Interpretation of the CEFR Companion Volume for developing rating scales in Cuban higher education

Claudia Harsch, University of Bremen

Ivonne de la Caridad Collada Peña, University of Informatics Sciences, Havana

Tamara Gutiérrez Baffil, University of Pinar del Río

Pedro Castro Álvarez, University of Informatics Sciences, Havana

Ioani García Fernández, University of Cienfuegos

Considering the need for improving English proficiency among Cuban university graduates, the Ministry of Higher Education (MES) implemented a new policy for teaching, learning and assessment of English proficiency. The policy adopted the CEFR (Council of Europe [CoE] 2001) as a proficiency framework, with the level B1 as the targeted attainment level. The CEFR needed to be adapted to suit the local context while operating within an internationally recognised framework. In 2017, the development of a valid and reliable proficiency exam was initiated. This work has been carried out by a network of Cuban teachers of English within the MES, coordinated by the University of Informatics Sciences, Havana, in collaboration with the University of Bremen, Germany. This article is a practice report of the process of developing rating scales for writing as part of the new exam. We explore the feasibility of using the CEFR Companion Volume (CEFR/CV; CoE 2018) descriptors as a basis for developing localised rating scales. Moreover, we describe the challenges faced during the process, which included creating more specific descriptors for the CEFR 'plus' levels (CoE 2001: 32, 181). Our insights show how the CEFR/CV (CoE 2018) descriptors can be adapted and how adaptation challenges can be overcome.

Keywords: rating scales, CEFR-based assessment, standardised testing, descriptor development, adaptation of the CEFR/CV for rating purposes

1 Introduction and background

The Ministry of Higher Education in Cuba (MES) introduced a national policy for English education in 2015 that considered the CEFR (CoE 2001) as the main proficiency framework, with B1 as the target level for university exit requirement. The attainment level (B1) was selected for a variety of reasons:

- the low proficiency level displayed by the majority of the new enrolments at university level who, in spite of the efforts made by Cuban general education, complete upper secondary education with poor English skills;
- the limited number of hours allotted to English in university undergraduate curricula, which makes it impossible to go beyond level B1 if "Below A1" is the starting point for many students; and finally,

- the fact that B1, as the "Threshold level" (CoE 2001: 34), in which language learners have acquired the beginning of an independence as users of the language, is the lowest level for university graduates to be able to start their professional lives with a possibility of continuing their training in English for academic and professional purposes through postgraduate education. This is considered a temporary phase since general education is also developing an improvement policy and will eventually upgrade the exit level of upper secondary schools.

One of the main issues when starting to implement the new policy was the lack of a proficiency exam available for certifying the exit requirement, given the impossibility of financial means to access international tests due to the budgetary and free nature of the education system in Cuba, which is subsidised by the state. In order to develop such an exam for Cuban higher education, a project was implemented in July 2017, the main goal of which is to develop a teaching and certification system for English so that Cuban language centres can reliably and validly certify students' English proficiency. The certification aims at international recognition through alignment to the CEFR (CoE 2001) proficiency levels. Partners in the endeavour are the MES, the University of Informatics Sciences (UCI) representing all Cuban universities, the University of Bremen, Germany, and the VLIR ICT for Development Network University Cooperation Program[1]. This way, we bring together local and global expertise to reflect local requirements while striving to adhere to international standards. The project included setting up a network of representatives of all Cuban universities (Cuban Language Assessment Network in Higher Education, abbreviated CLAN, which is part of LAALTA[2]).

The project encompasses three important objectives: first, developing assessment literacy among the CLAN teachers and preparing them for cascading this literacy in all universities; second, the development, validation, and implementation of the exam through a sustainable system; and third, research on assessment to support the first and second objectives.

So far, the first and second objectives have been addressed by means of six workshops, during which training and hands-on sessions for test development were provided, using the CEFR/CV (CoE 2018) as a framework for the CLAN members. In the workshops, all areas of assessment literacy in theory and practice have been covered. The members have had online working phases after each workshop, where they have collaboratively developed assessment materials and received feedback from each other and from the international trainers. The CLAN members have also been cascading their knowledge to other teachers in their institutions. Outcomes obtained so far include test specifications and item writer guidelines for the skills of listening, reading, speaking, and writing, along with the development of a bank of tasks for the four skills, as well as interlocutor guides for speaking. Based on this work, a small group of seven researchers (five of whom are the authors of this article) undertook the initial drafting of the rating scales for writing. In the next phases, speaking will be addressed, and the CLAN members will contribute to further refining the scales, following Harsch and Martin's (2012) development and validation approach, and Holzknecht et al.'s (2018) as well as Harsch and Seyferth's (2019) approach of involving teachers in developing tests.

This progress report describes the process of the initial rating scale development, focusing on the applicability of the CEFR/CV as a cornerstone, with a specific focus on the challenges faced and how we addressed them.

2 Rating scales development

Before outlining the actual development process, we will describe the basis, i.e. the assessment criteria and levels defined in the test specifications, as well as task characteristics that are relevant for the rating scale development.

---

1. The project was also supported financially by the British Council Cuba and UK, and ILTA.
2. LAALTA: Latin American Association of Language Testing and Assessment.

2.1 Basis: levels and criteria

The targeted level of the final exam is B1, as explained above; yet, in the first years, the exam should allow certifying students who can only demonstrate an A2 level. As part of the change management in the initial stage of the implementation of the policy, the Ministry decided to accept level A2 as exit requirement for a transitional period (2015-2021), until universities have been able to adjust to the new policy by creating all necessary human and material resources.

That is why the rating scale encompasses descriptors from A1+ to B1+. The decision to incorporate the so-called 'plus levels' in the scale is derived from the fact that the CEFR (CoE 2001) criterion levels (i.e. the main six levels) are too broad (Deygers and Van Gorp 2013: 4; Fulcher 2004: 258-259; Martyniuk and Noijons 2007: 6), even more so considering the narrow range of levels targeted in this project. We thus followed the CEFR's "branching approach" which suggests "cut[ting descriptors] into practical local levels" (CoE 2001: 32), i.e. adjusting the number of level subdivisions and hence the CEFR descriptors defining these sublevels to local needs. This way, we took into account the local context: teachers in Cuban higher education lack experience in working with analytic scales that span several levels. Accordingly, we introduced the plus levels A1+, A2+ and B1+ in order to provide more guidance and precision without making the scale too granular.

In the test specifications, the CLAN members defined the targeted skills, task characteristics, expected attributes of student performances, and an initial version of relevant assessment criteria. In order to decide which criteria to choose for rating written performances, the members considered the terms and concepts that teachers have traditionally used in Cuban teaching practice, in order to minimise the negative impact of change resistance amongst teachers when introducing the new system. The following criteria for assessing writing skills emerged:

- task fulfilment (TF, for interactive and productive tasks)
- coherence and cohesion (CC)
- vocabulary (VO, covering range and appropriateness)
- grammar (GR, covering range and accuracy)
- orthography (OR, covering spelling and mechanics).

We adapted the categorisations of the CEFR/CV to our local needs. With regard to the categories of interaction and production, for instance, we followed the CEFR/CV differentiation and developed productive and interactive writing tasks. Each exam includes one interactive and one productive task. These two aspects are also reflected in the rating scale category of task fulfilment, as will be explained in more detail below.

2.2 Methodology

The approach taken for development, validation and revision of the rating scales is an iterative one (Piccardo et al. 2019: 28), which was modelled on the research reported by Harsch and Martin (2012) and Harsch and Seyferth (2019). We are employing intuitive, qualitative and quantitative stages (CoE 2001; Fulcher, Davidson and Kemp 2011). Intuitive methods refer to approaches that "do not require any structured data collection, just the principled interpretation of experience", as the CEFR states (CoE 2001: 208).

We took existing descriptors, i.e. relevant descriptors from the CEFR/CV and from assessment scales in the context of CEFR-aligned exams, as a starting point. During the initial intuitive phase, a group of seven researchers selected existing descriptors for the targeted criteria/levels and then adapted formulations to avoid repetition or vagueness and to account for the local context (i.e. teaching styles, most common mistakes, as well as positive and negative transfer from native language).

The researchers tried out the initial scales with a few student samples, discussed reasons for digressions and revised the wording of the descriptors accordingly. In the next phase[3], the descriptors of the scale drafts will be qualitatively sorted into their targeted levels/criteria by the CLAN members. Then, the members will try the scales with student samples in a combined training and trial approach, in which qualitative and quantitative data will be analysed. Again, reasons for digressions will be discussed and descriptors adapted where necessary.

The focus of this contribution lies on the initial intuitive phase, as the main work with the CEFR/CV (CoE 2018), its adaptations and descriptor revisions took place during this phase.

2.3 Working with the CEFR/CV during the intuitive phase

The starting point for the rating scale development was the proficiency descriptors and the additional materials in the appendix of the CEFR/CV. Other scales consulted were the Aptis Speaking rating scale (O'Sullivan and Dunlea 2015), the IELTS speaking and writing band descriptors (IELTS 2013; IELTS 2016; and IELTS 2018) and the Pearson Global Scale of English Learning Objectives for Academic English (Pearson Education 2015). These scales were chosen because they have been widely valued and consulted by most of the faculty bodies in Cuban universities since the new policy was introduced. Appendix A shows the final draft of the rating scale (all sources are color-coded), with which we will go into training and validation with the CLAN members.

2.4 Compiling existing descriptors

In a first step, we considered the writing assessment grid in the CEFR/CV (CoE 2018: 173-174), which includes the following categories: Overall, Range, Coherence, Accuracy, Description and Argument. This categorisation, however, does not match our assessment criteria (see above). Hence, we selected relevant descriptors from the grid but placed them into the best fitting criterion in our assessment criteria system. As we do not use an Overall criterion in our analytic approach, we dropped this category. Instead, we focused on the criterion TF with a close reference to our test specifications and task demands; here, we mostly added our own descriptors regarding the message conveyed, the relevance of ideas, the language functions performed and genre requirements, as well as register and politeness conventions. The CEFR/CV scale on socio-linguistic appropriateness (CoE 2018: 138) contains some descriptors that we included (see appendix A, phrases in red); we also selected some of the IELTS (IELTS 2013 and 2016) descriptors (phrases in green in Appendix A). We dropped the CEFR/CV assessment grid categories of Description and Argument (CoE 2018: 173) since their content is already included in our TF criterion. Furthermore, we consulted the CEFR/CV scales on productive and interactive writing (CoE 2018: 75-80; 93-102); while they had provided helpful input for the test specifications, we found their descriptors too generic and abstract to be directly used in the rating scale.

We used the CEFR/CV's assessment grid category of Coherence (CoE 2018: 173) but inserted descriptors from the CEFR/CV scale Coherence and Cohesion (2018: 142) as well as our own additions regarding organisation, sequencing and topic progression. With regard to the CEFR/CV grids' categories of Range and Accuracy (CoE 2018: 173), we followed the local tradition in Cuba, i.e. treating them as sub-aspects of the wider categories of grammar and vocabulary, which was also laid down in the test specifications. Hence, we arranged the aspects of linguistic range and accuracy under our criteria Vocabulary (VO) and Grammar (GR). In VO, the term *accuracy* was replaced by *appropriateness*, to account for terminological use in the Cuban context, i.e. teachers here would regard students' vocabulary choice as a matter of socio-linguistic appropriateness rather than accuracy, which is strongly associated with grammar. For VO and GR, we also used the CEFR/CV scales Vocabulary Range, Grammatical Accuracy and Vocabulary

---

3. This work actually took place in a workshop in February 2020, just after the deadline for this article. We will publish the results elsewhere.

Control (2018: 132-134), as well as the occasional IELTS descriptor wording (IELTS 2013 and 2016). For our criterion Orthography (OR), we used the CEFR/CV scale Orthographic Control (2018: 137) and some IELTS descriptor wordings (IELTS 2013 and 2016). For all our criteria, we added statements on how to treat errors (in *italics*); following Harsch and Martin's insights (2012). These statements are intended to further guide the raters, because teachers in Cuba are traditionally used to focusing on error correction.

One of the challenges we found was adapting existing descriptors to the local context (see Appendix A where we color-coded all the different sources as well as the adaptations we undertook). Another major challenge was to describe the plus levels, as the CEFR/CV scales do not consistently provide them. Thus, we had to compare the existing descriptors of the CEFR criterion levels and formulate descriptors that would enable enough differentiation between them. We will discuss below (in Section 3) a detailed example of these challenges and how we overcame them.

2.5 Pre-trial

In the informal pre-trial, the researchers/authors used the initial rating scale drafts for the analysis of three student performances, each for an interactive and a productive writing task. The performances were elicited informally in the classroom by one of the researchers, who is also an active language teacher. The aim of the pre-trial was to evaluate the usability of the descriptors: they were evaluated for "clarity, [context-related] pedagogical usefulness" (North and Docherty 2016: 25), possibilities for constructive alignment and practicality, as well as consistency across the levels and the assessment criteria. In the pre-trial, we compared students' performances with the descriptors in the rating scale (Pollitt and Murray 1996) to place performances at levels, and we qualitatively discussed digressions and underlying reasons; i.e. we each gave explanations of our decisions, justified reasons why we placed a performance at a certain level and exchanged our justifications. After careful considerations of the different viewpoints, and careful re-analysis of student performance and descriptor wording, we revised the descriptors where necessary. Appendix A shows these revisions in blue; all deletions indicated in Appendix A also took place after this pre-trial. Most revisions happened in the criteria OR and GR, some in CC, and a few in VO.

We will use this draft of our rating scale for the next qualitative phase (see Section 4 below).

3 Discussion of the challenges with the CEFR/CV

We will now summarise the main challenges we faced and how we dealt with them when using the CEFR/CV and its proficiency scales/descriptors for developing rating scales.

Abundance of scales at different places: We found the fact that the CEFR/CV contains a wealth of scales for the productive/interactive skills, strategies and linguistic competences that may be quite overwhelming. This was exacerbated by the challenge of locating relevant scales (including the writing assessment grid in the Appendix) at different places in the CEFR/CV during the actual work with the CEFR/CV[4]. Appendix B gives an overview of the scales we consulted and their location in the CEFR/CV. Even when simultaneously working on several laptops, it was a constant search for relevant descriptors and scales. Here we would recommend a searchable online data bank of all CEFR/CV descriptors, where relevant ones could be compiled (along with a transparent source reference) to facilitate working with the CEFR/CV.

Different categorisations: As described above, the categorisations in the Writing Assessment Grid and other CEFR/CV scales differed from our assessment criteria. Moreover, the CEFR/CV's assessment grid categorisation also differs from the CEFR/CV scale system: the Assessment Grid differentiates range, coherence, accuracy, description and argument, while the CV scale system shows a much wider differentiation of language activities (written production, of which description and argument are sub-aspects, and written interaction, as well as strategies) and linguistic competencies (which subsume

---

4. One has to bear in mind that it is difficult in Cuba to print such large documents as the CEFR/CV.

range, accuracy and coherence, amongst many more aspects which are not covered in the Assessment Grid). This may be a natural phenomenon given the complexity of the construct of communicative competence, yet it does pose a challenge when the task is to compile relevant descriptors for a given set of writing assessment criteria.

Plus levels not always provided: Not all CEFR/CV scales consistently describe the plus levels. It proved difficult to develop suitable descriptors for these levels. We will provide an example in the next paragraph. It would also help to analyse actual student performances to fill the plus levels appropriately. We are planning to address this issue in the next step when we have a solid basis of student performances.

Inconsistent wording across scales and/or across levels: We found that some scales/materials (at different places) in the CEFR/CV address similar aspects but use different wording in descriptors that target the same level. Some descriptors (that appear in different scales) contain aspects that seem incoherent when comparing these aspects across different scales and levels. It was challenging to reach consistent interpretations of a given aspect (such as the nature and impact of errors) within one level and across the levels when comparing different scales (e.g., Grammatical Accuracy, Vocabulary Control, and the Writing Assessment Grid: Accuracy). Compare the following examples:

- CEFR/CV scale Grammatical Accuracy (2018: 133) states for level A2: "… still systematically makes basic mistakes …; nevertheless, it is **clear** what he/she is trying to say"; for level B1+: "Errors occur, but it is clear what he/she is trying to express." This aspect is not mentioned at B1, and there is no A2+ descriptor.
- CEFR/CV scale Vocabulary Control (2018: 134): no mention of the aspect of clarity of expression.
- CEFR/CV Writing Assessment Grid (2018: 174), criterion Accuracy, level A2: "… errors may sometimes cause **misunderstandings**"; level B1: "Occasionally makes errors that the reader usually can interpret correctly on the basis of the context." No plus levels are defined.

When comparing these statements, we found the aspects in bold (describing A2) contradictory (i.e. when there is a misunderstanding, it is *not* clear what one is trying to say). Moreover, we regarded the demand for clarity of what one wants to say too high for A2. When working on the target level A2+, we found it unfortunate that there are no A2+ descriptors in these scales. Our resolution was to make use of the IELTS (IELTS 2013 and 2016) band 4 descriptor[5]: "errors may cause strain on the reader" (IELTS 2013). We added this qualification at A2+ for our criteria VO, GR, and OR after it became clear in the pre-trial that we needed to qualify the kinds of errors we would expect and 'allow' at the different levels (for example, there are minor, non-impeding errors that are 'allowed' at B1+, while we would not expect systematic errors in basic sentence structures at this level; see the blue additions in Appendix A).

These issues were the main challenges we faced when working with the CEFR/CV (CoE 2018). In order to address these challenges, we resorted to different means, which can be summed up as follows:

- Reorganising CEFR/CV descriptors into the local assessment criteria.
- Adapting CEFR/CV descriptors (i.e. changing wording) to make levels coherent.
- Adding descriptors from other sources, particularly for the plus levels.
- Adding and adapting descriptors to account for the local context, both for criterion levels and plus levels.

4 Conclusions and outlook

Undoubtedly, the CEFR/CV provides a rich and informative source and starting point for rating scale development. Yet, one has to take into account the complexity of the CEFR/CV, its limitations and the

---

5. IELTS band 4 is actually targeting B1, which again seems in contradiction to the CEFR/CV descriptors on clarity of expression in the Grammatical Accuracy scale.

requirements of the local context. Each descriptor in the CEFR/CV has to be checked against the local test specifications (i.e. to see whether its content matches the content of the test specifications) and adapted accordingly to fit the local requirements. When adapting descriptors or writing additional ones, it is important to consult local experts and to take additional sources into consideration, such as assessment scales from other exams that are aligned to the CEFR. Particularly when the local rating scales require a finer granularity than the CEFR criterion levels, measures need to be taken to fill the plus levels with appropriate descriptors.

Based on our experiences with the CEFR/CV descriptors, we found it challenging to deal with the abundance of scales in the CEFR/CV, with differing categorisations across the CEFR/CV, with inconsistent wording within and across scales and levels, and with the fact that plus levels are not always provided. In order to overcome these challenges and to account for the local context, we reorganised CEFR/CV descriptors into our local assessment criteria, adapted CEFR/descriptors for more coherence, and added descriptors from other sources for the plus levels.

Any rating scale development is an iterative process with several rounds of revisions. It is advisable to use different methods to gain information on the validity and applicability of the new scale. In our case, we have covered the initial intuitive phase, using experts to compile, draft and trial the first version, leading to the first round of revisions. With the thus revised rating scales, we are entering the next phase, which includes a qualitative sorting exercise, i.e., the CLAN members will sort the descriptors into levels/criteria in order to validate the content and levels of the rating scales. Then, a benchmarking exercise will follow where the CLAN members will be trained to use the scales so that they can pass on this knowledge to their colleagues and roll out the new assessment approach at a national level.

5 References

Council of Europe. 2001. The Common European Framework of Reference for Languages: Learning, teaching, assessment. Cambridge: Cambridge University Press.

Council of Europe. 2018. *Common European Framework of Reference for Languages: Learning, teaching, assessment. Companion volume with new descriptors*. Strasbourg: Language Policy Division. https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989 (accessed 27 November 2019).

Deygers, Bart & Koen Van Gorp. 2013. The influence of the CEFR on rating scale design. In: Jozef Colpaert, Mathea Simons, Ann Aerts, Margret Oberhofer (eds.) *Book of Abstracts, "Language Testing in Europe: Time for a New Framework?"*, Antwerp: University of Antwerp., 57-59. https://www.academia.edu/3549147/The_influence_of_the_CEFR_on_rating_scale_design (accessed 12 February 2020).

Fulcher, Glenn. 2004. Deluded by artifices? The common European framework and harmonization. *Language Assessment Quarterly 1*(4), 253-266.

Fulcher, Glenn, Fred Davidson & Jenny Kemp. 2011. Effective rating scale development for speaking tests: Performance decision trees. *Language Testing 28* (1), 5-29.

Harsch, Claudia & Guido Martin. 2012: Adapting CEF-descriptors for rating purposes: Validation by a combined rater training and scale revision approach. In: *Assessing Writing* 17, 228–250.

Harsch, Claudia & Sibylle Seyferth. 2019: Marrying achievement with proficiency - developing and validating a local CEFR-based writing checklist. *Assessing Writing*. Online first: https://www.sciencedirect.com/science/article/pii/S1075293519301746 (accessed 20 March 2020).

Holzknecht, Franz, Benjamin Kremmel, Carmen Konzett, Kathrin Eberharter, Eva Konrad & Carol Spöttl. 2018. Potentials and challenges of teacher involvement in rating scale design for high-stakes exams. In Xerri, D., & Vella Briffa, P. (eds.), *Teacher involvement in high stakes language testing* (pp. 47-66). Springer.

IELTS. 2013. *IELTS TASK 1 Writing band descriptors* (public version). https://takeielts.britishcouncil.org/sites/default/files/ielts_task_1_writing_band_descriptors.pdf (accessed 12 February 2020).

IELTS 2016. *IELTS TASK 2 Writing band descriptors* (public version). https://takeielts.britishcouncil.org/sites/default/files/ielts_task_2_writing_band_descriptors.pdf (accessed 12 February 2020).

IELTS 2018. *IELTS Speaking: band descriptors* (public version). https://www.ielts.org/-/media/pdfs/speaking-band-descriptors.ashx?la=en (accessed 12 February 2020).

Martyniuk, Waldemar & José Noijons. 2007. *Executive summary of results of a survey on the use of the CEFR at national level in the Council of Europe Member States*. Strasbourg: Council of Europe.

North, Brian and Coreen Docherty. 2016. Validating a set of CEFR illustrative descriptors for mediation. *Research Notes*, *63*, 24-30.

O'Sullivan, Barry & Jamie Dunlea. 2015. *Aptis General technical manual version 1.0*. Technical Report TR/2015/005. British Council: London.

Pearson Education. 2015. *Global Scale of English Learning Objectives for Academic English*. http://www.b-li.ir/ar/06CEFR/4.GSE_LO_Academic_English.pdf (accessed 12 February 2020).

Piccardo, Enrica, Brian North & Tim Goodier. 2019. Broadening the scope of language education: Mediation, plurilingualism, and collaborative learning: The CEFR companion volume. *Journal of e-Learning and Knowledge Society*, *15*(1).

Pollitt, Alastair & Neil Murray. 1996. What raters really pay attention to. *Studies in language testing*, *3*, 74-9.

6 Biographies

Claudia Harsch is a full professor at the University of Bremen, specialising in language learning, teaching and assessment. She is also the Director of the Languages Centre of the universities in the Land Bremen. She has worked in Germany and in the UK and is active in teacher training worldwide. Her research interests focus on areas such as language assessment, language and migration, intercultural communication and the implementation of the CEFR. Claudia was president of the European Association of Language Testing and Assessment from 2016-2019.

Ivonne de la Caridad Collada Peña has been a teacher and teacher trainer for more than 30 years and has a master's degree in management. She has been the head of English departments, faculty dean, and is currently the director of the Language Centre at UCI. She has taught courses on linguistics, lexicology, ESP and EFL, among other subjects. Her research interests include TESOL methods, curriculum development, lexicology, grammar and language teaching and assessment. She is currently the Cuban coordinator of the project that is presented in this contribution.

Tamara Gutiérrez Baffil has been teaching for 22 years in higher education. She has a master Degree in Educational Science. Associate Professor and Language Centre Director. Currently researching on Language Centre Management as PhD dissertation and other searches in CEFR application in the Cuban context. Granted an International Diploma in Educational Planning and Administration and a given a Professional award in L2 teacher training. She has published in several journals. Teaches ESP, EAP, EPP. Participates in international and national scientific projects.

Pedro Castro Álvarez is a full professor at the University of Informatics Sciences, Havana, Cuba, where he has been deputy director of the Language Center. He has a doctorate in pedagogy from the Universidad Central de Las Villas in 2006 and a master's degree from the same institution in 2000. The professor has a 28-year experience as a teacher in Cuban higher education and has also taught in foreign universities. His research is related to the fields of educational technologies, curricular development and language assessment. He has published several journal articles and attended multiple conferences in Cuba and abroad.

Ioani García Fernández has been teaching for 18 years, 16 in Higher Education. She is an associate

professor and has a Master's Degree in Education from the University of Cienfuegos. She was certified with the Teaching Knowledge Test (TKT) and also received a Professional Award in L2 Teacher Training. She has participated in several international scientific events and published various articles related to the teaching of English as a Foreign Language. She is currently working as a professor at the Language Centre of the University of Cienfuegos.

Appendix A
Initial draft rating scale writing, after first trial

	Task Fulfilment	Coherence/cohesion	Vocabulary (range and appropriateness)	Grammar (range and accuracy)	Orthography (spelling and mechanics)
B1+	The message is clearly and appropriately conveyed. (CLAN) All ideas/content are relevant to the topic of the task (CLAN) Performs all the language functions required by the task (e.g., comparing, describing, explaining, justifying etc.) (Test specs page 8 and adapted from CV page 138). ~~Shows the required length.~~ Follows the conventions of the text type required by the task (CLAN). Uses an appropriate register (adapted from CV page 138) Shows salient politeness conventions (adapted from CV 138)	Uses a meaningful sequence of linked ideas, with adequate topic progression (TS, GE). Makes logical paragraph breaks, if required by task. (adapted CV p. 142) Uses various cohesive devices to establish cohesion throughout the text. (CLAN) Establishes more complex relations between ideas, e.g., Can introduce a counter-argument with 'however', cause and consequence, cause and effect (adapted form CV p. 142).	Uses a good range of topic-specific vocabulary related to the task (CV p 132-174). Uses vocabulary with reasonable precision. (adapted from CV page131) *May show occasional inaccurate word choices and collocations (adapted from IELTS band 7 and 8).* *Errors may occur when expressing more complex thoughts. (adapted CV 134)*	Uses a good range of simple structures and features with generally good control *though mother tongue influence* may be *noticeable.* Shows some complex grammatical features and syntactical structures, although not always correctly. *Errors may occur, but it is clear what he/she is trying to express* (CV p 133).	Spelling is accurate enough to not strain the reader. Punctuation generally follows conventions. *Spelling and punctuation may show mother tongue influence.* (adapted from CV 137).

	Task Fulfilment	Coherence/cohesion	Vocabulary (range and appropriateness)	Grammar (range and accuracy)	Orthography (spelling and mechanics)
B1	The message is generally clearly conveyed. (CLAN) The ideas/content are generally relevant to the topic of the task. (CLAN) Performs most of the language functions required by the task (e.g., comparing, describing, explaining, etc.) (Test specs page 8 and adapted from CV page 138). ~~Shows the required length.~~ Mostly follows the conventions of the text type/format required by the task (CLAN), *but the format may be inappropriate in places* (IELTS band 5). Shows awareness of the required register, *but may still be inconsistent in tone* (IELTS band 6). Generally follows salient politeness conventions, but not always appropriately (adapted from CV 138)	Mostly organizes ideas into a meaningful sequence, with adequate topic progression (TS, GE). ~~*May occasionally use unrelated or off-topic ideas (CLAN).*~~ Makes simple, logical paragraph breaks if required by task. (adapted CV p. 142) Links a series of shorter, discrete simple elements into a connected, linear sequence of points by using a limited number of cohesive devices (adapted CV p. 142)	Uses sufficient topic-specific vocabulary to express themselves on familiar topics. (CV page 132) Shows appropriate use of a wide range of ~~simple~~ basic, frequent vocabulary. (adapted from CV page 134) *Major errors may still occur when expressing more complex thoughts.* (CV page 134) *May use circumlocution and occasionally unclear expressions.* (adapted from CV page 131, 174)	Uses a range of simple grammatical features and sentence structures with reasonable accuracy. (adapted CV p. 133) Attempts a limited range of complex sentence structures or complex grammatical features, *though they may usually be incorrect.* (adapted from IELTS band 5) *In general, the reader can interpret the errors correctly based on the context.* (adapted from CV p. 174)	Produces generally intelligible spelling for most common words, *mother tongue influence is likely with less common words.* ~~Spelling,~~ Punctuation is ~~and layout are~~ accurate enough to be followed most of the time, but *mother tongue is likely to influence punctuation.* (adapted from CV p. 137)
A2+	The message gets across but with some limitations. In general, the ideas/content are related to the topic of the task. (CLAN) Performs basic language functions required by the task (e.g., describing, explaining, narrating); *may attempt the more complex ones, but not always successfully (e.g., comparing/contrasting ideas)* (Test specs p. 8 and adapted from CV p. 138). *May use an inappropriate format* (adapted from IELTS band 4). *May use an inappropriate tone (adapted from IELTS Band 4).*	Shows some organization of ideas and a clear attempt at topic progression (TS). *May still show some limitations in sequencing and text structure.* ~~*also off-topic ideas (CLAN)*~~ *Paragraph breaks may be missing.* Uses the most frequently occurring connectors to link simple sentences in order to tell a story or describe something as a simple list of points (CV p 142). *May use less frequent cohesive devices inappropriately. (CLAN)*	Uses basic, frequent vocabulary to express themselves in routine everyday situations (CV p. 132). Shows inaccuracies in word choice and collocation *that may occasionally cause strain for the reader.* (CLAN and adapted from *IELTS*) May *have to compromise the message and may use repetitions and circumlocutions (adapted from CV* 131 and CLAN).	Uses simple sentence structures and basic grammatical features (such as present perfect, continuous forms, modals). Systematic mistakes may still occur; *errors may sometimes cause strain on the reader* (adapted from IELTS Band 4), *but it is usually clear what s/he is trying to say.* (adapted from CV p. 133, 174). *May show attempts at more complex structures, but usually these are erroneous.*	Writes with reasonable phonetic accuracy, *but mother tongue is likely to be noticeable.* Punctuation is still *likely to be influenced by mother tongue.* (adapted from CV p. 137). *Errors may cause occasional strain on the reader. (adapted from IELTS band 4)*

	Task Fulfilment	Coherence/cohesion	Vocabulary (range and appropriateness)	Grammar (range and accuracy)	Orthography (spelling and mechanics)
A2	The message gets across but with some strain on the reader. The ideas/content are *not necessarily all* related to the topic of the task. (CLAN) Performs the more concrete language functions required by the task (e.g., social exchanges, invitations etc.). (Test specs p. 8). *Generally, the format may not yet be appropriate* (adapted from IELTS band 4). Apart from everyday polite forms of greeting and address, the tone *may be inappropriate* (adapted from CV page 138 and IELTS band 4).	~~Makes an attempt at organization and topic progression~~ (TS). Produces a list of points that are mostly in a logical sequence; not all are necessarily connected. *May show limitations in sequencing and text structure, ~~also off-topic ideas~~ (CLAN)* Links groups of words with simple connectors like 'and,' 'but' and 'because' (CV p 142). *May overuse connectors, may use other cohesive devices unsuccessfully. (CLAN)*	Shows ~~sufficient~~ limited basic vocabulary and memorized phrases to express basic communicative needs and to communicate limited information (adapted from CV p. 132 and 174). Shows frequent inaccuracies in word choice and collocation *that may cause strain for the reader.* (CLAN and adapted from IELTS)	Shows simple sentence structures, with memorized grammatical phrases and formulae. Still systematically makes basic grammar and syntax mistakes – for example tends to mix up tenses and forget to mark agreement, *which the reader may misunderstand* (adapted from CV p. 133, 174).	Writes with reasonable phonetic accuracy the most common words, *but not necessarily following standard spelling.* (adapted from CV. p. 137) Uses punctuation such as full stop, commas, question marks, *but not necessarily accurately.* *Errors in spelling and punctuation may cause strain for the reader. (adapted from IELTS band 5)*
A1+	The message only partly gets across and usually requires a sympathetic reader. (CLAN) Shows awareness of the required topic but the ideas are very limited. (CLAN) Performs only the most concrete language functions (e.g., establish social contact) (CLAN, adapted CV 138) Format and tone are mostly inappropriate. (CLAN)	Links words or groups of words with very basic linear connectors like 'and' or '~~then~~ because' (CV p. 142). Texts longer than short notes and messages generally show coherence problems that make them very hard or impossible to understand. (adapted from CV p. 174).	Shows a very basic range of simple vocabulary and memorized expressions related to particular concrete situations (CV p. 131-132) *May overuse certain words (CLAN)*	Shows only a few simple grammatical features and sentence patterns in a learnt repertoire (CV p. 133). *Errors are likely to be frequent and common.* (CLAN)	Writes only familiar words and short phrases used regularly with reasonable accuracy. Spells his/her address, nationality and other personal details correctly. Uses only basic punctuation (full stops and question marks (adapted from CV. p. 137)

Notes: Sources used by colour code:

CEFR Companion volume/relevant scales and level | *IELTS band descriptors* | own additions CLAN and test specs | revisions after first trial in small group

Appendix B
Overview of relevant scales in CEFR/CV (CoE 2018)

Writing	Number of scales	Pages in the CEFR/CV
Production activities and strategies	5	75-80
Interaction activities and strategies	7	93-102
Communicative language competences	12	133-143
Appendix 4: Written assessment grid	2	173-174