journal homepage: http://tujted.com/index.php

Structure and Typology of the Greek State Foreign Language Exam System Marking Scheme for Oral Production

Dimitris Zeppos[1]

Hellenic Open Univeristy, Greece.

Abstract

The Greek State Foreign Language Exam, KPG, carried out in November 2013 introduced a new marking sheet for the assessment of oral production of KPG-candidates based on a five-point Likert-type resembling response format. The paper introduces the KPG-format response sheet and compares it with the theoretical specifications of Likert scales through a bibliographic review of respective studies. The conclusion is reached that the KPG-format does not represent a genuine Likert or Likert-type scale but a Discrete-Visual-Analog-Scale-(DVAS) Response-Format resembling a Likert scale, which functions as an input device of performance descriptors for the assessment of oral production of foreign language examinees.

Key Words: Assessment, discrete visual analog scale, KPG, likert scale, response Format

Introduction

In November 2013, the Greek State Foreign Language Exam System (KPG, short for [el] Kratiko Pistopiitiko Glossomathias), introduced the new examinations for the integrated C1/C2 level (henceforth described as C-level) according to the *Common European Framework for Languages* of the Council of Europe to the state-owned foreign language exam system. This paper aims at presenting and discussing the structure of the marking scale for oral production which was used for the first time during this first session the of C-level exams, thus establishing its typology.

The scope of this paper is to describe and analyze the marking scale used for the evaluation of the oral production during the integrated C1/C2 language exams, which in the accompanying Examiner Pack is described as a "*Likert scale*". The aim of the paper is not to evaluate the validity of the scale or the procedure of collecting data through this instrument, but to show whether and to what extend the instrument itself conforms to the characteristics of the genuine Likert scale. For this, the

---

[1] Corresponding author: Phone: +306974649981
   E-Mail   : zepposd@sch.gr (Dimitris Zeppos)

KPG-format, as the scale will be called henceforth, is questioned for its theoretical background and its consistency to the Likert scale "specifications" rendered in the respective literature.

It is shown that the KPG-format does not represent a genuine Likert scale, but rather an input instrument in the form of a DVAS-Response Format for Likert-type items, which are used for the input of indicators in order for the examinee to be evaluated a posteriori, according to a specific mathematical procedure, which is not being rendered.

Research Problem

It is common practice to describe tested oral and written language achievement with various qualitative or quantitative measuring instruments, called *scales of measurement*[1]. Usual typologies of such scales include *nominal*, *ordinal*, *interval*, or *ratio* scales (A. Brown, 2007; J. D. Brown, 2011)[2]. Each of these scales can be used for a variety of reasons in order to fulfill specific research, assessment or measuring requirements.

Implementing oral examination procedures in order to assess language proficiency on a performance basis for foreign language users is part of the modern assessment practice. The examination of one's ability to orally interact in a foreign language has become one of the most important factors of estimating foreign language proficiency. According to Powers (2010) students examined in the traditionally tested language skills *Reading* and *Writing* did not always have the same interactive ability to equally perform in Speaking skills in real life. Based on this, researchers and language assessors alike have come to the result that

> *[…] oral production is commonly said to be the most complex ability to test, due to its specific features, the long time required for its assessment and the transient nature of the speech act.[…] (Roca-Varela & Palacios, 2013, p. 54).*

Due to the difficulties in obtaining and – more significantly – maintaining objectivity and reliability of the exam procedure and outcome, efforts have been made to monitor the behavior of oral examiners[3] in order to set up, pilot-test and implement objective assessment instruments, such as scales, marking sheets and various other types of response formats[4]. One of the prevailing "protocol" systems used in such efforts is a marking sheet set up to resemble the widely familiar Likert scale.

Obeying the standards, requirements and specifications of the *Common European Framework of Reference for Languages: Learning, teaching, assessment* (CEFR)[5] the relatively new state foreign language exam system of the Greek Ministry of Education, known as KPG, introduced a differentiated marking sheet to the Greek exam system, based on a Likert-type scale of assessment, which served as an interview protocol for the oral exam of all levels and was for the first time used for the integrated C-Level language exam carried out in Greece in November 2013. This paper compares the traditional and widely used *Likert scale* form of measurement specifications with the KPG-format of assessment.

To what extend the KPG-format used for the assessment of language proficiency implements none, any or all of the above patterns, shall be dealt with by answering the above question, and supplemented by the following:

Is the KPG-format a genuine "Likert scale", a simple list of "Likert-type items" comprised on one single sheet without satisfying all characteristics and specifications of a genuine Likert scale as described in Boone & Boone (2012) and Clason & Dormody (1994) or a "Likert Response format" (Carifio & Perla, 2007).

Methodology

This paper is a literary research on the description of marking sheets, Likert(-type) scales and -items and respective theory about response formats using printed discrete visual analog input devices. An attempt is made to analyze the most common characteristics of measurement scales and the design of check-in input devices used in questionnaires. A bibliographic analysis of the "Likert scale" and its alternatives is carried out and the specific characteristics of such scales, their similarities and differences are discussed. By researching the bibliography on such measurement scales, no respective

publication concerning the newly established KPG-format has been found which could be utilized for further discussion in this paper.

Since the marking-sheet of the KPG language proficiency exam is to be discussed, a brief history of the actual exam as well as a short description of the exam modules is given, based on the documentation presented by the Greek Ministry of Education and the two Greek Universities that endorse the exam. Finally, the structure and logic of the actual KPG marking-sheet is described and parallels are drawn to the theoretical basis of the genuine Likert scale, in order to establish whether the KPG marking-sheet can be classified as a Likert scale, a Likert-type item Scale or a Response Format of a Discrete Visual Analog Attitude Scale.

Discussion on Scales

Likert scales

The name "*Likert*" scale derives from the American psychologist *Rensis Likert* (1932) who first introduced this instrument (Brown 2011: 10-11). A Likert scale is composed of a series of four or more sequential Likert items that are combined into a single composite score/variable during the data analysis process (Boone & Boone, 2012). As they state, "*the items are used to provide a quantitative measure of a character or personality trait*" (ibid).

Common to this kind of measuring instrument is that each query, which is presented in the form of a statement of attitude measure (Hitchcock & Porter, 2004), utilizes a number of response options, also called *anchors* (i.e. Grosshans & Chelimsky, 1993), usually odd in number and numbered from 1 through 5, 7, or 9 (van Elst, 2013, p. 81). Other types of such scales are set up with an even number of anchors, e.g. 2, 4 or 6, leaving out the "midpoint" choice, widely regarded to be a "neutral" point of the scale (J. D. Brown, 2011; Carifio & Perla, 2007; Garland, 1991; Norman, 2010; Raaijmakers, Hoof, Hart, Verbogt, & Vollebergh, 2000).

The main characteristic of a Likert scale is that every subsequent item of a query set is logically subordinate to the previous item of the same set, thus producing a series of descending (or deepening) logically structured query.

In short, as can be derived from Uebersax' (2006) definition, a *genuine* "*Likert*" scale consists of a series of subsequent and interrelated question items in form of short exclamatory or descriptive statements which produce an internally logical investigation of the query under examination. The respondent is required to state his "*[...] level of agreement with a series of attitude statements*" (Hitchcock & Porter, 2004, para. 1) for the "*intensity question*" (ibid) where the numbering represents an *ordinally ranked attribute of agreement* (van Elst, 2013, p. 81), usually expressed in descriptive nominal characteristics ranging from "strongly disagree" to "strongly agree". Likert scales are generally used to measure attitude, therefore they very often are also called "*attitude tests*" (Chomeya, 2010, p. 400).

Likert-type items

Likert-*type* items are often mistakenly called Likert items because the former are very closely conforming to a number of characteristics of the latter, including layout and wording. In the past decade a number of scholars (Clason & Dormody, 1994; Gadermann, Guhn, & Zumbo, 2012; Gliem & Gliem, 2003; Pearse, 2011; Uebersax, 2006, to name just a few) have identified a number of confusing references of the use of the term Likert-*type* item, which is interchangingly used with the term Likert item.

As Clason & Dormody (1994) point out in their work, Likert items have a very strong interrelated logical structure, while Likert-*type* items represent a set of independent question sets which offer the respondent the opportunity to state "*to which of several ordered alternatives they belong*". Therefore, Likert-*type* items have the same visual layout as Likert items, but Boone & Boone (2012) and Brown (2011) conclude that they lack interior dependency.

Another difference to the genuine Likert items is the fact that Likert-*type* items allow for ordinal measurement while the genuine Likert items accommodate interval measurement as shown in Boone

and Boone (2012). Extensive discussion on this statistical issue is described in Brown (2011), where various alternative solutions are proposed.

As far as the analysis of the item responses is concerned, Carifio & Perla (2007) show that Likert-*type* items allow for the separate statistical analysis of each item, since, as mentioned above, they represent non-inter-correlated subjects of investigation.

In short, as Uebersax (2006) also reasons in his seminar notes, Likert-*type* items resemble Likert items in layout, number of response anchors and the presence of attitude level choices. The difference to the *genuine* Likert items is that they do not correlate internally nor do they have a logical descending (or deepening) structure, thus focusing on different weight of interest for each item.

Likert Response Formats

A measurement scale consists of a number of questions or, in the case of Likert questionnaires, statements which the respondent is required to answer by choosing from a given set of responses. Likert response formats are very commonly used in social science measurements, where

[…] respondents are asked to indicate their level of agreement with an item by choosing one of a given number of ordered response categories, e.g., with five categories ranging from 'strongly agree' to 'strongly disagree' […] (Gadermann et al., 2012, p. 2).

Very commonly, Likert-*type* response formats are also used for other types of questionnaires, such as *semantic differential item scales*, as Uebersax (2006) describes them, where the layout is the same as in Likert scale response formats but the cohesion of the question items differ.

In short, Response Formats are a given type and layout of response choices, accompanied by either verbal, numerical or visual descriptors with specific edge points, in which case the Response Format accommodates differential item scales. In case of the responses being constrained to specific values which do not allow for the responder to deviate, such response formats facilitate *discrete visual analog scales*, since the climax of the anchor points is depicted and defined in the order of the verbal or numerical response choices offered.

KPG in short history and description

Greece has a long tradition in endorsing and requiring foreign language exams for all ages and levels of language expertise. A series of "imported", non-Greek, language exams were established during the last decades. These exams were conducted according to specific examination regulations set out by foreign language centers like the Cambridge University for *English as a Foreign Language* (EFL) or the Goethe Institut for *German as a Foreign Language* (GFL).

This tradition was later on supplemented by efforts of private Greek language teaching institutes, which in 1981 established an interior examination body named *Panhellenic Federation of Language School Owners* (P.A.L.S.O.), through which they intended to certify the foreign language proficiency of their students[6].

The public requirement for language certification was one of the reasons, why the Greek Ministry of Education decided to set up a nation-wide examination system for the certification of foreign language knowledge in 1999. As Prof. B. Dendrinos of the Research Center of Language Teaching, Testing and Assessment (RCEL) stated in 2009, the KPG

[…] offers exams in European languages of 'value' to the Greek society. It is a product of and a service by the Greek state at no material or symbolic profit. Exams are offered to people living, studying and working in Greece. […] [7].

In 2003, the KPG exams were administered nation-wide for the first time for four European languages, namely English, French, German and Italian[8]. These first exams were on level B2 according to the 6-scale language proficiency level format introduced by the CEFR, while the other levels were gradually added during the following years. Spanish was added in 2010 while Turkish became the sixth language to be examined through the KPG system in 2011. Both latter languages were examined only on levels B (integrated) and C1 when they were first included to the exam system.

Every year there are two exam periods (November and May) for all levels, with the exception of the integrated A1/A2 level, which is examined only in May of each year since 2008. Since November 2013 the latest change of the exam is the integration of the two C levels (C1 and C2) into one integrated C level for all languages, except for Turkish, thus introducing the C2 level, since this was the only remaining level of the CERFL 6-level-scale description for the KPG exam system which had not been part of the exams.

In all exams the same examination modus defined by the specifications given by the Central Examinations Board (CEB)[9] is implemented. Common characteristic of all exam levels is the modular system used throughout the languages and their respective levels. So, all exams consist of four modules[10]:

The first module, called "Reading comprehension and language awareness" aims at assessing candidates' understanding of the overall or partial meanings in a text, to make inferences or intelligent guesses on the basis of the text. Moreover, some items are designed to assess their ability to make language choices that are correct and appropriate to the linguistic, discursive and social context.

The second module named "Writing and written mediation" tests candidates' ability to produce written discourse and function as mediators through written production. It consists of 2 activities: 1 activity for written production where candidates are asked to interact with and respond to an English text and 1 for written mediation based on a Greek text.

The third module is called "Listening comprehension". This module tests the candidates' ability to understand standard spoken language on both familiar and unfamiliar topics normally encountered in personal, social, academic and vocational life. Candidates always have the opportunity to listen to the recordings twice before responding.

Finally, the fourth module, which is the one of interest for this paper, is called "Speaking and oral mediation". This module tests the candidates' speaking production and, in more detail, the ability a) to state with precision, justify and support views, ideas or opinions on social issues and practices (Activity 1), b) to state attitude, opinion or view on a specific stimulus, in response to certain lead-in questions (Activity 2) and c) to mediate complex information and discuss a difficult subject, using information from a Greek source text (Activity 3).

Introduction to the KPG "C-level Oral Production Evaluation Scale"

As mentioned above, a Likert-type "input format" is the main instrument used by the KPG exam system to evaluate the candidates' oral production of the foreign language.

In the KPG, this format functions on a two-dimensional base. The first dimension refers to the overall accomplishment of the candidate in the three activities required, while the second dimension measures six different oral skills on a can-do basis, according to the CEFR descriptions for each language proficiency level. The actual printout of the KPG-format is set up to be used by two examiners at the same time, one being the (interacting) interlocutor and the other the (non-interacting) rater of the candidate. The paper-scale is printed in Greek, since it is part of a Greek state examination system[11]. The following description of the actual marking sheet assumes that the reader does not have the necessary knowledge of the Greek language to fully comprehend the instructions given, which makes it necessary to describe the marking sheet together with the annotations in all detail and with the assistance of translations of crucial termini.

The unique characteristic of the KPG-format, as mentioned above, is the fact that it combines two table-like scoring lists for each candidate on the same piece of paper, divided into two main categories. The sheet is divided into nine rows of five items each resembling Likert-type statements. The first three rows measure the overall production of the candidate for each of the three activities examined in the test (Activity 1: debate, Activity 2: one-sided talk: description/disputation of attitudes and Activity 3: oral production and mediation). The remaining six rows address equal numbered specific language awareness, linguistic, semantic and grammatical aspects of the oral proficiency of the speaker. It is assumed that the following description details the first column of the two-column scoring sheet of the exams.

Each item of the response format consists of a "*stimulus component, a response/scoring component and a context component*" (Carifio & Perla, 2007, p. 114). The first is not actually printed

on the marking sheet but accompanies it in the examiners' booklet given to the examiners previous to the exams.

The stimulus components of the marking scale are the verbal descriptions of each activity or skill interrogated. In detail, the descriptive stimuli are[12]:

- for Activities 1 through 3 (items 1, 2 and 3):
- How well did the candidate cope with the required activity?
- for each of the language skills examined:
- Articulation and intonation (item 4)
- Lexical range and control (item 5)
- Grammatical accuracy (item 6)
- Appropriateness of language choices (item 7)
- Interactional skills (item 8)
- Cohesion, coherence of speech and fluency (item 9)

Each response/scoring component includes five assessment anchors numbered "1" through "5". Therefore the scale in use is a "midpoint" scale (Carifio & Perla, 2007; Garland, 1991; Raaijmakers et al., 2000; Wakita, 2004) offering the evaluator the choice to deliver evaluation points "2" and "4". These can be regarded as midpoint values for the main anchors "1" and "3" or "3" and "5" respectively. Anchor "3" itself poses as the midpoint for the two utmost anchors "1" and "5".

According to the instructions in the Examiner Pack of the November 2013 exams, each point value refers to a specific verbal description, which does not represent an attitude statement though, and is not depicted in the KPG-format itself. In detail, these values are as follows[13]:

Value (1) → not satisfactory
Value (2) → marginally satisfactory
Value (3) → moderately competent
Value (4) → very satisfactory
Value (5) → completely satisfactory

The exact computing string of the automated OCR-rating system has not been revealed to either the examiners or the candidates, obviously in order to avoid influence on the examiners' judgment. Instead the instructions given to the examiners state that anchor "1" delivers a "possible" evaluation of "non-pass" for the lower C-level (former C1 according to the scale set by the European Council), while "3" translates into "pass" for level C1. Anchor "5" would respectively produce the highest possible mark for the specific row, acknowledging CEFR "C2" output.

KPG-format

Comparing it with Likert scale, Likert-type scale or Response Format specifications as pointed out in Uebersax (2006), Likert scales are characterized by the presence of verbal labels for each response anchor, printed above each respective anchor. The KPG-format, on the other hand, conforms to the definition of *Semantic Differential Item Scales*, since the response anchors are described with numeral values which imply "specific metric relations" (ibid) among the response levels, while the stimuli are printed on a separate location and detached from the actual check-points. Such scales which constrain the respondent to a specific set of pre-specified levels are called *Discrete Visual Analog Scales (DVAS)* (ibid).

Furthermore, Likert scale items articulate the interrogated point of interest in form of a statement, which allow for the respondent to position him/herself on the attitude scale denoted. The KPG-format displays a series of questions in the form of question sentences (with question mark), which clearly differentiates the KPG-format from the Likert scale, even though the accompanying instructions in the Examiners' Pack (Κεντρική Επιτροπή Θεμάτων - Central Board of Examination KPG, 2013, pp. 5–6) deliver specific assessment statements, the purpose of which is to assist the assessors in choosing the appropriate pre-described response according to the oral production of the candidate.

Likert-*type* items, which - as discussed previously - are a set of non-interrelated questions in form of statements, nevertheless conform to a number of Likert scale characteristics. However, as can be seen in the actual KPG-format, the layout of the Likert-type item, namely the presence of verbal labels for each item superseded by interval numerical response anchors, is not implemented. Here each item is accompanied by a question sentence, placed in the centre of the printed form in order to accommodate both assessors and thus avoiding unnecessary use of space on the printed sheet. Nevertheless, the actual stimulus type is not a "statement" but a "question" which leads to the conclusion that the numbered scale given next to each stimulus is a ranking anchor from "worse" to "best".

The centered position of the item stimulus itself leads the assessor to relate the question to the production of the candidate in a sequential procedure. This fact does conform to the original idea of Likert for a linear placement of attitudes on "*cut-off points*" (Pearse, 2011, p. 160). Clearly, the KPG-format uses Likert-*type* items in the sense of offering specific descriptors for the assessor, but it is also very obvious that these items function as guidelines so that the assessor can place the oral production on the continuum of the assessment line formed by the increasing numbers of the anchors.

Given the fact that for every exam procedure there must be an instrument of gathering usable data, the format for the KPG exams obviously assists in collecting unified assessment estimates for every candidate in an ordered and predefined manner. Therefore, the KPG-format as used by the examiners during the November 2013 run, functions as an input-tool for the oral exam. Furthermore, according to the accompanying handouts of the introductory seminars conducted in November 2013 by the National Kapodistrian University of Athens (EKPA), which is one of the scientific advisors and exam system designers of the KPG[14], the configuration of the KPG-format is set up in such a way that it accommodates complex mathematical strings to calculate the overall oral achievement of the candidate[15]. The data collected is computed *a posteriori* by means of specific OCR software[16], producing the final score of the examinee. This score, as also mentioned in the accompanying instructions of the Examiners Pack (2013, p. 5), can and should not be predictable by the assessors.

Finally, as can be seen in the description of the query stimuli presented above for each of the items given in the KPG-format, none of the stimuli represents *attitude* statements, which is the main obligatory characteristic to genuine Likert scale questionnaires. Furthermore, as can also be instituted in the instructions given to the examiners prior to the actual sitting, the actual KPG-format input data does not function as an *additive attribute* (van Elst, 2013, p. 81) for the overall assessment of the candidate.

Result and Conclusion

After analyzing the characteristics, prerequisites and descriptive details found through bibliographic research, it has been established that the KPG-format used in November 2013 does not form a Likert scale in its original sense.

The basic characteristics of a Likert scale, which are interrelated logical connections of each item in a group of statement items as well as the verbal labels for each item, are not satisfied in this format. Furthermore, the scores denoted in this format are not used as the actual assessment of the candidate. Instead, they function as indicators for an OCR-driven mathematical processing of the oral achievement of the examinee.

Despite the fact that the KPG-format is described as a Likert scale in the actual instruction material included in the Examiner Pack (Κεντρική Επιτροπή Θεμάτων - Central Board of Examination KPG, 2013, p. 5), it does not conform to the characteristics cited for a Likert scale, even though it allows specific responses on the five-anchor numeric bar for each Likert-type resembling item requested. According to the previous literary analysis it is obvious that the KPG-format is actually a DVAS-Response-Format, not a genuine Likert scale.

References

Boone, H. N. J., & Boone, D. A. (2012). Analyzing likert data. *Journal of Extension*, *50*(2), 30.

Brown, A. (2007). An investigationof the rating process in the IELTS oral interview. *IELTS Collected Papers*, 49–84.

Brown, J. D. (2011). Likert items and scales of measurement? *JALT Testing & Evaluation SIG Newsletter*, *15*(March), 10–24.

Carifio, J., & Perla, R. J. (2007). Ten Common misunderstandings, misconceptions, persistent myths and urban legends about likert scales and likert response formats and their antidotes. *Journal of Social Sciences*, *3*(3), 106–116.

Chomeya, R. (2010). Quality of psychology test between likert scale 5 and 6 points rungson chomeya department of educational psychology and guidance, Faculty of Education, *Journal of Social Sciences*, *6*(3), 399–403.

Clason, D., & Dormody, T. (1994). Analyzing data measured by individual Likert-type items. *Journal of Agricultural Education*, *35*(4), 31–35.

Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data : A conceptual , empirical , and practical guide. *Practical Assessment, Research & Evaluation*, *17*(3).

Garland, R. (1991). The Mid-point on a rating scale : is it desirable? *Marketing Bulletin*, *2*(RN 3), 3–6.

Gliem, J. A., & Gliem, R. R. (2003). Calculating, Interpreting, and reporting cronbach's alpha reliability coefficient for likert-type scales, (1992), 82–88.

Grosshans, W., & Chelimsky, E. (1993). *Developing and using questionnaires*. United States General Accounting Office (GAO).

Hitchcock, A., & Porter, K. (2004). The Likert Scale. *Methodology*, 7–10.

Karavas, E., & Delieza, X. (2009). On site observation of KPG oral examiners: Implications for oral examiner training and evaluation, *3*(1), 51–77.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22*, 5–55. doi:2731047

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education : Theory and Practice*, *15*(5), 625–32.

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

Pearse, N. (2011). Deciding on the scale granularity of response categories of likert type scales : the case of a 21-point scale. *Electronic Journal of Business Research Methods*, *9*(2), 159–171.

Powers, D. E. (2010). The case for a comprehensive, four-skills assessment of english-language proficiency. *Educational Testing Service*, (14).

Raaijmakers, Q. A. W., Hoof, A. Van, Hart, H. 't, Verbogt, T. F. M. A., & Vollebergh, W. A. M. (2000). Adolescents' midpoint responses on Likert-type scale items : neutral or missing values? *International Journal of Public Opinion Research*, *12*(2), 208–216.

Roca-Varela, L., & Palacios, I. M. (2013). How are spoken skills assessed in proficiency tests of general English as a Foreign Language? A preliminary survey 1. *Interantional Journal of English Studies*, *13*(2), 53–68.

Uebersax, J. (2006). Likert scales: dispelling the confusion. *Statistical Methods for Rater Agreement Website*. Retrieved from http://john-uebersax.com/stat/likert.htm

Valax, P. (2011). *The Common European Framework of Reference for Languages: A critical analysis of its impact on a sample of teachers and curricula within and beyond Europe*. University of Waikato.

Van Elst, H. (2013). Foundations of Descriptive and Inferential Statistics. Karlshochschule International University.

Wakita, T. (2004). Assessment of the distance between categories in rating scales by using the item response model. *Shinrigaku Kenkyu : The Japanese Journal of Psychology*, *75*(4), 331–338.

Wylie, E. (2002). An overview of the International Second Language Proficiency Ratings. *ISLPR Overview*. Griffith University: Centre for Applied Linguistics and Languages (CALL).

Κεντρική Επιτροπή Θεμάτων - Central Board of Examination KPG. (2013). Εξέταση προφορικού λόγου επιπέδου Γ (Γ1&Γ2) - Αγγλικά - Νοέμβρης 2013. Αθήνα: Υπουργείο Παιδειας και Θρησκευμάτων.

---

[1] Such measurement scales are put to use for language proficiency tests like the TOEFL® test (http://www.ets.org/toefl), the Cambridge English Language Assessment tests of the University of Cambridge (http://www.cambridgeenglish.org/gr/exams-and-qualifications) and the German Goethe Institute language exams (http://www.goethe.de/lrn/prj/pba/deindex.htm), to mention just a few.

[2] For more on scales of measurement and further examples, see Brown (2011), Grosshans & Chelimsky (1993), Norman (2010), O'Sullivan, Weir, & Saville (2002) and Wylie (2002).

[3] For an example description of such observation efforts also see Karavas & Delieza (2009).

[4] For a detailed description of the most common exam schemes used for oral assessment also see Roca-Varela & Palacios (2013).

[5] http://www.coe.int/t/dg4/linguistic/CADRE1_EN.asp . For a more detailed description of the CEFR also see Valax (2011).

[6] For further details on P.A.L.S.O. visit http://www.linkedin.com/company/palso-federation and http://www.palso.gr/?i=portal.en.federation-history (last visited in November 2013).

[7] http://rcel.enl.uoa.gr/kpg/gr_kpgcorner_feb2009.htm (last visited in November 2013)

[8] For more information on the history of the implementation of the KPG exams see also http://rcel.enl.uoa.gr/sapig/gr_sapig_and_kpg.htm (last visit in November 2013 – Greek language only)

[9] See details of test specifications and regulations at http://rcel.enl.uoa.gr/kpg/organisation.htm (last visited in October 2013).

[10] The short descriptions of the four modules are taken from the "ORAL EXAMINER INFORMATION PACK" for level C1 as published at http://rcel.enl.uoa.gr/files/KPG/english/C1_level_oral_examiner_info_pack.pdf (last visit in October 2013). For detailed information on all levels also see http://rcel.enl.uoa.gr/kpg/exam_train.htm.

[11] A detailed description – in Greek and English– together with an example of the November 2013 exams is given at http://rcel.enl.uoa.gr/files/kpg/english/tests/2013-11/EN_C_NOV13.zip (English, last visit in November 2013) and http://www.minedu.gov.gr/publications/docs2013/Kpg_GE_c1_c2.zip (German, last visit in November 2013).

[12] The translations of the Greek stimulus descriptions of items 1 through 9 are borrowed from the Examiner Pack of the May 2013 exams for the C1 level for the respective assessment responses (http://rcel.enl.uoa.gr/files/KPG/english/C1_level_oral_examiner_info_pack.pdf, last visited in November 2013).

[13] The translations of the Greek descriptions of the values (2) and (3) are borrowed from the Examiner Pack of the May 2013 exams for the C1 level for the respective assessment responses. The remaining three values (1), (4) and (5) are translations by the author of this paper.

[14] The other advisor to the Exam system is the Aristotle University of Thessaloniki (http://kpg.auth.gr/index.php/el/).

[15] This mathematical scheme, though, shall not be discussed in this paper.

[16] Hence the layout and the use of pre-described check boxes, the color red for the print and the obligation of the examiners to use either black or blue ink for the completion.