 7th Primary School of Arta. She has also worked as an English teacher in IEK (Institute of Vocational Training), KEK (Vocational Training Centre) and the State Nursing School of Arta. She holds an MA in Postgraduate Specialisation of Teachers of the English Language from HOU.

Angeliki DAPHNI holds a degree in English Language from the University of Athens and a Med in TESOL from the Hellenic Open University. She has worked as an EFL teacher in IEK, private language centres and schools as well as in state schools of all levels in cycladic islands, north Euboea and Athens. She has been teaching in Dionysus senior high school since 2006.

Eleni DASKALOGIANNAKI graduated from the Faculty of English Studies at the University of Athens. She holds an MA in Teaching English as a Foreign Language from the Hellenic Open University in Patras. Her research interests concern the effectiveness of weblogs in the English writing classroom. She has been teaching English for about eight years in Secondary Education in Greece.

Xenia DELIEZA holds a B.A. from the Faculty of English Studies, National and Kapodistrian University of Athens, and an M.A. degree in Media Technology for TEFL from the University of Newcastle upon Tyne, U.K. Since 1999 she has been an EFL teacher in secondary education in Greece. She has taught both in the private and the public sector and since 2003 she has been a state EFL teacher in secondary education in Greece. For four years she was seconded at the Faculty of English Studies to be part of the RCEL team, which carries out research and prepares the English exams for the KPG, of which she is still a member. Presently, she is also doing her PhD on interlocutor discourse in KPG oral exams.

Georgia EFTHYMIOU holds an MA in TESOL from the Hellenic Open University of Patras. She has been teaching English for about 15 years. For the last 11 years she is working as an EFL teacher in a Greek state primary school. She is also an examiner for the oral part of the National Certificate of Languages (KPG) exams since 2005. She is interested in the methodology of teaching and testing young learners and in Comenius programmes, in many of which she has participated herself.

Jenny LIONTOU is a PhD student at the Faculty of English Studies, University of Athens. She holds a B.A. in English Language & Literature and an M.A. in Lexicography: Theory and Applications, both from the same faculty. She also holds an M.Sc. in Information Technology in Education from Reading University, UK. She has worked as an expert Item Consultant for AvantAssessment, USA, a research assistant at the Research Centre for English Language Teaching, Learning and Assessment (RCEL) and as an oral examiner and script rater for various examination boards. She has made presentations in national and international conferences and has published papers in the aforementioned fields. Her current research interests include theoretical and practical issues of EFL reading comprehension.

Efimia KARAYIANNI holds a B.A in English Literature from the University of Athens and an M.ed in TESOL from the Hellenic Open University (HOU) in Patras. She has been teaching English for about 15 years and she is currently teaching English in the 4th State Gymnasium of Arta, Epirus. Her research interests concern assessment of the English language in the state schools as well as Educationa Technology and its implementation in the EFL classroom.

Sophia KOUZOULI is a teacher of English at the 1st Primary School of Pyrgos. She holds a B.A. in English Language and Literature and an M.Ed. in Teaching English as a Foreign Language from the H.O.U., Patra. She has been teaching English for more than 20 years. Her fields of interest include language teaching methodology, learning theories, educational assessment and diversity in the classroom. She has recently joined the "English for Young Learners" project conducted by the RCEL research team of the Faculty of English Studies at the University of Athens as an evaluator of the teaching materials developed for the English language teaching programme to first and second grade pupils.

Elena MELETIADOU is currently a PhD candidate in Applied Linguistics at the Department of English Studies, University of Cyprus. She holds a MEd in TESOL from the Hellenic Open University and an M.A. in Theoretical and Applied Linguistics from the University of Cyprus. She has been teaching EFL, ESL, GFL (Greek as a foreign language) and FFL (French as a foreign language) for almost twenty years in a variety of contexts in France, Cyprus and Greece. Her research interests include classroom-based language assessment, alternative assessment, collaborative language learning and teacher training.

Eleftheria NTELIOU works as an EAP/ESP Instructor in the University of Thessaly. She holds a B.A. in English Language and Literature from Aristotle University, Thessaloniki and a Master's Degree in Applied Linguistics from the University of Reading, UK. Her current PhD research focuses on Language Testing and Assessment, with a special emphasis on the interaction between task design and performance in the KPG oral exams. She has been an oral examiner and examiner trainer in the KPG exam system since 2005. Her research interests also include methods of assessing the teaching and learning of academic English as well as material design.

Sophia PAPAEFTHYMIOU-LYTRA is Professor of Applied Linguistics and Chairperson in the Faculty of English Studies, University of Athens. She is also Head of the M.Ed. in TESOL programme at the Hellenic Open University. Her publications include five books and numerous articles in journals and edited volumes dealing with issues in learning and teaching English as a foreign language and teacher education and training. Her current research interests include classroom discourse analysis and learning, learning and communicating strategies, linguistic and cultural awareness and learning, self-

access, autonomous and distance learning, the role of L1, assessment and certification, teacher education and training, adult foreign language education.

Vanda PAPAFILIPPOU studied English Language and Literature in the Aristotle University of Thessaloniki, Greece. In 2008 she completed her MA Educational Studies in the University of Warwick, UK where she specialised in special educational needs and assessment. Currently, she is about to finish her ESRC-funded PhD in the Graduate School of Education, University of Bristol, UK. Her narrative-inquiry-based PhD explores how the 'Greek candidate' is constituted as a subject by society, ideology and discourses around and promoted by English language tests, as these operate in Greece. Her research interests include sociology of education, sociology of assessment, philosophy of education, test validity and epistemology. She has also worked as a TESOL teacher in Greece and as a teaching assistant in the Graduate School of Education, University of Bristol.

Spiros PAPAGEORGIOU is a Research Scientist at Educational Testing Service in Princeton, New Jersey, USA. He has also worked as a Language Assessment Specialist at the University of Michigan. His research interests include mapping language test scores to the Common European Framework of Reference (CEFR) levels and listening assessment, and he has published a number of journal articles and book chapters on these topics. Spiros has also been an active member of the language testing community, serving as member-at-large of the Executive Boards of the International Language Testing Association (ILTA) and the Midwest Association of Language Testers (MwALT) in the USA. He was the recipient of the Robert Lado Memorial Award in 2007 and the Jacqueline Ross TOEFL Dissertation Award in 2009.

Sauli TAKALA is Professor (emeritus) at the Centre for Applied Language Studies, University of Jyväskylä, Finland. He obtained his PhD focusing on vocabulary learning in EFL in 1984 at the University of Illinois at Urbana-Champaign. He took an active part for over 40 years in the national research and development work on language education in Finland and participated in Nordic cooperation in this area. He coordinated the IEA International Study of Writing in the 1980s and helped to plan and coordinate the EU DIALANG project in the late 1990s. He has had a long association with the Council of Europe modern language project and is currently consultant for its European Centre for Modern Languages. He is a founding member of EALTA and served as its second president.

Dina TSAGARI is a Lecturer in Applied Linguistics/TEFL in the Department of English Studies, University of Cyprus. Prior to that she was a tutor at the M.Ed. in TESOL programme at the Hellenic Open University. Her research interests and publications relate to language testing and assessment, language teaching and learning, teacher education, adult and distance learning, language course and materials design. She has participated in a number of research projects in Greece, Cyprus and Hong Kong and is a coordinator of the EALTA-CBLA SIG group.

Stavroula VLANTI holds a BA in English Language and Literature from the University of Athens and a Master's Degree in Education for Teachers of English to Speakers of Other Languages from the Hellenic Open University. She has worked as an English language teacher in primary and secondary education. For the last nine years, she has been teaching in the 9th Junior High School of Aigaleo.