eLADE: The University of Granada Bi-level B1/B2 Accreditation Exam. Construction and Validation Process

Aurora Biedma Torrecillas, Lola Chamorro Guerrero, Alfonso Martínez Baztán, Adolfo Sánchez Cuadrado and Sonia Sánchez Molero

Abstract: The Centre of Modern Languages of the University of Granada offers the eLADE B1/B2 (*Examen en Línea de Acreditación de Dominio de Español*[1]), the first e-Test of Spanish Proficiency to be completely reliable. This test is aligned with the CEFR and complies with the standards for best-practice assessment of international institutions. It is also recognized by all the Universities belonging to the Associations of Language Centres in Higher Education in Spain (ACLES) and Europe (CERCLES). In this article we will describe the test, its specifications and administration, together with the construction and validation process that make it the first online test of Spanish proficiency to be fully fair, validated and reliable. All this information is aimed for potential candidates and policy-makers, as well as other test developers, for the sake of transparency and good testing practices.

Key words: Spanish, language accreditation, online testing, test specifications

Abstract: El presente artículo describe el proceso de elaboración y validación del examen Elade B1/B2 (Examen en Línea de Acreditación de Dominio de Espaol), el primer examen en línea de espaol completamente validado. Este examen es desarrollado por el equipo de evaluación del Centro de Lenguas Modernas de la Universidad de Granada y se realiza cumpliendo con los más exigentes estándares y códigos de buenas prácticas en evaluación de los organismos internacionales, por lo que está reconocido por varias instituciones oficiales (como ACLES y CERCLES). En el presente artículo se describe sus especificaciones, así como su proceso de creación y validación con el fin de poner a disposición de todos los potenciales candidatos y responsables académicos, así como creadores de exámenes, la información necesaria, dentro de un marco de transparencia de buenas prácticas.

Abstrakt: Centrum moderních jazyků na Univerzitě v Granadě nabízí eLADE B1/B2 (Examen en Línea de Acreditación de Dominio de Español), první Proficiency e-test pro španělštinu, který je zcela spolehlivý. Tento test je v souladu s CEFR a splňuje standardy nejlepšího hodnocení mezinárodních institucí. Je také uznáván všemi univerzitami patřícími k Asociaci jazykových center ve vyšším vzdělávání ve Španělsku (ACLES) i v Evropě (CERCLES). V tomto článku popisujeme tento test, jeho specifikace a administraci spolu s procesem konstrukce a validace, které z něj dělají první on-line test Spanish Proficiency, který je zcela spravedlivý, legální a spolehlivý. Všechny tyto informace jsou zaměřeny na potenciální kandidáty a činitele stejně jako na další tvůrce testů za účelem transparentnosti a zlepšování testovacích postupů.

---

[1] Spanish Testing and Accreditation Team (Aurora Biedma Torrecillas, Lola Chamorro Guerrero, Alfonso Martínez Baztán, Adolfo Sánchez Cuadrado, Sonia Sánchez Molero). Multimedia and design: (César Amador Castellón, Jesús Puertas Melero, José Rodríguez Vázquez)

---

1 Rationale behind an online Spanish accreditation exam

In today's global world, the demand for second language learning is an ever-growing concern. One particularly important requirement is the accreditation of the more widely-spoken languages such as English and Spanish, which has led to the development of standardized officially-recognized proficiency tests both for education and the job market.

The increased mobility among European students fostered by the Bologna Plan and the recent implementation of the Erasmus Plus programme beyond Europe have both brought about a considerable increase in the demand for Spanish language proficiency tests. At the University of Granada (UGR), it was felt necessary to address these issues, particularly given the high demand for graduate and postgraduate courses from students from different international programmes. The university faced the problem of how to satisfactorily certify students' level of Spanish, both to determine their ability to participate in courses and for the purpose of awarding financial grants. As in other European universities, the University of Granada established a series of benchmarks necessary for the correct assessment of students' language level.

While some prospective students were already accredited by the Cervantes Institute's internationally-recognized DELE exam, most either did not have access to an official examination centre or could not make the limited exam dates, while others simply presented unofficial qualifications which were deemed unacceptable for certification purposes. This situation put the university in the difficult position of having to award mobility grants without knowing if the student in question did in fact have the necessary level of Spanish. In response to this problem, the Department of International Relations assigned the University of Granada's Modern Language Centre the task of developing a bi-level B1/B2 proficiency test, which was subsequently carried out by the Spanish Department's specialist test development team in strict accordance with the directives established by the CEFR, EALTA, ALTE and the European Council (CoE, 2001; CoE, 2009; CoE, 2011; ILTA, 2000; ALTE, 2001; EALTA, 2004). The first administrations of this test were paper-based and took place in the University of Granada Modern Language Centre.

As a result of the complications encountered by some foreign students in taking a local paper-based test in Spain, it was decided to create an equivalent online proficiency test that would fulfil those students' accreditation requirements. In this way, students would be able to certify their language level from their own universities or education establishments, thereby securing their grants before travelling to Spain.

Furthermore, it was subsequently realized that other Spanish universities were also facing the same problem and it was therefore decided that the University of Granada should apply for national ACLES (Asociación de Centros de Lenguas de Enseanza Superior) accreditation status in order to be able to help them offer a viable proficiency

test to their own mobility students. Accreditation status was granted by ACLES and CERCLES in 2012 and subsequently by the CRUE (the Committee of Spanish University Rectors) in 2014.

More recently, demand from other quarters has also increased as knowledge of the ease of application and reliability of the eLADE exam has grown. This interest comes both from educational establishments such as secondary schools and language centres, and private individuals interested in accrediting their language ability.

2 General Description

The eLADE exam is an online, bi-level test which certifies candidate performance at either B1 or B2 level in listening and audiovisual comprehension, reading comprehension, and both written and spoken production, and interaction as defined in the CEFR.

Below is a brief description of the scales and descriptors used in the development of tasks for each of the linguistic competences assessed in the eLADE test.

Listening comprehension

A B1 candidate is expected to understand the main ideas of clear, standard speech dealing with everyday topics. A B2 candidate should be able to understand both abstract and concrete topics, as well as those specialized topics related to the candidate's professional or academic specialism. In both cases, candidates should be able to understand conversations between native speakers, conferences or presentations, announcements or instructions, broadcasts and recorded material. The difference between B1 and B2 in this regard is the length of the discourse and the level of complexity of the ideas expressed, even though in both cases there should always exist some form of explicit discourse structure (as opposed to the C1 level, where this is not a requirement). B1 and B2 level candidates should also understand audio-visual material such as most TV programmes and films, even though slow, clear articulation and greater visual support will be necessary at B1 level.

Reading comprehension

A B1 level candidate should be able to understand simple texts, which are either of a general nature or related to the candidate's field of expertise, to an adequate level. A B2 level candidate should be able to work easily with a large variety of texts due to a high level of reading vocabulary, even though their understanding may well be hampered by colloquial language and idiomatic expressions. The test aims to verify that candidates are capable of adapting themselves to the situation at hand, to different text types and reading goals such as scanning, searching, understanding instructions.

Written production

A B1 level candidate is expected to be able to produce simple texts with an adequate level of cohesion concerning everyday topics or topics of interest to the candidate, as well as certain other text types (i.e. letters, reports, notes, messages, forms etc.). A B2 level candidate is required to produce more complex texts, not only of a general nature but also those pertaining to the candidate's field of expertise, in which he or she should be able to express an argument as well as bring together or re-write information from several different sources.

Spoken production

A B1 level candidate is expected to maintain a short monologue with a degree of fluency in which he or she describes topics of interest, recounts experiences or provides simple arguments. In terms of spoken interaction, the candidate should be able to exchange information about standard topics and possess the necessary linguistic repertoire to be able to deal with everyday situations. To this end, a candidate is also expected to understand a native interlocutor both in formal and informal conversations. Once again, the difference between B1 and B2 concerns topic type, which in B2 contains a wider range of specialist topics as well as the necessary level of detail and complexity and the inclusion of arguments and relevant examples to illustrate the ideas expressed. In spoken interaction with native speakers a B2 level candidate should be able to maintain a conversation fluidly and naturally enough so as not to cause any strain between participants.

3   Description of the tests, tasks and items

The exam consists of four parts designed to evaluate six communicative linguistic competences. These are: reading comprehension, listening comprehension, written production and interaction, and spoken production and interaction. The exam lasts three hours fifteen minutes and all tasks contribute to the attainment of a B1 or B2 grade. Indeed, for accreditation at either level the candidate must pass all four parts of the exam.

The Table 1 shows the number and length of tasks in each part of the exam.

Tab. 1: *The number and length of tasks in each part of the exam*

Listening Comprehension	Reading Comprehension	Written Production and Interaction	Spoken Production and Interaction
5 listening/ audiovisual tasks	5 reading tasks	2 writing tasks	3 oral production and interaction tasks
45 minutes	75 minutes	60 minutes	10–15 minutes
Online	Online	Online	Via Skype or on-site at CLM

Listening Comprehension: 45 minutes

The listening comprehension part consists of five listening or audio-visual tasks. Depending on their content, at least two of the tasks are set at B1 level and at least two at B2 level. The duration of each recording varies between two to five minutes. Recordings are sourced from news items, reports, conferences and adverts containing descriptions, arguments, explanations or instructions. The social register employed is standard (informal, formal and/or academic).

In total, the tasks consist of 25–30 items (5–10 per task). Each recording is heard twice. Item types on comprehension tasks are the following: a) three- or four-option multiple choice questions; b) multiple match questions; c) short answer questions with a maximum of five words.

Reading Comprehension: 75 minutes

The reading comprehension part consists of five reading tasks. Depending on their content, at least two of the tasks are set at B1 level and at least two at B2 level. In total, there are between thirty to thirty five items (5–10 per task). Texts length ranges from 250–380 words at B1 and from 350–500 words at B2.

The texts chosen are taken both from personal and public domains and are all written in a standard social register (informal, formal and/or academic). As far as possible, text selection aims to cover different text types (e.g. emails, articles, editorials, reviews, adverts, guides and instructions among others), different linguistic functions, as well as providing a variety of topics.

Item types for reading comprehension tasks are the following: a) three- or four--option multiple choice questions; b) short answer questions with a maximum of five words; c) True (T), False (F) plus justification; d) multiple match 1: reinsertion of a previously-extracted word or piece of text in its original position; d) multiple match 2: matching text and items from two separate columns.

Written Production and Interaction: 60 minutes

The written production and interaction part consists of writing two texts, the first of which can be completed at B1 level, while the second requires B2 level language. The B1 task is 15–200 words long and the B2 task is 250–300 words long.

Candidates must produce two text types: a) a neutral or informal letter/e-mail, or narrative; b) a formal letter/e-mail, article, report or essay. Each text type demands a specific type of discourse – descriptive, narrative, explanatory, or argumentative (this last only at B2).

Spoken Production and Interaction: 10–15 minutes

The spoken production and interaction part of the test consists of three tasks: the first two are designed to evaluate B1 level while the third task evaluates B2.

The first task consists of an interview of between 2–3 minutes long which covers personal questions about the candidate (country of origin, work, studies, likes, etc.). The second task is between 2–4 minutes long and is structured as a role-play interaction in which the candidate must be able to successfully resolve a dispute. Task three is a presentation or monologue on a controversial topic of 4–5 minutes in length in which the candidate must explain and argue their point of view and which serves as a springboard for further discussion with the interlocutor for 2–3 minutes. This task is not improvised as the candidate has 10 minutes to prepare prior to the beginning of the interview. For tasks 2 and 3 the candidate may choose between three options.

Marking Criteria

Both the written and oral production and interaction parts of the exam are marked following CEFR recommendations. Accordingly, the marking scales for the marking of written production and interaction the following criteria are taken into account: 1. task fulfilment; 2. coherence; 3. accuracy; 4. range. Furthermore, spoken production and interaction scales also take into account fluency (5.) and interaction (6.).

4   Test development and piloting

The eLADE exam is subjected to protocols and guidelines of good practice at all stages of its planning and design procedure in order to assure the following: a) its conformity to the CEFR; b) its validity, fairness and reliability; c) its appropriateness of level; d) its aptness for the correct evaluation of its intended use in the accreditation of Spanish language proficiency at B1 and B2 levels. (Alderson, 1998; Bachman, 1990; Bachman, 1996; Bachman, 2004).

Both the written and spoken production and interaction parts are developed based on an analysis of content and linguistic functions relative to language use context and the required ability at B1 and B2 levels. Tasks are developed to be functionally valid for the candidates doing them, particular care being taken that they do not contain any elements which would either favour or penalize any one type of candidate. All speaking and writing tasks are trialled with students, natives and expert judges to check the clarity of the instructions and the suitability of the level of the prompts and capacity to elicit the expected samples of spoken and written language. Elicited responses should provide sufficient evidence in terms of quantity and quality to be able to give an accurate and fair assessment of a skill when applying relevant assessment scales and criteria. Both spoken and written production tasks are double marked. Crucially, the above procedures are subjected to critical analysis during benchmark-

ing sessions in order to avoid the potential biases which may be encountered in this type of test and standardize exam protocol. To this end, the team has developed and continues to revise its test specifications, scales and marking criteria, double marking protocols, mark sheets, and revision protocols. Furthermore, statistical analysis is used to determine both intra- and inter-rater [JS1] reliability of both production parts of the exam.

Similarly, reading and listening comprehension test design is subjected to critical analysis by expert judges, L1 speakers, and L2 speakers. The selection of source material, text-mapping by expert judges to identify relevant information, task design and item revision processes are all carried out according to previously established protocols in order to ensure both face and content validity. Tasks are then piloted and subjected to reliability analyses. Reading and listening comprehension tasks are subjected to both classical analysis using SPSS and Rasch analysis using the WIN-STEPS program.

5 Test Administration

While the eLADE test may be taken at the CLM, the University of Granada's modern language centre, it can also be administered in any centre around the world which has undertaken a prior agreement with the university to this effect.

Where no test centre is available near a candidate's place of residence, they may request that their local academic institution apply for accreditation through our web page. To be eligible for accreditation, academic institutions are required to have access to the necessary technical resources specified in the administration guidebook and to sign a confidentiality agreement in which examiners agree not to disclose any details of exam content. Similarly, each exam centre must assign an administrator who will follow test procedures before, during and after the exam administration in order to confirm the candidate's identity and ensure they sit the exam individually without either outside help or the use of reference materials. While this individual need not speak Spanish in order to administer the test, they must have a basic grasp of English as the administration protocols are written only in English or Spanish.

Upon candidate registration, the examining centre concerned is assigned a date and time for the oral exam. Administrators are not required to undergo any special training but do however receive detailed instructions about exam protocol for the administration of each part of the test, including timing, necessary technical resources, and candidate instructions.

6   Design of the online exam: Technical aspects of the eLADE test

Compatibility

From the very beginning, the eLADE test was designed to be administered at a global level. For this reason, one of the main challenges was to design a platform that would allow its administration using computers with low resources and which would be easily accessible from any device with internet access. To that end, the platform has been specially designed using the Web 2.0 protocol in order to guarantee the greatest degree of compatibility possible with candidates' devices.

Security

Candidates must pass through a two-tier security process in order to sit the exam. Firstly, access to the exam web page is kept secret until just before the exam. Secondly, a username and password are required to log onto the webpage. While the webpage may be previously accessed by administrators in order to test compatibility, the exam itself cannot be accessed without the introduction of the candidate's personal access details. To further improve exam security, the platform itself is run from a secure server.

Technical considerations

Video reproduction uses the *webm* codec in order to ensure a high degree of compatibility with browsers. This codec is decoded by the browser itself and does not require any external software.

In order to deal with any keyboard problems resulting from the specific nature of the Spanish alphabet and punctuation system, the exam platform includes a small tool for the introduction of Spanish characters and diacritic marks. The function keys on candidates' keyboards are also disabled remotely to avoid possible input errors.

Candidates' answers are saved and updated in real time to an online database. This data is 100 % recoverable should a technical problem occur. Due to the fact the exam is both timed and monitored, it can be re-started from exactly the point in which the error occurred.

7   Validity Statistics

This section provides information on the use of statistical analysis to assess exam validity both as a whole and for individual component tasks and items in order to help ensure it does indeed test its objectives at the correct levels. The eLADE exam is subjected to both classical (SPSS) and Rasch (WINSTEPS) analysis. As well as reliability data, these programs allow the collection of data about mean scores, discrimination,

population correlation statistics, item facility value, and standard measurement error, among others, providing detailed information with which to develop and revise exam material.

Test Reliability and Item Discrimination

Reliability is the extent to which a test measures what it is supposed to and the degree of confidence we have that there is no undue influence by elements outside the area of linguistic ability being measured, helping us to ensure that evaluation is stable and only affected by increases in linguistic ability. Reliability is measured using statistical software on a scale between -1 and $+1$, where $+1$ is the ideal result and indicates that candidates who repeat a test twice without having increased ability will obtain the exact same results. For acceptable reliability, a test should have a reliability of .7 or above. The overall reliability score for first eLADE exam administration was .909, with individual reliability of .883 and .752 for the listening and reading parts respectively.

Discrimination

Discrimination is the ability of an item to differentiate between candidates of higher and lower language proficiency. Good items are those which are only answered correctly by candidates of the correct level and are answered incorrectly by candidates of lower ability. However, whether it is due to bad or ambiguous item design, or for other reasons, this does not always occur and lower ability candidates may answer correctly while those candidates who do have the level actually get the item wrong. An item of this type is considered bad as it does not tell us which candidate has the higher ability level and should therefore be discarded. A good discrimination index in CTT ranges from .250 to 1, with 1 as the ideal (Green, 2013). To date, all items which contribute to the eLADE exam have had a discrimination index of between .250 and .643.

Tasks to be included in final version of the eLADE exam are chosen through a selection process in which the reliability, discrimination, standard error scores from the statistical analysis of the piloted tasks are used to decide which items are appropriate for creating a meaningful bi-level B1/B2 test which will provide enough information to extrapolate a fair grade. Of the ninety five listening and reading items first developed, forty-three (45.2 %) were discarded as not having the correct level of difficulty, not having optimal discrimination or because they did not contribute to the overall reliability of the test. Of the final fifty two items used for the administration, the listening part contained twenty-nine items and the reading part twenty three.

Cut Scores

In order to ensure that the exam and its component tasks and items do indeed assess B1 and B2 proficiency levels, Rasch analysis is employed to calculate individual item difficulty and standard error independently of the test population as a whole.

Rasch results help us make important decisions about the final exam format by providing information on item difficulty, measurement error, discrimination, the number of levels that can be assessed using a bank of items, test form (that is, its ability to evaluate its intended elements, which requires the elimination of those items which do not conform to the general test model). Together with expert judgement, Rasch analysis helps us to decide on cut scores for the levels to be tested.

Item level and relationship to the CEFR levels is established according to the Table 2 (North and Jones, 2009), where B1 items are considered to be those which have a logit value between -1.23 and 0.72 and B2 items have a logit value of between 0.72 and 2.80. Only items with an acceptable difficulty level and measurement error scores are selected for inclusion in the final exam.

Tab. 2: *Item level and relationship to the CEFR levels*

Level	Cut-off	Range on Logit scale
C2	3.90	
C1	2.80	1.10
B2+	1.74	1.06
B2	0.72	1.02
B1+	−0.26	0.98
B1	−1.23	0.97
A2+	−2.21	0.98
A2	−3.23	1.02
A1	−4.29	1.06
Tourist	−5.39	1.10

Correlation between the Different Parts of the Test

Once tasks have been designed with appropriate items, they are tested to see if each of the four parts of the exam do indeed serve to assess different types of abilities. Here, the correlation coefficient shows us the extent to which two parts of the test contain similar elements which overlap and which would therefore constitute a higher percentage of the material assessed in the exam. If the aim is to design tests which assess different abilities (e.g. listening assessment vs. speaking assessment), responding to these tasks should elicit the use of different cognitive features, strategies and knowledge sets. Correlation should ideally be low, between 0.4 and 0.6.

For the eLADE test, correlation between speaking and listening comprehension parts is 0.6, an indication that they do indeed measure different abilities. The same method

is also used to test the individual relationships between each and every other part of the exam.

Setting cut scores and grading

In the listening and reading comprehension parts of the test, cut scores (the number of items a candidate must answer correctly to pass the exam at each level) are established using the Angoff method (Cizek, 2011; Council of Europe, 2009) which consists of two rounds of assessment in which item difficulty level is decided by expert consensus. The question is which items a candidate performing at the minimum required level should be able to answer correctly. From each round of discussion, the judges' average response is taken as the proposed cut score. During this process, Rasch statistics are also taken into account to provide further perspective (see Figueras, 2011).

Once cut scores have been established, judges also take into account and discuss other variables such as pass rates, item facility values and measurement error and re--establish cut scores accordingly by consensus. As previously mentioned, the eLADE exam uses Rasch results as a further analysis tool in a third round of discussion to help determine individual item difficulty (Figueras, 2011).

The spoken and written production and interaction parts of the exam are evaluated using a holistic scale specially developed in accordance with CEFR criteria. The written production and interaction scale has the follow criteria: task fulfilment, coherence, range, and precision. The spoken production and interaction scale contains the following criteria: fluency, coherence, monologue, precision, range and interaction. Both speaking and writing tasks are double marked. In the spoken production and interaction test, one assessor acts as interlocutor. All assessors undergo periodic standard-setting sessions.

The eLADE exam does not compensate between skills and candidates must pass all four skills at the desired level in order to obtain accreditation. However, the test results detail the grades obtained for each specific skill as well as that of the final level achieved. Therefore, it is perfectly possible for a candidate to achieve a B1 overall while obtaining a higher mark in one or more particular skill. Individual skill grades may be carried over during the next two exam administrations, which allows candidates wanting to improve their result to retake the exam during this time and improve their overall result.

8 Conclusions

The University of Granada eLADE exam has grown beyond its initial intended purpose as a local mobility accreditation test to become an accreditation test for any person interested in certifying a CEFR language proficiency level in Spanish for either

personal or professional reasons. The exam has a simple, attractive online format, is easily and globally accessible and is validated according to internationally agreed criteria. All of these factors contribute to making the eLADE exam a useful, reliable and necessary tool for both the education community and the public at large.

Thanks to recent technological developments, we now have the opportunity to give anyone who wishes it the chance to obtain an L2 accreditation at CEFR B1 and B2 from anywhere in the world. This technology has the same potential to allow us to provide further accreditation at A1, A2, C1 and C2 proficiency levels. It is our belief that the future provision of valid, easy-access exams such as the eLADE test for all CEFR levels will be a major contribution to the promotion of the Spanish language at a global level.

References

ALDERSON J. CH., CLAPHAM, C., & WALL, D. (1995). *Language Test Construction and Evaluation*, Cambridge: Cambridge University Press.

ASSOCIATION OF LANGUAGE TESTERS IN EUROPE (ALTE) (2007) Principles of Good Practice for ALTE Examinations, [PDF], [online: 21/12/2007].

ALTE. (2001). www.alte.org/attachments/files/good_practice.pdf, 2001.

BACHMAN, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

BACHMAN, L. (2004). *Statistical Analyses for Language Assessment*. Cambridge: Cambridge University Press.

BACHMAN, L., & PALMER, A. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.

CIZEK, G., & BUNCH, M. (2007). *Standard Setting*. California: Sage Publications, Thousand Oaks.

COUNCIL OF EUROPE. (2001). *Common European Framework of Reference for Languages Learning, teaching, assessment*. Strasbourg: Council of Europe.

COUNCIL OF EUROPE. (2009). *Relating Language Examinations to the CEF, Manual Preliminary Pilot Version. Introduction and Feedback to the pilot phase*, [PDF]. Strasbourg: Council of Europe.

COUNCIL OF EUROPE. (2011). *Manual for Language test Development and Examining*, [PDF]. Strasbourg: Council of Europe.

EUROPEAN ASSOCIATION FOR LANGUAGE TESTING AND ASESSMENT (EALTA). (2004). *Directrices de EALTA para una buena práctica en el desarrollo de pruebas y la evaluación de lenguas*, [en línea 5/12/2007].

EALTA. http://www.ealta.eu.org/documents/archive/guidelines/Spanish.pdf.

FIGUERAS, N., KAFTANDJIEVA, F., & TAKALA, S. (2011). Relating a Reading Comprehension Test to the CEFR Levels: A Case of Standard Setting in Practice with Focus on Judges and Items. *The Canadian Modern Language Review/La Revue Canadienne des langues vivantes*, 69, 4, (November / novembre), 359–385

GREEN, R. (2013). *Statistical Analyses for Language Testers*. Palgrave McMillan.

INTERNATIONAL LANGUAGE TESTERS ASSOCIATION (ILTA). (2000). *Code of Ethics for ILTA*, [online: 2/6/2012], www.iltaonline.com/images/pdfs/ILTA_Code.pdf

MARTÍNEZ BAZTÁN, A. (2011). *La evaluación de lenguas: garantías y limitaciones*. Barcelona: Octaedro.

NORTH, B., & JONES, N. (2009). *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). Further Material on Maintaining Standards across Languages, Contexts and Administrations by exploiting Teacher Judgment and IRT Scaling*, [PDF], Strasbourg: Council of Europe. [online: 20/2/2014].

http://www.coe.int/t/dg4/education/elp/elp-reg/Source/Publications/Standards_language_EN.pdf

Bionote

Aurora Biedma Torrecillas, e-mail: xxx, Centro de Lenguas Modernas de la Universidad de Granada
She is a Doctor in Spanish Philology from the University of Granada. She has been a teacher of ELE and Literature at the CLM (U. of Granada) since 1993 and has also taught at the University of Heidelberg (Germany). She has been a member of the ACLES Accreditation Commission since 2013.

Adolfo Sanchez Cuadrado
He holds a BA in Translation and Interpretation from the University of Granada and he gained a Master's Degree from the University of Delaware (USA). He has been a teacher of ELE and Translation at the CLM (U. of Granada) since 2000. He is an accredited evaluator for the DELE exam at all levels.

Lola Chamorro Guerrero
She is a Doctor in Spanish Philology from the University of Granada. She has been a teacher of ELE at the CLM (U. of Granada) since 1993. She collaborated in the development of exams in Spanish for the European Certificate (1992–1999), financed by the EU Lingua programme.

Alfonso Martinez Baztan
He has a European Doctorate in Spanish Philology from the University of Granada, having done his thesis on the evaluation of languages. In 2010 his book *The evaluation of languages: guarantees and limitations* (Octaedro) was published. He has been a teacher of ELE since 1989, and has worked at the CLM (U. of Granada) since 1996.

Sonia Sanchez Molero
She graduated in Spanish Philology at the University of Granada and holds a Master's degree in Applied Linguistics in Teaching Spanish as a Foreign Language from the Antonio de Nebrija University. She is accredited as an examiner for the DELE exams at B1 and B2 levels.