

# REALL: Rubric for the Evaluation of Apps in Language Learning

## Resumen

En la última década se ha generalizado el uso de rúbricas o plantillas para una evaluación estandarizada en educación, con varias ventajas asociadas a su uso: una evaluación más objetiva, comprensión clara de los criterios utilizados, homogeneización de las expectativas y características deseables de los trabajos de los alumnos, etc. En consonancia con esto ha habido diversos intentos de creación de rúbricas para evaluar “apps” educativas (véase, por ejemplo, Avatar Generation, 2012 o Santiago, 2012), pero no se ha avanzado mucho en el área específica de la enseñanza de lenguas extranjeras. Nuestra contribución pretende llenar ese vacío mediante la presentación de una rúbrica que incluye criterios tanto educativos como lingüísticos.

Con esta finalidad se ha creado una rúbrica siguiendo el formato de las rúbricas analíticas que permite prestar una mayor atención a los aspectos específicos de la enseñanza y aprendizaje de lenguas, tal y como ha sido definido por el Marco Común Europeo de Referencia para las Lenguas o MCER –comprensión y producción oral, comprensión y producción escrita, interpretación y traducción– ((Consejo de Europa 2001); y proporcionar descriptores detallados para cada categoría.

Dicha rúbrica está basada en versiones anteriores (Arús-Hita, Rodríguez-Arancón y Calle-Martínez, en prensa; Rodríguez-Arancón, Arús-Hita y Calle-Martínez, en prensa) desarrolladas para la evaluación pedagógica de aplicaciones educativas móviles en general, fruto del trabajo de ATLAS (Artificial Intelligence for Linguistic ApplicationS), un grupo de investigación consolidado formado por 17 investigadores de diferentes universidades españolas, dentro de su proyecto SO-CALL-ME (Entorno móvil de aprendizaje de lenguas basado en ontologías sociales y realidad aumentada, en sus siglas inglesas).

Tanto la rúbrica inicial como la orientada a la enseñanza de lenguas que se presenta en esta comunicación están basadas en una guía de criterios de calidad para la evaluación y creación de objetos de aprendizaje (Fernández-Pampillón et al. 2011). La aplicación combinada de los criterios de calidad y los descriptores del MCER dan como resultado una rúbrica que no sólo facilita la evaluación de aplicaciones para lenguas extranjeras ya existentes, sino que también se convierte en una valiosa guía para la creación de “apps” nuevas. La discusión y conclusión de este artículo proporcionan evidencia de la aplicación de la rúbrica a la evaluación real de las “apps” más comunes en este campo en los diferentes sistemas operativos. Además, la conclusión enfatiza el potencial de la rúbrica y sus descriptores para mostrar puntos débiles y fuertes de este tipo de aplicaciones.

**Palabras clave:** Aprendizaje móvil, idiomas, evaluación, rúbrica, enseñanza de lenguas extranjeras

## Abstract

Rubrics, or documents for standardized assessment have been generalized in education in the past decade, and several benefits can be drawn from their use: a more objective assessment, a clear understanding of the criteria used, a homogenization of expectations and desirable features, etc. Thus, there have been several attempts to create rubrics for evaluating educational apps (see, for example, Avatar Generation, 2012 or Santiago, 2012) but not much has been done specifically in the field of Foreign Language Teaching (FLT). Our contribution seeks to fill that gap by presenting a rubric which includes criteria that are educational but also linguistic.

To that end, a template was created following the format of an analytic rubric, enabling to focus on the specific dimensions of language teaching and learning, as defined by the Common European Framework of Reference for Languages or CEFR (Council of Europe 2001): oral reception and production, written reception and production, interpretation and translation; and providing detailed descriptors for each category. This rubric is based on previous ones (Arús-Hita, Rodríguez-Arancón and Calle-Martínez, in press, Rodríguez-Arancón, Arús-Hita and Calle-Martínez, forthcoming) for the pedagogic assessment of mobile educational applications in general, developed by members of ATLAS (Artificial inTelligence for Linguistic ApplicationS), a consolidated research group formed by 17 researchers from different Spanish universities, within their project SO-CALL-ME (Social Ontology-based Cognitively Augmented Language learning Mobile Environment).<sup>1</sup> Both the initial rubric and the one geared to FLT, i.e. the rubric presented in this paper, are based on a set of quality criteria previously established. This quality guide was developed following and adapting a previously existing guide of quality criteria for the assessment and creation of Learning Objects (Fernández-Pampillón et al. 2011).

The combined application of the quality criteria and the CEFR dimensions results in a rubric that not only allows the evaluation of existing FLT apps but also provides valuable guidance in the creation of new ones. The discussion and conclusion provide evidence of the application of the rubric to the actual evaluation of the most commonly used FLT apps for MALL (Mobile Assisted Language Learning) in the different operating systems available (mainly Android and iOS). Furthermore, the concluding part of our paper emphasizes the potential of the rubric and its descriptors to pinpoint constraints but also affordances in apps for FLT.

**Keywords:** Mobile learning, languages, assessment, rubric, Foreign Language Teaching

## Introduction

Mobile learning can have various meanings for different groups of people. Superficially, it appears from the outside to be learning via mobile devices such as smartphones, MP3 players, laptops and tablets. Certainly, these are important in enabling mobile learning. But mobile learning is more than just using a mobile device to access content and communicate with others - it is about the mobility of the learner. Mobile learning can be defined as the processes, both personal and public, of coming to know through exploration and conversation across multiple contexts amongst people and interactive technologies (Sharples et al., 2007). Little by little, mobile learning is taking force in the field of education, which uses increasingly more portable tools as a support in the classrooms. Mobile devices are not only for the benefits of schools; numerous projects that enhance their educational use are being carried out outside the traditional classrooms.

The concept of *lifelong learning* is a term associated to mobile learning and at the same time to change. *Lifelong learning* means the training throughout the life cycle of a person. This is the key element of the new century and it is linked to the concepts of “educational society” and “knowledge society” which pursue to raise the level of awareness of as many people capable and willing to learn so that everyone can better understand the nature of things. The aim of *lifelong* education should be to provide the means to achieve a better balance between work and learning.

According to Paine (2011) mobile learning offers many benefits for learning. The access anytime, anywhere makes learning available in new situations. It can happen during ‘dead times’, that is, while travelling or waiting for a meeting to start. It fits many different learning

---

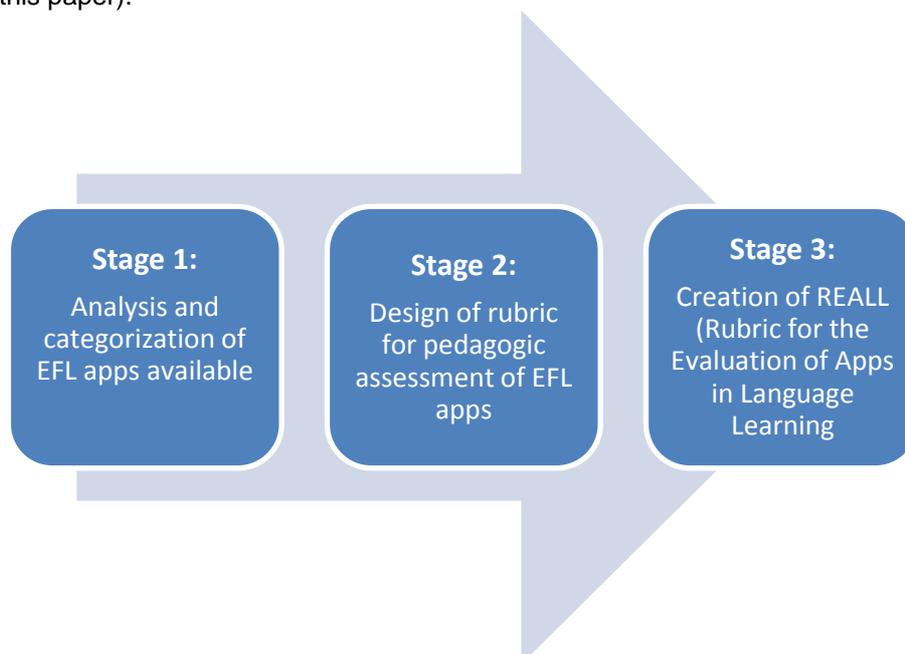
<sup>1</sup> With support from the Spanish Ministry of Science and Innovation (ref. no. FFI 2011-29829).

styles, such as, reading, listening to podcasts, contributing to discussions. All these are means for offering learning on mobile devices.

Thus, there have been several attempts to classify educational apps and categorize them using standards or rubrics which provide a more objective assessment, a clear understanding of the criteria used, a homogenization of expectations and desirable features, etc. (see, for example, Avatar Generation, 2012 or Santiago, 2012) but not much has been done specifically in the field of Foreign Language Teaching (FLT). Our contribution seeks to fill that gap by presenting a rubric which includes criteria that are educational but also linguistic.

This paper describes research undertaken within the SO-CALL-ME (Social Ontology-based Cognitively Augmented Language learning Mobile Environment) project, which has a double purpose. Firstly, it is planned to design and develop a theoretical framework for a new, hybrid mode of computer-assisted language learning: social and ubiquitous, incorporating augmented reality techniques and accessible from the latest handheld devices (smartphones, tablet PCs, etc.). This will enhance flexible, adaptive, interactive, practical learning, very much related to everyday communicative socio-cultural contexts and the use of (foreign) language. Secondly, it is intended to design and develop a linguistic ontology of visual learning objects which will boost foreign language learning, avoiding the problems caused by other learning materials which are largely textual, static and de-contextualised from our surrounding socio-cultural reality. The underlying hypothesis is that the increasing sophistication of mobile devices can be a real asset for foreign language learning, which is convenient because of its portability and widespread use among professionals and higher education students, but also efficient and pedagogically rigorous.

In this sense, and as a starting point for the development of MALL (Mobile Assisted Language Learning) applications for EFL (English as a Foreign Language), within the context of the SO-CALL-ME research project, our paper offers an examination of both the qualities and limitations of the most outstanding MALL applications in the market at the moment by assessing their characteristics from a pedagogic and linguistic point of view. This research is being developed in subsequent stages: stage 1 comprises an analysis of the EFL apps available and a categorization; stage 2 consists in the design of a rubric for the pedagogic assessment of EFL apps and stage 3 involves the creation of a rubric which is specifically linguistic (REALL, the focus of this paper).



**Figure 1: Research stages**

## Pedagogic assessment of mobile learning apps for EFL

### Stage 1: Analysis and categorization of EFL apps available

The objective of this particular phase of the project was to analyze some of the over 28,000 educational applications for mobile devices available in the market at the moment.<sup>2</sup> This would represent a starting point from which to develop our own apps after gaining knowledge and insights into the features that are effective and suitable for learners using MALL. This original assessment phase did not focus on the technical specifications of the apps, but rather on their pedagogic goals, in a most general sense. No in-depth methodological analysis of any particular app was therefore intended at that stage. In order to carry out this evaluation process, two templates were created, and shared through Google Drive: the first was a table with two columns and an extendible number of rows where each of three evaluators could indicate the app assessed and their URL to avoid any possible repetitions. The second template consisted in an in-house created rubric with three criteria and a scale from one to five. The intention was to keep the rubric simple and very much geared towards our project's specific needs. The purpose was to assess as many apps as possible within a relatively short space of time and guarantee homogeneity in the process. The three criteria considered were: 1) the apps' cognitive value; 2) similarity of the app with the pedagogic aims of the SO-CALL-ME project; and 3) complementarity with the pedagogic aims of the SO-CALL-ME project. Each rubric was also accompanied by a brief description of the app and a final evaluative remark.

A total of 67 EFL apps were assessed, combining the study of the information available on the websites describing each app and, whenever possible, tested on a mobile device –i.e. when they were free to download. Each evaluator assessed different apps, which has the advantage of providing information about a larger number of them but also the potential disadvantage of less reliable assessments. However, the comparison of the rubrics in the only two cases in which two evaluators accidentally assessed the same app proved to show rather similar criteria of analysis.

The conclusion of this first phase was that a high number of apps presented technical problems at the time of downloading or starting them. In fact, more than one third of the apps downloaded by the evaluators proved not to work properly or not to work at all. Concerning software, the vast majority of apps assessed were available for Apple devices –iPhone, iPad and, sometimes, iPod Touch – and around one in four were also available for Android; very few were *only* available for the latter; and other operating systems such as BlackBerry OS, Bada or Ovi seem to be much less targeted by app developers. A few of them could also be directly run from the Internet on a conventional computer.

Regarding prices, three marketing approaches were defined: the most expensive apps which were in fact mobile versions of traditional dictionaries, textbooks, vocabulary or grammar tests, etc. with a price as high as 30 euros. A second group of apps downloadable for a small amount –usually around one euro, and rarely above three euros– such as Cambridge's *English Monstruo*, and those apps with an initial free sample pack and the possibility to download further packs for a small amount (as e.g. the British Council's *LearnEnglish Grammar*). A final group of English courses such as *Busuu* or EF's *EnglishTown*, where the price depends on the needs of the user and/or seasonal offers.

The apps could also be categorised in several groups: a) Games, very often aimed at children, e.g. the apps available from Cambridge English Online; b) app versions of dictionaries, handbooks and textbooks, e.g. Cambridge's EFL methods, dictionaries, etc.; c) apps providing vocabulary, grammar and/or pronunciation practice, such as *My Word Book*, *Johnny Grammar's Quiz Master*, *60 Second Word Challenge* or *Sounds Right*, some which allow the practise of different skills beyond mere drilling or quizzing in the form of listening comprehension by means of podcasts, e.g. *Listen-to-English* and *A Cup Of English*, and apps allowing conversation practice, e.g. *English Feed*, even with other users, e.g. *The Language Campus*; d) the

---

<sup>2</sup> <http://www.eduapps.es>

adaptation of online courses such as *Busuu* and EF's *EnglishTown* to mobile devices; e) most closely related to the interests and goals of the SO-CALL-ME project are those apps exploiting the use of language in context and presented in a variety of ways, such as podcasts –e.g. *Learn English, Talking Business English*– videos –e.g. *Learn English Audio & Video, Conversation English*– films –e.g. *English Attack*– and cartoons –e.g. *Big City Small World*. It is also worth mentioning the existence of apps such as the mobile version of *Voxi*, where users select the situation in which they need to use their English and the app tells them the expressions to be used, although the output is rather limited.

A last item resulting from the assessment which will be very relevant for the future development of our own apps concerns those features which differentiated some apps from the rest and provided and added value. For instance, the drag-and-drop facility available, e.g. *Learn English Grammar*, the possibility to draw with your finger, as in *Premier Skills*, connectivity with social networks, as offered by *Language City, Learn English, 60 Second Word Challenge* and *Tongue Mystery English*, and, finally, a feature we found particularly appealing from a pedagogical point of view, i.e. the inclusion of an Avatar, as in Cambridge's *Quiz up*. As Cohen (2007) states: "Avatars are excellent for online education. They provide the human interaction that is natural in classrooms and in the traditional learning environment".

### **Stage 2: Design of an evaluation rubric for the pedagogic assessment of EFL apps**

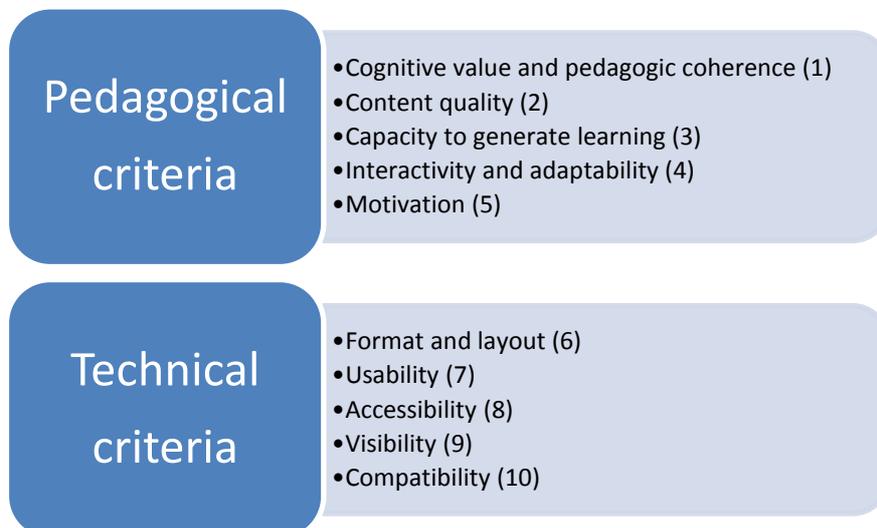
In "a pedagogic assessment of mobile learning applications" (Arús, Rodríguez-Arancón, Calle, in press) we reported on the assessment carried out on a number of MALL applications in the context of EFL so as to gain a global overview of the teaching and/or practising points they cover. Our assessment was made by means of a rubric created in-house. This rubric geared the assessing task towards the specific needs of the SO-CALL-ME project, and reflected a quantitative rather than a qualitative approach.

We first assessed EFL apps focusing not on their technical specifications but on their pedagogic goals, in a most general sense. To that effect, two templates were created: a) a list of apps assessed and the URL from which such app is available, so each evaluator would know what apps had already been dealt with by the two other evaluators and thus avoid repetitions; b) an in-house created rubric with three criteria and a scale from one to five for each of the criteria. The purpose of this rubric was to guarantee homogeneity in the assessment process and to provide a means for relatively fast assessment. The rubric was therefore kept as simple as possible, and very much geared towards our project's specific needs. The three criteria considered were: 1) the apps cognitive value; 2) similarity of the app with the pedagogic aims of the SO-CALL-ME project; and 3) complementarity with the pedagogic aims of the SO-CALL-ME project. Each rubric was to be accompanied by a brief description of the app and a final evaluative remark.

A total of 67 EFL apps were assessed, combining the scrutiny of the information available on the websites describing each app and, whenever possible, tested on a mobile device –i.e. when they were free to download and once downloaded the apps ran well. The results obtained from the assessing process gave us an idea of the qualities and limitations of the apps evaluated, as a first step in the development –within the context of our project- of other apps that may fill some existing gaps. Pending a more in-depth assessment of specific apps, the quantitative scrutiny allowed us to ascertain the limited scope of many of the existing products. It is fact that they tend to provide a rather fragmented language practice: some vocabulary here, some grammar there, etc. Some of the MALL apps evaluated, however, do provide more contextualized practice. It is precisely some of these apps that we look at more in detail in "The use of current Mobile learning applications in EFL" (Rodríguez-Arancón, Calle & Arús, forthcoming).

In that paper we report on the work carried out in order to develop the necessary tools to evaluate and create educational apps. A quality guide and a rubric were the results of such work. The guide, based on the one created by Fernández-Pampillón *et al.* (2012) for the creation of learning objects, encompasses the quality criteria for the evaluation and creation of educational apps. The app quality guide takes the ten criteria used by Fernández-Pampillón *et*

al. and adapts them to the characteristics and goals of educational apps. An important aspect of this guide is that it combines pedagogical criteria with technical ones. The ten quality criteria are pedagogical (Cognitive value and pedagogic coherence; Content quality; Capacity to generate learning; Interactivity and adaptability; and Motivation) and technical (Format and layout; Usability; Accessibility; Visibility; and Compatibility) as can be seen in figure 2 below.



**Figure 2: Quality criteria for the creation of digital learning objects**

The sub-criteria within each criterion have also been adapted to meet the needs of educational applications. For instance, one of the points within this first criterion for the evaluation of Learning Objects refers to the existence of a metadata file specifying goals, skills, etc. Since this kind of files are specific to learning objects but irrelevant to apps, no mention of metadata files is made in our quality guide.

Based on this guide, a new rubric was designed to facilitate the app evaluation process. The information in the cells is based on the specifications made in the quality guide. The way in which we proceeded was to first fill in the cell corresponding to the maximum marks, i.e. 5, with the fulfillment of all the subcriteria and gradually slacken such fulfillment as we move down the scale, till the minimum marks, i.e. 1, is reached, where none of the sub-criteria is fulfilled. Table 1 shows a row in the rubric, corresponding to one of the ten criteria.

**Table 1: Criterion 3 in the educational app evaluation rubric**

	1	2	3	4	5
3. <i>Capacity to generate learning</i>	Contents do not help to achieve learning goals or autonomous learning	Contents help autonomous learning but not clearly the achievement of the initial learning goals	Contents help to achieve the learning goals but neither autonomous learning nor relating old knowledge to new knowledge	Contents help to achieve the learning goals but not autonomous learning OR not relating old knowledge to new knowledge	Contents help to achieve the learning goals, autonomous learning and relating old knowledge to new knowledge

Five of the 67 previously evaluated EFL apps with the highest marks, i.e. with the highest potential to serve as sources of inspiration for the apps to be developed, were chosen for a

preliminary evaluation: *Englishfeed*, *Speakingpal*, *Clear Speech*, *LearnEnglish Audio & Video*, *LearnEnglish Elementary Podcasts* (see figure 3 below). As stated in our paper, this number is still too small to statistically measure the evaluators' agreement, yet the results obtained seem to show consistency between the two evaluators. Another interesting fact is that, pending further evaluation, criterion 4 –Interactivity and adaptability– seems to be the weakest one in the apps evaluated. This comes as no surprise, as the specifications of this criterion in our quality guide include some of the essential requisites for successful FLT, e.g. contextualized teaching, which are also the ones with which FLT methods have traditionally struggled.



Figure 3: Apps chosen to be assessed in stage 3

Because the weakest point in the evaluated apps has to do with key methodological issues, we found it was necessary to tackle those aspects and therefore zero in on EFL-specific methodology as a prior step to the design and development of EFL apps. We therefore looked at the Common European Framework of Reference (CEFR henceforth, Council of Europe, 2001) in order to establish a benchmark that was specifically linguistic.

### Stage 3: Rubric for the evaluation of apps in language learning (REALL)

The CEFR has become in over a decade the key reference for anyone involved in learning, teaching or assessing foreign languages: educational administrators, course designers, teachers, teacher trainers, examining bodies, etc. It provides categories and educational levels with detailed descriptors which facilitate the elaboration of curricula and materials for FLT. It is thus a valuable tool to be incorporated in the evaluation of EFL apps and so it was considered by the authors of this paper. The implementation of the CEFR was done in such a way that it complemented the previous stages of research and meant an added value to the pedagogic assessment that had been already fulfilled. The process is shown in figure 4 below:

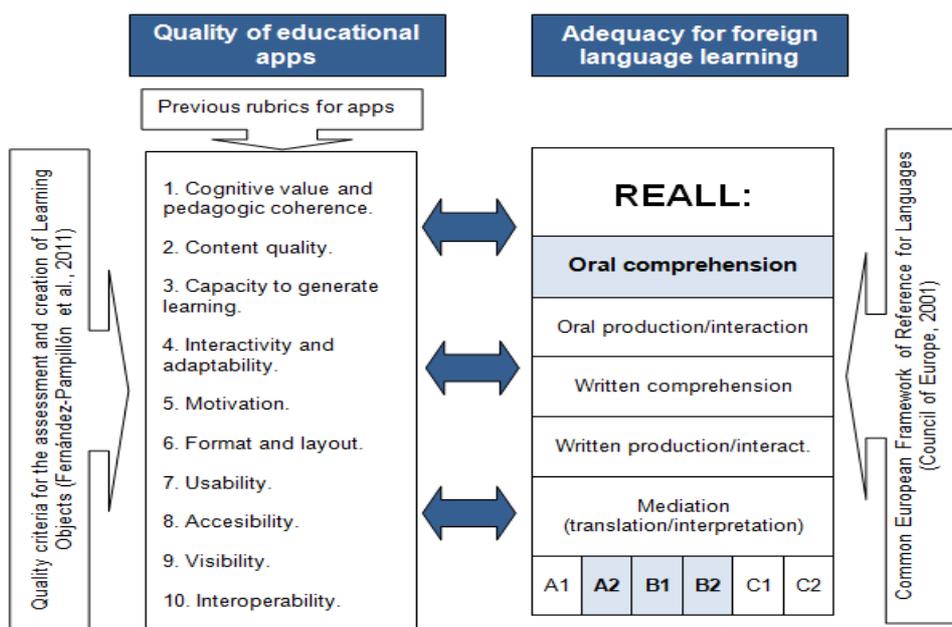


Figure 4. Pedagogic and linguistic evaluation of apps for foreign languages

The CEFR breaks language competence into three differentiated levels (Level A, basic user; Level B, independent user; Level C, proficient user) which can be further sub-divided into two, resulting in a total of six levels: A1 or breakthrough, A2 or waystage, B1 or threshold, B2 or vantage, C1 or effective operational proficiency and C2 or mastery. For the purposes of this research we have focused on levels A2-B2, which are the ones that cover the majority of the EFL learners and users. Table 2 below shows the descriptors for those levels in the CEFR global scale. The words or phrases in bold letters show the key terms highlighted in order to create REALL.

**Table 2. Common Reference Levels: Global scale**

B2 (Independent user)	Can <b>understand</b> the main ideas of <b>complex text on both concrete and abstract topics</b> , including <b>technical discussions</b> in his/her field of <b>specialisation</b> . Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and Independent disadvantages of various options.
B1 (Independent user)	Can <b>understand</b> the main points of <b>clear standard input on familiar matters</b> regularly encountered in <b>work, school, leisure, etc.</b> Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.
A2 (Basic user)	Can <b>understand</b> sentences and <b>frequently used expressions</b> related to <b>areas of most immediate relevance</b> (e.g. <b>very basic personal and family information, shopping, local geography, employment</b> ). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.

SO-CALL-ME has a clear focus on oral competence, which is why the development of REALL gave priority to this skill. The starting point has been oral reception, but the rest of the language activities described by the CEFR will follow (oral production and interaction, written reception, written production and interaction, interpretation and translation). In this line, the CEFR descriptors for listening competence were analysed and highlighted accordingly. Table 3 shows an excerpt of those levels:

**Table 3. Common Reference Levels: Listening**

B2 (Independent user)	Can understand <b>extended speech and lectures</b> and follow even complex lines of argument provided the topic is reasonably familiar. Can understand <b>most TV news and current affairs programmes</b> . Can understand the majority of <b>films in standard dialect</b> .
B1 (Independent user)	Can understand the main points of <b>clear standard speech</b> on familiar matters regularly encountered in work, school, leisure, etc. Can understand the <b>main point of many radio or TV programmes on current affairs or topics of personal or professional interest</b> when the <b>delivery</b> is relatively <b>slow and clear</b> .
A2 (Basic user)	Can understand <b>phrases and the highest frequency vocabulary</b> related to <b>areas of most immediate personal relevance</b> (e.g. <b>very basic personal and family information, shopping, local area, employment</b> ). Can catch the <b>main point in short, clear, simple messages and announcements</b> .

The CEFR includes the description of three projects that put this methodological approach into practice: The Swiss research project, the DIALANG scales and the ALTE “Can Do” statements (for further information see CEFR annexes, Council of Europe, 2001). They are mostly user-

centred, whereas the research shown in this paper is material-centred, since it focuses on the EFL apps. However, out of the three, the DIALANG scales for listening provided some useful information that could be transferred to featuring FLT materials and resources and was consequently selected. Table 4 shows an extract of these scales:

**Table 4: The DIALANG Scales for listening**

	<b>What types of text I understand</b>	<b>What I understand</b>	<b>Conditions and limitations</b>
B2 (Independent user)	<b>All kinds of speech on familiar matters. Lectures.</b> Programmes in the <b>media</b> and <b>films</b> . Examples: technical discussions, reports, live interviews.	Main ideas and specific information. Complex ideas and language. Speaker's viewpoints and attitudes.	<b>Standard language and some idiomatic usage, even in reasonably noisy backgrounds.</b>
B1 (Independent user)	<b>Speech on familiar matters and factual information. Everyday conversations</b> and discussions. Programmes in the <b>media</b> and <b>films</b> . Examples: operation instructions, short lectures and talks.	The meaning of some unknown words, by guessing. General meaning and specific details.	<b>Clear, standard speech.</b> Will require the help of visuals and action. Will sometimes ask for repetition of a word or phrase.
A2 (Basic user)	<b>Simple phrases</b> and expressions about things important to me. <b>Simple, everyday conversations</b> and discussions. <b>Everyday matters in the media.</b> Examples: messages, routine exchanges, directions, TV and radio news items.	Common everyday language. Simple, everyday conversations and discussions. The main point. Enough to follow.	<b>Clear and slow speech.</b> Will require the help of sympathetic speakers and/or images. Will sometimes ask for repetition or reformulation.

All this resulted in REALL, a rubric which has been used to evaluate the linguistic adequacy of EFL apps for listening. It follows the same pattern as the previous rubric: the information in the cells takes the quality guide as a reference starting point and the cell with the maximum marks contains all the sub-criteria, which are thinned out until 1, the minimum mark. An extra column was added to indicate the cases in which none of the descriptors were applicable. The categories chosen are the following: level, types of texts, topics and delivery. An example of the fourth category, delivery, is shown in table 5 below:

**Table 5: Delivery in REALL**

	1	2	3	4	5	N/A
<i>Delivery</i>	Language difficulty, clarity and speed mix different levels. If	Language difficulty, clarity and speed rarely belong to the same level. If	Language difficulty, clarity and speed tend to belong to the same level. If	Language difficulty, clarity and speed usually belong to the same level. If	Language difficulty, clarity and speed belong to the same level. If	

	adaptive, delivery is clearly not well adapted	adaptive, delivery rarely corresponds to the right level	adaptive, delivery more often than not corresponds to the right level	adaptive, delivery usually corresponds to the right level	adaptive, delivery corresponds to the right level	
--	--	--	---	---	---	--

The evaluating process was parallel to the one used in the pedagogic assessment of stage two: two evaluators analysed the five chosen apps (*Englishfeed*, *Speakingpal*, *Clear Speech*, *LearnEnglish Audio& Video*, *LearnEnglish Elementary Podcasts*), in order to ascertain their linguistic adequacy according to the CEFR. Again, the number is too small to reach definitive conclusions, but it served the authors to pilot REALL and show the consistency between the two evaluators, since there were minimum discrepancies between the evaluators (see appendix).

All five apps cater for A2-B2 language learners and the most salient aspect of the evaluation is that only two of them “pass” this linguistic assessment, achieving more than half of the possible marks: *Learn English Elementary Podcasts* and *Speaking pal* seem to be the most comprehensive ones, since they both obtain high scores in stage 2 and stage 3 evaluations. See table 6 for a comparative study.

**Table 6: Comparative study of stage 2 and stage 3 evaluations**

Pedagogic assessment		Linguistic assessment	
Name of the app	Score	Name of the app	Score
<i>Speakingpal</i>	91	<i>LearnEnglish Elementary Podcasts</i>	37
<i>LearnEnglish Elementary Podcasts</i>	89	<i>Speakingpal</i>	29
<i>Englishfeed</i>	82	<i>LearnEnglish Audio&amp; Video</i>	19
<i>LearnEnglish Audio&amp; Video</i>	81	<i>Clear Speech</i>	14
<i>Clear Speech</i>	64	<i>Englishfeed</i>	13

## Conclusions

After this three-stage research that involved the categorization of EFL apps available in the market, the design of a rubric for the pedagogic assessment and a specifically linguistic rubric for a subsequent evaluation, it can be concluded that the pedagogic and technical quality of the app does not necessarily go hand in hand with its linguistic value and adequacy for EFL teaching and learning: Only two of the five apps that got the highest score in the pedagogic assessment achieved a reasonably good score when applying REALL. This evaluation made clear the fact that apps initially attractive to the user of MALL are not necessarily backed up by a sound linguistic content that is adequate for steady language learning. This should provide “food for thought” for all those involved in the design of language apps, making us reflect on the importance of both dimensions when creating an app for FLT.

It goes without saying that a further, more ample sampling and quantitative research is needed in order to reach definitive conclusions, but this is a sound starting point in that direction. Both rubrics (stage 2 and stage 3) have been sufficiently piloted and are being currently fine-tuned in order to be re-used and adapted for the design of further rubrics that cover the rest of the CEFR competences and can help us reach a full picture in the assessment and evaluation of language learning apps, which ideally should result in a theoretical framework for the design of successful, pedagogically and linguistically sound EFL apps.

## References

- Arús-Hita, Jorge, Pilar Rodríguez-Arancón and Calle-Martínez, Cristina (in press) "A pedagogic assessment of mobile learning applications". In Proceedings of ICDE 2013, *Mobilizing Distance Education*, UNED, Madrid.
- Avatar Generation (2012) "Rubrics for evaluating educational apps". Retrieved from <http://www.avatargeneration.com/2012/09/rubrics-for-evaluating-educational-apps/> (accessed 15/04/2013).
- Cohen, Alan (2007): "Avatars and Education", in Classrooms without Walls, blog. Retrieved from <http://acohen843.wordpress.com/2007/11/11/avatars-and-education/> (accessed 17/04/2013).
- Council of Europe (2001). *Common European Framework of Reference for Languages*. Cambridge: Cambridge University Press.
- Fernández-Pampillón Cesteros, Ana, Elena Domínguez Romero and Isabel de Armas Ranero (2011) *Herramienta para la revisión de la Calidad de Objetos de Aprendizaje Universitarios (COdA): guía del usuario. v.1.1*. Madrid: e-prints Complutense. Retrieved from <http://eprints.ucm.es/12533/> (accessed 18/04/2013).
- Paine, C *et al.* (2011) How mobile technologies are changing the executive learning landscape. UNICON. Ashridge.
- Rodríguez-Arancón, Pilar, Jorge Arús-Hita y Calle-Martínez, Cristina (forthcoming) "The use of current Mobile learning applications in EFL". In Proceedings of IETC 2013, Kuala Lumpur, Malaysia.
- Santiago, R. (2012). Una revisión de la taxonomía del aprendizaje y las apps educativas en el contexto del Mobile-Learning. En S. Tralbaldo (ed.), *10 años de vivencias en educación virtual*. Buenos Aires: Net-Learning.
- Sharples, M., et al. (2007) 'Mobile Learning: Small devices, Big issues'. In Sharples, M., et al. (eds.) *Technology-Enhanced Learning*.

## Appendix

### Stage 2 evaluations (evaluator 1 in red, evaluator 2 in blue):

<i>Englishfeed</i>		<i>Speakingpal</i>		<i>Clear Speech</i>	
Criterion	Mark	Criterion	Mark	Criterion	Mark
1)	3/5	1)	5/4	1)	5/5
2)	3/5	2)	5/5	2)	3/3
3)	4/4	3)	5/5	3)	3/3
4)	2/4	4)	4/4	4)	3/2
5)	4/5	5)	5/5	5)	2/3
6)	3/5	6)	5/5	6)	2/2
7)	4/5	7)	5/5	7)	3/3
8)	5/5	8)	5/4	8)	4/4
9)	3/5	9)	4/4	9)	4/4
10)	4/4	10)	3/4	10)	3/3
TOTAL (100)	82	TOTAL (100)	91	TOTAL (100)	64

<i>LearnEnglish Audio&amp; Video</i>	
Criterion	Mark
1)	4/4
2)	5/5
3)	3/4
4)	2/3
5)	3/4
6)	5/5
7)	5/5
8)	4/4
9)	4/4
10)	4/4
TOTAL (100)	81

<i>LearnEnglish Elementary Podcasts</i>	
Criterion	Mark
1)	5/5
2)	5/4
3)	5/5
4)	3/3
5)	5/4
6)	4/5
7)	5/5
8)	5/5
9)	4/4
10)	4/4
TOTAL (100)	89

### Stage 3 evaluations (evaluator 1 in red, evaluator 2 in blue):

<i>Englishfeed</i>		<i>Speakingpal</i>		<i>Clear Speech</i>	
Criterion	Mark	Criterion	Mark	Criterion	Mark
Level	2/3	1)	1/2	1)	3/3
T.Text	2/NA	2)	NA/NA	2)	3/4
Topics	1/NA	3)	1/NA	3)	4/4
Delivery	2/3	4)	4/4	4)	4/4
TOTAL (40)	13	TOTAL (40)	14	TOTAL (40)	29

<i>LearnEnglish Audio&amp; Video</i>	
Criterion	Mark
Level	1/2
T.Text	5/5
Topics	3/4
Delivery	2/3
TOTAL (40)	19

<i>LearnEnglish Elementary Podcasts</i>	
Criterion	Mark
1)	5/4
2)	4/5
3)	4/5
4)	5/5
TOTAL (40)	37