MODERN GREEK LANGUAGE E-DIAGNOSTIC TESTS

Thomais Rousoulioti\textsuperscript{1}, Rania Voskaki\textsuperscript{2} and John N. Kazazis\textsuperscript{2}

\textsuperscript{1}Aristotle University of Thessaloniki, \\
\textsuperscript{2}Centre for the Greek Language

Abstract
The Centre for the Greek Language (CGL) has designed the Modern Greek Language e-Diagnostic tests (MOGEDs). MOGEDs are online testing applications, available for teachers of Modern Greek as a second or foreign language (L2). They are mainly addressed to adult potential candidates for CGL’s language exams, willing to assess their language competence level. MOGEDs are compliant with the standard levels (A1-C2) of the Common European Framework of Reference for Languages (CEFR) (Council of Europe 2001) as adapted for Modern Greek. In this paper, the structure of MOGEDs will be analysed and compared to equivalent e-diagnostic tests in terms of the technical architecture adopted. MOGEDs have been developed within the framework of educational technology, taking into account (a) the CGL’s technical expertise in that field in relation with (b) state-of-the-art content design principles and (c) current trends in Information and Communication Technology.

1. Introduction

Greek as a second or foreign language (L2) is learned and taught in many countries all over the world. The Centre for the Greek language (CGL) is the official state-recognized research centre which certifies the Modern Greek Language Attainment and awards the official “Certificate of Attainment in Modern Greek”. For the purposes of the CGL exams and the promotion of the Greek language in general, the CGL cooperates with about 200 exam centres all over the world. Within this framework, and in order to support all those who are interested in participating in the examinations for the Certificate of Attainment in Modern Greek, the CGL has designed a set of e-diagnostic tests.\textsuperscript{1}

Modern Greek e-Diagnostic tests (MOGEDs) are an online testing application, available for teachers of the Modern Greek language as L2, but mainly for adult potential candidates for the CGL exams looking to determine their level of language competence. Greek is one of the least widely spoken languages outside Greece. This
is confirmed by surveys carried out inside (e.g. Ανδρούλακης 2008) and outside Greece (European Commission 2006, Eurostat 2010). Using this online testing application, people from all over the world can find out what level of Greek they have achieved before they take the CGL exams in order to have their language knowledge and awareness certified.

The aim of this article is to present the framework adopted for the design of the aforementioned e-tests rather than serve as a model for the design of diagnostic tests. The paper is structured as follows: In section 2 we present basic assumptions about the role and function of language diagnostic tests, the definition and the purpose of diagnostic (2.1) and the basic principles governing the design of a diagnostic test (2.2). In Section 3 we explain the methodology adopted for the characteristics of the e-diagnostic tests to take on their final form (3.1) as well as the pilot testing and test validation (3.2). In section 4 we describe the layout and function of the e-diagnostic tests and, finally, in section 5 we discuss the limitations of this research project and conclude with suggestions for further improvement.

2. Literature review

In the Glossary of the Association of European Examiners (The Association of Language Testers in Europe, hereinafter ALTE), a diagnostic test is defined as “a test which is used for the purpose of discovering a learner’s specific strengths or weaknesses. The results may be used in making decisions on future training, learning or teaching.” (ALTE 1998: 142).

According to Alderson (2005: 4), who led the development of a well-known diagnostic test, DIALANG, a diagnostic test is the type of test that seems to be closer to the learning process than any other. This happens perhaps because a diagnostic test is designed on the basis of a specific content and field that is covered either by a language curriculum or by a particular language proficiency theory. Bachman (1990: 60) holds that diagnostic tests can be based either on a linguistic approach or on a syllabus.

A key feature of a diagnostic test is to provide feedback (Ypsilandis 2002) to the students who use it in order to transform the examination process into a learning process (Υψηλάνης 2009: 87). The “philosophy” and the overall ultimate goal of a diagnostic test is mainly pedagogical, learner-centered and supportive of learning, as
it provides reports on what the student should improve. Υψηλάνης (2009) also suggests the design of an online diagnostic adaptive test that, in addition to the diagnostic process, will provide information (strategies, techniques) on “how the learner will improve their learning” (ibid.: 87). Therefore, a diagnostic test may be purely diagnostic and advisory, but may also provide a mark or rating (from level A1 to level C2) depending on the objectives it serves.

Summarising how the results of a diagnostic test may be exploited, Davies et al. (1999) state that the information obtained from diagnostic tests can be used at the beginning of a syllabus to classify, select or even design the programme itself. Also, diagnostic tests could be used in addition to the initial preparation phase of a language training programme or at the end of a preparatory course playing the role of a placement test. As Alderson et al. (1995: 12) argue, a diagnostic test can provide information on those areas where students need further assistance. For this reason, diagnostic tests should be general-oriented and give us information on whether the learner, for example, needs particular help in one of the four basic language skills, or even more specific information, identifying their weaknesses, for example in the use of grammar. Due to the complex nature of language, it is rather difficult to construct language tests that identify the strengths and weaknesses of students. This is the reason why there are very few tests that really function as diagnostic tests. Each test, under certain conditions, could diagnose learners’ language abilities (Bachman 1990: 60) but there is one type of test that seems to be very close to the diagnostic, the placement test. Hence, as Alderson (2005: 6) states, there is considerable confusion in the literature between diagnostic and placement tests. Thus achievement and proficiency tests are used in the classroom quite often and systematically to perform a diagnostic purpose.

It would no doubt be very useful to administer exams that allow the identification of the learners’ strengths and weaknesses both by teachers as well as by the learners themselves. In particular, for high-stakes language tests, such as university entrance exams or long-term residence permit exams, the consequences of failing are critical and therefore the design of high quality diagnostic language tests is extremely important. However, the role of diagnostic tests in L2 teaching is rather neglected (Kunnan and Jang 2009: 610). Lee (2015: 295) argues that, in the future, the development of diagnostic tests has to take into consideration the following: a) where diagnosis is quite often needed, b) dynamic assessment, c) existing and future models
of cognitive diagnostic assessment, d) technological innovation in e-assessment and scoring and e) feedback research.

2.1 E-diagnostic tests: Definition and scope
Most researchers in e-assessment suggest that the whole process of e-assessment, from its design to its completion and provision of feedback should be implemented electronically (Alruwais et al. 2018: 34). Stödberg (2012) lists five categories of questions commonly used in e-assessment: a) close-ended questions, such as multiple-choice or matching questions, b) open-ended questions, c) portfolios, d) products, such as computer programmes and e) discussion between students. The same study concludes that close-ended questions are the most frequently used in e-assessment.

The scope of MOGEDs is pre-assessment of language receptive skills in order for the test takers to participate in the exams for the Certificate of Attainment in Modern Greek. It is therefore possible for anyone interested in these exams to initially identify their strengths and weaknesses in specific language skills and then proceed to apply for the most appropriate level of language exams for these skills. In addition, MOGEDs can be used by teachers and curriculum developers to diagnose and determine the level of language proficiency of the students more accurately and, based on the information they will collect, to better organize the syllabus. MOGEDs are freely available at http://www.greek-language.gr/certification/tests/index.html

2.2 Designing an e-diagnostic test
Designing an e-diagnostic test requires both deep knowledge of language assessment and technology use. For the design of MOGEDs, the framework of Suvorov and Hegelheimer (2013) has been adopted in order to define a computer-assisted language test as any test delivered via a computer or a mobile device.

This framework consists of nine attributes and their corresponding categories, as shown in Table 1.
Table 1: Framework for the description of computer-assisted language tests (Suvorov and Hegelheimer 2013)

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Directionality</td>
<td>Linear, adaptive, and semi-adaptive testing</td>
</tr>
<tr>
<td>2 Delivery format</td>
<td>Computer-based and Web-based testing</td>
</tr>
<tr>
<td>3 Media density</td>
<td>Single medium and multimedia</td>
</tr>
<tr>
<td>4 Target skill</td>
<td>Single language skill and integrated skills</td>
</tr>
<tr>
<td>5 Scoring mechanism</td>
<td>Human-based, exact answer matching, and analysis-based scoring</td>
</tr>
<tr>
<td>6 Stakes</td>
<td>Low stakes, medium stakes, and high stakes</td>
</tr>
<tr>
<td>7 Purpose</td>
<td>Curriculum-related (achievement, admission, diagnosis, placement, progress) and non-curriculum-related (proficiency and screening)</td>
</tr>
<tr>
<td>8 Response type</td>
<td>Selected response and constructed response</td>
</tr>
<tr>
<td>9 Task type</td>
<td>Selective (e.g., multiple choice), productive (e.g., short answer, cloze task, written and oral narratives), and (selective) interactive (e.g., matching, drag and drop)</td>
</tr>
</tbody>
</table>

According to the aforementioned framework, MOGEDs are computer-assisted language tests that adopt standard test theory and they are linear in terms of their directionality, namely, in MOGEDs the participants are administered all test items in the same order and are allowed to revisit previous items if they wish to change their answers before they submit the test.

Furthermore, language tests administered with the help of computers can be divided into computer-based tests (CBTs) and Web-based tests (WBTs). In terms of their delivery format, MOGEDs are Web-based tests, as an assessment of test takers’ performance in an online format. Ockey (2009), among other researchers, had
predicted that due to rapid technological advances, WBTs would be more popular and further developed in the future.

One of the advantages of regular computer-assisted language testing may be media density, namely, the possibility to integrate different media formats. However almost all e-diagnostic tests quite often use a single medium (e.g. an audio-only listening test or a text-based reading test) and less often multimedia (e.g. a listening test with a video or a reading test with text and images). Although many researchers believe that the use of multimedia can increase the degree of authenticity of the assessment process, others, like Douglas and Hegelheimer (2007) warn that the use of multimedia can threaten the validity of the computer-assisted language tests, as we may not be sure that we assess what we have to assess.

Concerning target skills, computerized language tests can assess either a single language skill (listening, reading, speaking or writing) or a set of integrated language skills, which may reveal the test takers’ overall ability in language knowledge and use and enhance the authenticity of the assessment process (Ockey 2009). According to Alderson (2005: 1), diagnostic tests are “more likely to be discrete-point than integrative”, because they focus on diagnosing linguistic competence on only one skill (listening, for instance), while proficiency tests examine the language proficiency of the individual as a whole. MOGEDs assess only receptive skills (listening or reading) and only one of these skills in each set of activities.

As regards the scoring mechanism, in MOGEDs test takers’ answers are matched with the predetermined correct answers and this is the reason that most test items are in the form of multiple-choice questions. The automatic scoring calculation available to test takers, as in the case of MOGEDs, frees up time for other activities for teachers (Fulcher 2013: 218) and gives the test taker the opportunity to know about their performance immediately.

When it comes to stakes, according to Roever (2001), computer-assisted language testing may have:

1. low stakes (little or no consequences) for test takers, when it is used mainly for practicing, self-studying and assessment,
2. medium stakes (has implications for classroom assessment of students),
3. high stakes (important consequences for test takers’ lives at the level of education, professional promotion or citizenship acquisition).
According to the characteristics mentioned above, it is obvious that MOGEDs are a set of low stake tests, since candidates take them in order to determine if they are well-prepared to take the exams for the Certificate of Attainment in Modern Greek and, more specifically, their possible level of attainment in this language.

Concerning the purpose of the test, Carr (2011: 6) classifies tests in two broad categories: curriculum-related and non-curriculum related tests. Some of the roles a curriculum-related test performs are admission to a programme, placement at a specific level, diagnosis of the test takers’ strengths and weaknesses, performance assessment and achievement of the objectives of the programme. MOGEDs are curriculum-based tests. They are used for the diagnosis of the test takers’ strengths and weaknesses and simultaneously they classify test takers into a particular language level which shows if they possess the receptive skills necessary for the respective level at the Certificate of Attainment in Modern Greek.

The response type of a computer-delivered test is divided into two major categories of answers: selected and constructed responses (Parshall, Davey and Pashley 2000). Selected responses involve tasks requiring the test taker to choose a correct answer from a list of options, just like the multiple-choice questions. In the case of constructed responses, test takers have to develop their own answers and to produce a short or an extensive text. For the sake of practicality and immediate feedback, MOGEDs are mostly based on close-ended questions.

There is a great number of types of tasks that can be designed for computerized language tests. Task types can be divided into three broad categories:

1. selective (multiple-choice questions, yes/no or matching questions),
2. productive (written and oral narratives, short answer tasks, and cloze tests), and
3. interactive (matching, drag and drop).

A computer-based test enables the design of tasks suitable only for computer. Alderson (2005), for example, describes 18 experimental e-items that were designed for the DIALANG project (https://dialangweb.lancaster.ac.uk/). DIALANG is a low stakes computer-based diagnostic test which is available in 14 European languages, including Modern Greek. MOGEDs consist only of a selective task type in order to as much as possible simulate the paper placement test (concerning receptive skills only) that the stakeholders are required to take if they wish to enrol in a Modern Greek
course. Also the selective task type employed in MOGEDs allows their completion relatively quickly in the preset time period.

It is quite normal for the above described attributes to interact with each other. For example, stakes interact with the selection of scoring and delivery format. High stakes tests prefer the CBT (Computer-Based Testing) format and combine automated and manual scoring, while low stakes tests usually adopt WBT (Web-Based Testing) format and opt for automated scoring. MOGEDs are consistent with the latter.

3. Methodology

3.1 MOGEDs: Design features

Before designing MOGEDs, a thorough online research was conducted during which language tests for English, French, German, Spanish, Italian and Polish were studied. MOGEDs follow the classification of the levels of the Certificate of Attainment in the Modern Greek Language from A1 to C2 (http://www.greek-language.gr/certification/node/119.html) which is based on CEFR (Council of Europe 2001).

In a way, MOGEDs constitute a test equivalence (Αντωνουπούλου, Βεντούρης and Τζοπάνογλου 2015) of the exams for the Certificate of Attainment in Modern Greek concerning the receptive language skills. At this trial stage of their implementation, MOGEDs consist of 24 tasks, 4 for each language level from A1 to C2. There are six separate sets of tests for certification levels from A1 to C2 respectively on the diagnosis of the level of test takers’ receptive skills. The diagnosis of test takers’ performance in receptive skills includes oral and written comprehension. MOGEDs also include tasks that examine the use of language from level B2 to C2, otherwise known as test takers’ language awareness.

The texts that have been chosen for reading comprehension are mostly authentic. According to Morrow (1977: 13) “An authentic text is a stretch of real language, produced by a real speaker or writer for real audience and designed to convey a real message of some sort”. However, sometimes, particularly in low language levels (A1-B1), texts are semi-authentic, simplified to fit the needs of lower-level learners but also in this case they come from authentic language sources. Likewise, texts for listening comprehension are either authentic for high language levels (B2-C2), or semi-authentic texts that have been elaborated on in terms of the language level and
have been recorded in studios with the help of actors for low language levels. McNamara (2000: 131) defines authenticity as “the degree to which test materials and test conditions succeed in replicating those in the target situation”. It is quite clear that the degree of authenticity differs from one language level to another and from one situation to another, graded accordingly.

The online environment allows the completion of each language task within a specified time and gives immediate feedback to the user. It therefore provides exactly the information that will be useful to the user to identify their weaknesses and support their learning procedure.

3.2 MOGEDs: pilot study/validation

In order to choose and check the kind of tasks for MOGEDs, a pilot study was conducted. Sixty students of Modern Greek as L2 from every language level took part in the pilot study. The pilot study took place at the School of Modern Greek Language, Thessaloniki, Greece, the “St. Kliment Ohridski” University, Department for Language Teaching and International Students (DLTIS), Sofia, Bulgaria and the Greek Community of Rome and Lazio (Comunitá Ellenica di Roma e Lazio), Rome, Italy.

Participants in the pilot study of the e-diagnostic test successfully completed it by 68%. The performance of the participants in the pilot study was taken into account in the final form of MOGEDs. Therefore, changes were made to diagnostic tasks, some of which had to be replaced, or to specific diagnostic questions that were either unclear or inappropriate for the particular language level. In addition, the performance of test takers in the exams of CGL for the Certificate of Attainment in Modern Greek was also taken into consideration, with emphasis on tasks that seem to have the greatest degree of difficulty. Test takers’ performance is important because in reality “we cannot see or measure language ability at all, we only observe and measure performance, and on the basis of the performance of our test takers make inferences about their language ability” (Douglas 2010: 9-10).

Cronbach’s alpha was used to measure the internal reliability of MOGEDs and the measurement yielded fairly reliable values, ranging from 0.8 to 0.85 for each set of items per test. External validity did not apply, because test takers had access to feedback and could improve their performance each time they retook the test. Messik (1989) refers to validity as an “evaluative judgment of the degree to which empirical
evidence and theoretical rationales support the adequacy and appropriateness of intercorporations and actions based on test scores or other modes of assessment”. On this basis, scientific interest shifts from the measurement tool itself to its use. Regarding criterion validity in MOGEDs, students completed the diagnostic test that was in accordance with the language level at which they attended classes.

Construct validity is the degree to which a test measures what it claims. In MOGEDs receptive skills are mostly examined with items in which potential candidates have to choose the correct answer among others according to the test structure in the exams of CGL for the Certificate of Attainment in Modern Greek. In this case, construct validity exists given the limitations set by the technical implementation of an e-test.

Content validity (topics, length of texts, items and images) (Πετρίδου and Καραγιώργη 2017: 92) was checked with the help of 11 test developers for the exams of the Certificate of Attainment in Modern Greek regarding receptive skills. Scoring validity was not an issue since MOGEDs do not examine productive skills. As previously mentioned, all tasks related to the examination of receptive skills are based on the method of choosing the correct answer among words, phrases, sentences or texts, while grading is based on an automatic scoring system.

4. MOGEDs layout and function

The MOGEDs homepage ensures a user-friendly navigation. The levels from A1 to C2 appear on the menu on the top left and in the centre of the screen (fig. 1).

*Figure 1: Excerpt of MOGEDs’ homepage*
Test takers can choose the level and the skill (reading, listening comprehension or use of language), in which they want to be assessed by clicking on the corresponding field. They can repeat the e-diagnostic tests as many times as they wish, within time limits which have also been set according to the duration of the exams for the Certificate of Attainment in Modern Greek. At the bottom left of the screen, a timer is displayed reminding the test taker of the remaining time (fig. 2).

**Figure 2:** MOGEDs’ timer

MOGEDs are a kind of drill-and-practice educational software as they are based on question and answer interactions, providing appropriate feedback to the candidate (Κουτσογιάννης 2007). We could parallel the material offered with a corresponding past papers book, with the exception that the software does not require the intervention of the teacher (Κουτσογιάννης 2007). The correction and feedback is done by the software and can provide each test taker with an overview of their performance at the end of the test. Some of the advantages of drill-and-practice educational software are that it provides a) immediate feedback, b) a range of numerous repetitions “without exhausting the machine’s patience” and c) the ability to evaluate test takers’ performance (Κουτσογιάννης 2007).
As previously mentioned, MOGEDs usually adopt the typology of multiple-choice questions, consisting of three to four distractors with or without a drop-down menu, cloze tests questions, “Yes or No” or “True or False” questions. MOGEDs are interactive as they automatically provide feedback, a score and justification of both correct and wrong answers to test takers in order to make them aware of the errors they have made (Alderson et al. 2015, Jang and Wagner 2013). While providing feedback in the case of a correct answer, a frame pops up with the word “right” written in green and the justification for the given answer. If the answer is not correct, feedback is provided in exactly the same way with the word “wrong” in red and the justification for classifying the answer as incorrect. This procedure is necessary for a diagnostic test (Noijons 2013: 54).

Automatic scoring calculations available to test takers free up time for the teachers to carry out other activities. MOGEDs scoring methods have been tried out during piloting (Fulcher 2013: 219). Also, a special field, entitled “comment” is provided to both teachers and test takers to write their opinion. The comments submitted constitute important feedback to be taken into consideration for possible improvements and updates of MOGEDs (fig. 3).

![Figure 3: MOGEDs’ feedback and automatic scoring](image-url)
In terms of the characteristics of MOGEDs, as has already been mentioned, it is possible to repeat an e-test as many times as the test taker wishes. Moreover, they contain a management system that keeps track of the scores. Each user of the e-test has a personal account in order to have access to the set of e-tests and be able to pick up a test from where they stopped. All accumulated scores are saved in their account for easy access.

5. Limitations and prospects

In the near future, the enrichment of MOGEDs with more tasks per skill is required to ensure the validity of the tests. A set of e-tests for the certification of level A1 for young learners (8 to 12 years old) will also be designed and added, in accordance with the examinations for the Certificate of Attainment in Modern Greek organized by CGL. The purpose of the e-diagnostic test is a first, preliminary assessment of the degree of proficiency in Greek. Therefore, every future candidate of the exams for the Certificate of Attainment in Modern Greek should be given the opportunity to initially diagnose their strengths and weaknesses in receptive language skills and then submit an application for participation at the most appropriate language level.

The completion of the test is indicative. It provides potential candidates with indications for the exam results and does not mean in any case that success in the Greek language proficiency certification exams organized by CGL is guaranteed. Also, as the language level of candidates in productive skills cannot be checked by this application of automatic assessment, candidates’ performance in productive skills is suggested to be assessed by their teachers.

Furthermore, integration of literary texts appropriate for each language level in order for MOGEDs to comply with the CEFR Companion volume (2018) is considered imperative. Finally, a further improvement of MOGEDs will be the provision of justification for both correct and wrong answers in order to help test takers enhance their language knowledge. The feedback provided needs to be reinforced by theoretical, mainly morphosyntactic, information about the students’ errors and references to online supporting material. Upon completion of the above steps, MOGEDs will take on their final form.

As a final note, it should be pointed out that MOGEDs do not seek to be used as standard diagnostic tests. However, they indicate the intention of a state research
institution which conducts language proficiency certification exams to support potential participants in these exams as much as possible. The advantages of MOGEDs are that test takers may choose the language level and the skill in which they want to be assessed and have the opportunity to repeat the e-diagnostic tests as many times as they wish, as long as they are within the time limit. All those involved in the teaching of L2 Greek can benefit from educational technology, computers and the internet to promote and encourage students towards e-pre-assessment, before taking part in the high stakes exams for the Certificate of Attainment in the Modern Greek Language.

Notes

i For the working group see: http://www.greek-language.gr/certification/tests/index.html#pagebottom
iii https://dialangweb.lancaster.ac.uk/
iv Placement tests are not the object of this study.

Thomais Rousoulioti holds a PhD in didactics, specialized in teaching and assessment of Greek as a second or foreign language. She is a member of Special Teaching Staff at the Department of Italian Language and Literature, Aristotle University of Thessaloniki (since 2017) and a fellow Researcher at the Centre for the Greek Language in the division for the Support and Promotion of the Greek Language (since 2010). Currently is also a postdoctoral researcher in applied linguistics. She is a member of the Greek Applied Linguistics Association, of EALTA and ALTE.

thomaisr@itl.auth.gr

Rania Voskaki holds a PhD in Computational Linguistics, Automatic Translation, Computational Lexicography and processing of Language Resources. She is a research fellow at the Centre for the Greek Language, Divisions on Lexicography and Support and Promotion of the Greek Language (since 2007), a member of the “Natural Languages Processing Unit” of the “Translation and Language Processing Laboratory” of the French Department, Aristotle University of Thessaloniki (since 1999) and a member of the “Institute Gaspard Monge Laboratory” at the University of Paris-Est Marne-la-Vallée (since 2003).

rvoskaki@hotmail.com
Professor John Kazazis studied Classical Philology at the Aristotle University of Thessaloniki (AUTh), the University of Illinois and Yale University. Fellow at Harvard’s Centre for Hellenic Studies and Visiting Professor, University of Maryland. 1979-2014: Professor of Classical Philology, AUTh. From 1994 to date: Deputy Director and later Director of the “Centre for the Greek Language” and Director of its Division of Lexicography. 2010-2014: Director of the Greek National Council for Primary and Secondary Education. Honorary Professor at Kuban State University, Russia. He wrote extensively on Herodotus, Demosthenes, Archaic Greek Epic, Greek Epigram, and the History of Greek Studies in the West and Russia. As Redaction Director he edited vols. 15-21 of the “Kriaras” Lexicon of the Medieval Vulgar Greek Literature 1100-1699.

jkazazis@gmail.com

References


