_____

Assessing writing in English in Cuban Higher Education

Evaluación de la escritura en inglés en la Educación Superior Cubana

Dr. C. Tamara Gutiérrez Baffil[1], http://orcid.org/0000-0001-8303-5206

M. Sc. Ivonne de la Caridad Collada Peña[2], https://orcid.org/0000-0001-9367-3586

[1] Universidad de Pinar del Río, Pinar del Río, Cuba

[2] Universidad de las Ciencias Informáticas, La Habana, Cuba

tamara@upr.edu.cu
yvonne@uci.cu

Abstract

Objective: This article aims at discussing the elements studied in the process of developing rating scales for writing according to the test specifications defined to assess it at levels A2 and B1 of the CEFR (since these are the two levels targeted by the test).
Methods: Analysis and synthesis were used as theoretical methods, and an iterative approach was followed for the development, validation, and revision of rating scales.
Result: Test specifications for assessing the writing skills in a national standardized test were obtained as well as item writer guidelines, which include all the necessary orientations for task developers to have consistency and uniformity, and the rating scales.
Conclusion: The rating scales developed for writing assessment become a valuable tool for constructive alignment between curriculum development, instruction, classroom assessment, and national proficiency testing. Based on some of the most internationally recognized descriptors and scales, they respond to the higher education local needs and expectations to describe in a standardized qualitative way the observed student performances.

Keywords: foreign language instruction, speaking, language skills, skills assessment, achievement rating

Resumen

Objetivo: El objetivo de este trabajo es exponer los elementos estudiados en el proceso de desarrollo de las escalas de evaluación de la escritura según las especificaciones de la prueba definidas para evaluarla en los niveles A2 y B1 del MCER (por ser los dos niveles a los que se dirige la prueba).
Métodos: Se utilizaron el análisis y la síntesis como métodos teóricos, y se siguió un enfoque iterativo para el desarrollo, la validación y la revisión de las escalas de calificación.
Resultado: La creación de las especificaciones de las pruebas para evaluar las habilidades de escritura en una prueba nacional estandarizada, las directrices para los redactores de tareas con todas las orientaciones necesarias para que estas tengan coherencia y

_____

_____

uniformidad, y las escalas de calificación.

Conclusión: Las escalas de calificación desarrolladas para la evaluación de la escritura constituyen una valiosa herramienta para lograr una alineación constructiva entre el desarrollo del plan de estudios, la enseñanza, la evaluación en el aula y la prueba nacional de aptitud. Basadas en algunos de los descriptores y escalas más reconocidas a nivel internacional, estas responden a las necesidades y expectativas locales de la educación superior para describir de forma cualitativa estandarizada los rendimientos observados de los estudiantes.

Palabras clave: enseñanza de las lenguas extranjeras, habilidades orales, evaluación de habilidades.

Introduction

The policy for improving the teaching process of English in Cuban Higher Education (MES) arises from the need to achieve a competent professional in English at level B1 of the Common European Framework of Reference for Languages (CEFR).

Studies conducted on this teaching process from 2011 to 2014 showed that the competence level of university graduates does not meet the expectations and needs of the society although the Higher Education Ministry has implemented teaching strategies in this language, as well as approaches and methods from the most traditional to the most up-to-date ones. For these reasons, the MES has promoted a paradigmatic change in English language teaching with the policy that conceives English as an exit requirement, bringing about changes in curriculum, teaching, and assessment practices.

At the beginning of the implementation of the new policy, one of the main problems identified was the non-existence of a standardized test to certify the exit requirement due to the impossibility of having the financial means to access international tests due to the budgetary and free nature of the Cuban educational system, which is subsidized by the state.

Therefore, in July 2017, an innovative project began to be implemented with the main objective of developing a system for teaching and certifying English, so the country's language centers could reliably and validly certify the students' English proficiency by developing such an exam for Cuban higher education.

To this end, and due to the growing importance of the Common European Framework of Reference for Languages (Council of Europe, 2001, 2018), the MES assumes it as a framework of competence aligning itself to internationally recognized reference frameworks.

_____

The CEFR describes learner proficiency in foreign languages on six ascending levels of proficiency[1] for a range of different aspects of communicative competence.

Since its publication in 2001, the CEFR has been applied in curriculum reforms in all European education systems as well as in many countries around the world. It is important to highlight what in this respect the Council of Europe makes clear: "the framework provides a common basis for the development of language syllabuses, curriculum guidelines, examinations, textbooks, etc." (Council of Europe, 2001, p. 1), that is to say, it is a global framework and allows adaptation to local contexts, becoming one leading framework also for higher education in the world and all major proficiency tests and certificates.

The above-mentioned project was undertaken by a group of 40 teachers of English from all universities in Cuba, the Cuban Language Assessment Network (CLAN) with the guidance of Prof. Claudia Harsch, from the University of Bremen, and financial support from that German University, MES, the University of Informatics Sciences (UCI), the VLIR ICT for Development Network University Cooperation Program, the British Council Cuba and UK, and the International Language Testing Association (ILTA).

This article aims at discussing the elements studied in the process of developing rating scales for writing according to the test specifications defined to assess it at levels A2 and B1 of the CEFR (since these are the two levels targeted by the test). In the process some results of developing test specifications to assess these skills and the process of developing rating scales for assessing writing in the Cuban tertiary education system are described.

Methods

To conduct the study, theoretical methods were used such as analysis-synthesis to study the theory and practice behind language assessment as a process in language teaching and learning, particularly for writing skills. Expert training, consultation, and joint elaboration were used in eight workshops with CLAN[2] members to obtain reliable results for the context.

For the following stage (development, validation, and revision of rating scales) the approach taken is iterative (Piccardo, North, & Goodier, 2019, p. 28), modeled on the study reported by Harsch & Martin (2012) and Harsch & Seyferth (2019), intuitive, qualitative and quantitative stages (Council of Europe, 2001); Fulcher, Davidson & Kemp 2011) were employed.

Initial results of the project which consist in assessment literacy in eight workshops, development of test specifications for the four skills in the national standardized exam, item writer guidelines, task development (in a process of development, feedback, and revision individually, by region,

_____

[1] The six main levels are labeled A1, A2 (beginners), B1, B2 (independent users), and C1, C2 (competent users). Each of these levels can be further differentiated in a so-called plus level, e.g. B1+, indicating a level between B1 and B2.
[2] Cuban Language Assessment Network in Higher Education.

_____

and collectively), among other outcomes.

The starting point for the selection and adaptation of descriptors for test specifications at the initial intuitive phase was the analysis of the existing descriptors of the CEFR/CV (Council of Europe, 2018). A decision was made for which criteria to consider in the scales. Later, the descriptors were reformulated considering the local context characteristics such as teaching styles, common errors, as well as positive and negative transfer from the mother tongue to avoid repetition or vagueness. The writing tests were then designed with these features in mind.

A pre-trial followed, and then a group of six researchers drafted a first version of the rating scales, taking into consideration the above-mentioned test specifications and other assessment scales in the context of the CEFR-aligned examinations.

Another pre-trial followed. A sample of thirty students from the University of Pinar del Río was selected to do the writing test.

Then, the CLAN group developed an online workshop (due to COVID-19 constraints). In this new session a scale sorting exercise was developed, to validate the accuracy of descriptor wording. In addition, three new samples were thoroughly analyzed for consensus building and benchmarking.

Results and Discussion

As a result of the project, the test specifications were designed for assessing the writing skill in a national standardized test, the item writer guidelines were created including all the necessary orientations for task developers to have consistency and uniformity, and the rating scales were developed to place the students' behavior at a given level.

The test specifications for assessing the writing skill in a national standardized test broadly included two tasks: one interactive and one productive both eliciting one and more of the language functions in the test specifications. The interactive includes letters, emails (to respond to a specific person and an initial text), and letter of application (responding to a job advert). The productive consists of writing to a general audience, without having to respond to one specific recipient (reports, descriptions, essays, brochures, narratives, notes, etc., posts, blogs, etc.).

The construct is aligned to Cuban learning and teaching objectives, as expressed in local curriculum, and defined in terms of targeted learning outcomes, describing the language subskills required to meet the expected outcome successfully (e.g. describe familiar objects and places, people and their routines, hobbies, and activities, everyday processes, health conditions) very basic events in the past, using simple connectors.

The topics areas to be covered comprise mostly general, professional, or academic topics accessible to a general audience with a concrete to slightly abstract nature, avoiding controversial or distressing topics that could affect students' performance in an exam situation.

_____

Authenticity and reliability in the writing test are predicted to take place in social, academic, and professional scenarios both in Cuba and abroad, which includes interactions with non-native and native English speakers. The prompts are designed taking into account sources, topics, nature of the content, length, and can be presented as pictures, hints/suggestions (in key words), simple graphs, charts, tables, simple letters, or emails (below level, 80-100 words maximum).

The discourse types comprise narrative, descriptive, instructive, expository, simple argumentative texts, with a length from 100 to 130 words per task and the time for writing total (for two tasks), 45 minutes for each task, about 20 minutes.

The item writer guidelines include all the necessary orientations for task developers to have consistency and uniformity, and the rating scales developed.

Although the target level of the final exam is B1, the exam should allow students who can only demonstrate an A2 level in the first years to be certified. For this reason, in the initial phase of the policy implementation, the Ministry decided to accept level A2 as an exit requirement for a temporary period (2015-2021), until universities can adapt to the new policy by creating all the necessary human and material resources. Therefore, the rating scales established levels covered from A1+ to B1+.

As can be seen, incorporating the so-called 'plus levels' in the scales is because the CEFR (Council of Europe, 2001) criterion levels (the six main levels) are too broad (Deygers & Van Gorp, 2013, p. 4; Fulcher, 2004, pp. 258-259; Martyniuk & Noijons, 2007, p. 6), and for the project's purposes, a narrow range of levels is necessary. Therefore, the "branching approach" suggested by the CEFR was followed to "cut descriptors down to practical local levels" (Council of Europe, 2001, p. 32), i.e. to adjust the number of level subdivisions and hence the CEFR descriptors defining these sub-levels to local needs.

The CLAN was in charge of defining the target competencies, task characteristics, expected attributes of student performances, and an initial version of the relevant assessment criteria in the test specifications. They also considered the terms and concepts that have traditionally been used in Cuban teaching practice when deciding on the criteria to be chosen for marking written performances, which minimize the negative impact of teachers' resistance to change when introducing the new system. As a result, the evolving criteria for assessing writing skills were task fulfillment, coherence and cohesion, vocabulary (range and appropriateness), grammar (range and appropriateness), and orthography (spelling and mechanics), which are defined by descriptors on five successive half-levels of the CEFR (A1+ to B1+).

The first stage of rating scale development is described in an article published in 2020 by the five researchers who developed the first draft of the scales (Harsch & Seyferth, 2019).

In that initial intuitive phase, the starting point was for the proficiency descriptors and the additional materials in the appendix of the CEFR/CV (Council of Europe, 2020). Other scales

_____

consulted were the *Aptis Speaking rating scale* (O'Sullivan & Dunlea, 2015), the IELTS speaking and writing band descriptors (IELTS 2016), and the *Pearson Global Scale of English Learning Objectives for Academic English* (Pearson English, 2019). 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. Table 1 shows the final draft of the rating, with which we will go into training and validation with the CLAN members.

Table 1. Final draft rating scale writing, after the first trial (CEFR Companion volume/relevant scales and level | *IELTS* |own additions CAG and TS test specs |after first trial in small group WS6)

	Task Fulfilment	Coherence / cohesion	Vocabulary (range and appropriateness)	Grammar (range and accuracy)	Orthography (spelling and mechanics)
B1+	The message is clearly and appropriately conveyed. (CAG) All ideas/content are relevant to the topic of the task (CAG) 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). Follows the conventions of the text type required by the task (CAG). 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. (CAG) Establishes more complex relations between ideas, e.g. 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 5). 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.* (CAG, adapted from CV 137).
B1	The message is generally clearly conveyed. (CAG) The ideas/content are generally relevant to the	Mostly organizes ideas into a meaningful sequence, with adequate topic progression (TS, GE).	Uses sufficient topic-specific vocabulary to express themselves on familiar topics. (CV page 132)	Uses a range of simple grammatical features and sentence structures with	Produces generally intelligible spelling for most common words, *mother tongue influence is likely with less*

topic of the task. (CAG) 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). Mostly follows the conventions of the text type/format required by the task (CAG), *but the format may be inappropriate in places* (IELTS band 5). *Shows awareness of the required register, but may still be inconsistent* (IELTS band 5). Generally follows salient politeness conventions, but not always appropriately (adapted from CV 138)	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)	Shows appropriate use of a wide range of 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)	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 IELTS band 5) *In general, the reader can interpret the errors correctly based on the context*. (adapted from CV p. 174)	*common words*. (CAG) Punctuation is 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. (CAG) 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/*	Shows some organization of ideas and a clear attempt at topic progression (TS). *May still show some limitations in sequencing and text structure. (CAG)* Paragraph breaks may be missing. (CAG) Uses the most frequently occurring connectors to link simple sentences in order to tell a story or describe something like a	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.* (CAG and adapted from IELTS) May *have to compromise the message and may use repetitions and circumlocutions*	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, but it is usually clear what s/he is trying to say*.	Writes with reasonable phonetic accuracy, *but the mother tongue is likely to be noticeable*. Punctuation is still *likely to be influenced by the mother tongue*. (CAG, adapted from CV p. 137). *Errors may cause occasional strain on the reader*. (CAG, *(adapted from IELTS band 4)*

	contrasting ideas) (Test specs page 8 and adapted from CV page 138). *May use an inappropriate format* (IELTS band 4). *May use an inappropriate tone (IELTS Band 4).*	simple list of points (CV p 142). *May use less frequent cohesive devices inappropriately. (CAG)*	*(adapted from CV 131 and CAG).*	(adapted from CV p. 133, 174). *May show attempts at more complex structures, but usually, these are erroneous.*	
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. (CAG) Performs the more concrete language functions required by the task (e.g.social exchanges, invitations, etc.). (Test specs page 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)*.	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 (CAG)* Links groups of words with simple connectors like 'and, 'but' and 'because' (CV p 142). *May overuse connectors, may use other cohesive devices unsuccessfully. (CAG)*	Shows 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*. (CAG 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. (CAG) Errors in spelling and punctuation may cause strain for the reader. (adapted from IELTS band 4)*
A1+	The message only partly gets across and usually requires a sympathetic reader. (CAG) Shows awareness of the required topic but the ideas	Links words or groups of words with very basic linear connectors like 'and' or 'because' (CV p. 142). Texts longer than short notes and	Shows a very basic range of simple vocabulary and memorized expressions related to particular concrete situations (CV p. 131-132)	Shows only a few simple grammatical features and sentence patterns in a learnt repertoire (CV p. 133).	Writes only familiar words and short phrases used regularly with reasonable accuracy. Spells his/her address, nationality, and

_____

are very limited. (CAG) Performs only the most concrete language functions (e.g. establish social contact) (CAG, adapted CV 138) The format and tone are mostly inappropriate. (CAG)	messages generally show coherence problems that make them very hard or impossible to understand (adapted from CV p. 174).	*May overuse certain words (CAG)*	*Errors are likely to be frequent and common.* (CAG)	other personal details correctly. Uses only basic punctuation (full stops and question marks (adapted from CV. p. 137)

The second stage of the qualitative method included raters' training and scale validation. It is developed in two different workshops in which the following outcomes were obtained:

- revision of the rating scales tailored to the Cuban Higher Education (CHE) according to the writing assessment criteria.
- restating the assessment criteria: Task Fulfilment (register, topic, text type), coherence/cohesion, vocabulary (range-appropriateness), grammar (range-accuracy), and orthography (spelling-mechanics).
- three scripts samples were analyzed in each workshop, aiming at validating scales and identifying benchmarks.
- proposals for descriptor wording improvement (5th version of rating scales)

Finally, an online course on assessment literacy was taught to directors of language centers from all universities in February 2020, with a high level of satisfaction among participants.

Conclusions

The starting point in aligning the curricular expectations in the Cuban Higher Education with international proficiency frameworks are transparent test specifications based on not only international reference frameworks such as the CEFR but also on the needs of the Cuban context.

The tests specifications established describe target competencies, task characteristics, and expected attributes, which are the basis for developing the exam.

The rating scales developed for writing assessment become a valuable tool for constructive alignment between curriculum development, instruction, classroom assessment, and national proficiency testing. Based on some of the most internationally recognized descriptors and scales, they respond to the higher education local needs and expectations to describe in a standardized qualitative way the observed student performances.

_____

References

Council of Europe. (2001). *Common European Framework of reference for languages: Learning, teaching assessment*. Retrieved july 3, 2020, from http://www.coe.int/lang-CEFR

Council of Europe. (2018). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment Companion Volume With New Descriptors*. Retrieved April 12, 2020, from www.coe.int/lang-cefr

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Retrieved May 5, 2021, from www.coe.int/lan

Deygers, B. & Van Gorp, K. (2013). The influence of the CEFR on rating scale design. In J. Colpaert, M. Simons, A. Aerts, & M. Oberhofer (Eds.), *Book of Abstracts, "Language Testing in Europe: Time for a New Framework?"* (pp. 57-59). Antwerp: The University of Antwerp. Retrieved May 20, 2021, from https://www.academia.edu/3549147/The_influence_of_the_CEFR_on_rating_scale_design

Fulcher, G. (2004). Deluded by artifices? The common European framework and harmonization. *Language Assessment Quarterly, 1*(4), 253-266. Retrieved September 12, 2021, from https://www.tandfonline.com/doi/abs/10.1207/s15434311laq0104_4

Fulcher, G., Davidson, F. & Kemp, J. (2011). Effective rating scale development for speaking tests: Performance decision trees. *Language Testing, 28*(1), 5-29. Retrieved September 12, 2021, from doi: https://journals.sagepub.com/doi/10.1177/0265532209359514

Harsch, C. & Martin, G. (2012). Adapting CEF-descriptors for rating purposes: Validation by a combined rater training and scale revision approach. *Assessing Writing*(17), 228–250. Retrieved September 12, 2021, from https://www.sciencedirect.com/science/article/abs/pii/S1075293512000244

Harsch, C. & Seyferth, S. (2019). Marrying achievement with proficiency developing and validating a local CEFR-based writing checklist. *Assessing Writing*(43), 10-43. Retrieved May 20, 2021, from https://doi.org/10.1016/j.asw.2019.100433

IELTS. (2016). *IELTS TASK 2 Writing band descriptors (public version)*. Retrieved May 20, 2021, from https://takeielts.britishcouncil.org/sites/default/files/ielts_task_2_writing_band_descriptors.pdf

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

Europe.

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

Pearson English. (2019). *Global Scale of English Learning Objectives for Academic English Flipping Book*. Retrieved May 20, 2021, from http://www.bli.ir/ar/06CEFR/4.GSE_LO_Academic_English.pdf

Piccardo, E., North, B. & Goodier, T. B. (2019). Broadening the Scope of Language Education: Mediation, Plurilingualism, and Collaborative Learning: the CEFR Companion Volume. *Journal of E-Learning and Knowledge Society*(15), 17–36. doi:10.20368/1971-8829/1612

Síntesis curricular:

Tamara Gutiérrez Baffil: Licenciada en Educación. Especialidad Lengua Inglesa. Doctora en Ciencias. Máster en Ciencias de la Educación. Profesora Auxiliar y directora del Centro de Idiomas de la Universidad de Pinar del Río. Tiene 23 años de experiencia en la Educación Superior. Premio Profesional en formación de profesores de L2. Participa en proyectos científicos internacionales y nacionales sobre gestión de procesos, enseñanza del inglés y aplicación del MCER en el contexto cubano, que se presenta en esta contribución.

Ivonne de la Caridad Collada Peña: Licenciada en Filología. Master en Gestión. Profesora Auxiliar y directora del Centro de Idiomas de la Universidad de las Ciencias Informáticas. Ha sido profesora durante más de 30 años, directora de departamentos de inglés y decana de la facultad. Ha impartido cursos de lingüística, lexicología, ESP y EFL, entre otras materias. Sus intereses de investigación incluyen métodos de TESOL, desarrollo curricular, lexicología, gramática y enseñanza y evaluación de idiomas. Actualmente es la coordinadora cubana del proyecto que se presenta en esta contribución.

Declaración de responsabilidad autoral:

Tamara Gutiérrez Baffil: Aportó el diseño de la investigación, colección, análisis e interpretación de los datos. Tuvo a su cargo la redacción del artículo.

Ivonne de la Caridad Collada Peña: Participó en el análisis e interpretación de los datos, Tuvo a su cargo la revisión y corrección final del artículo.