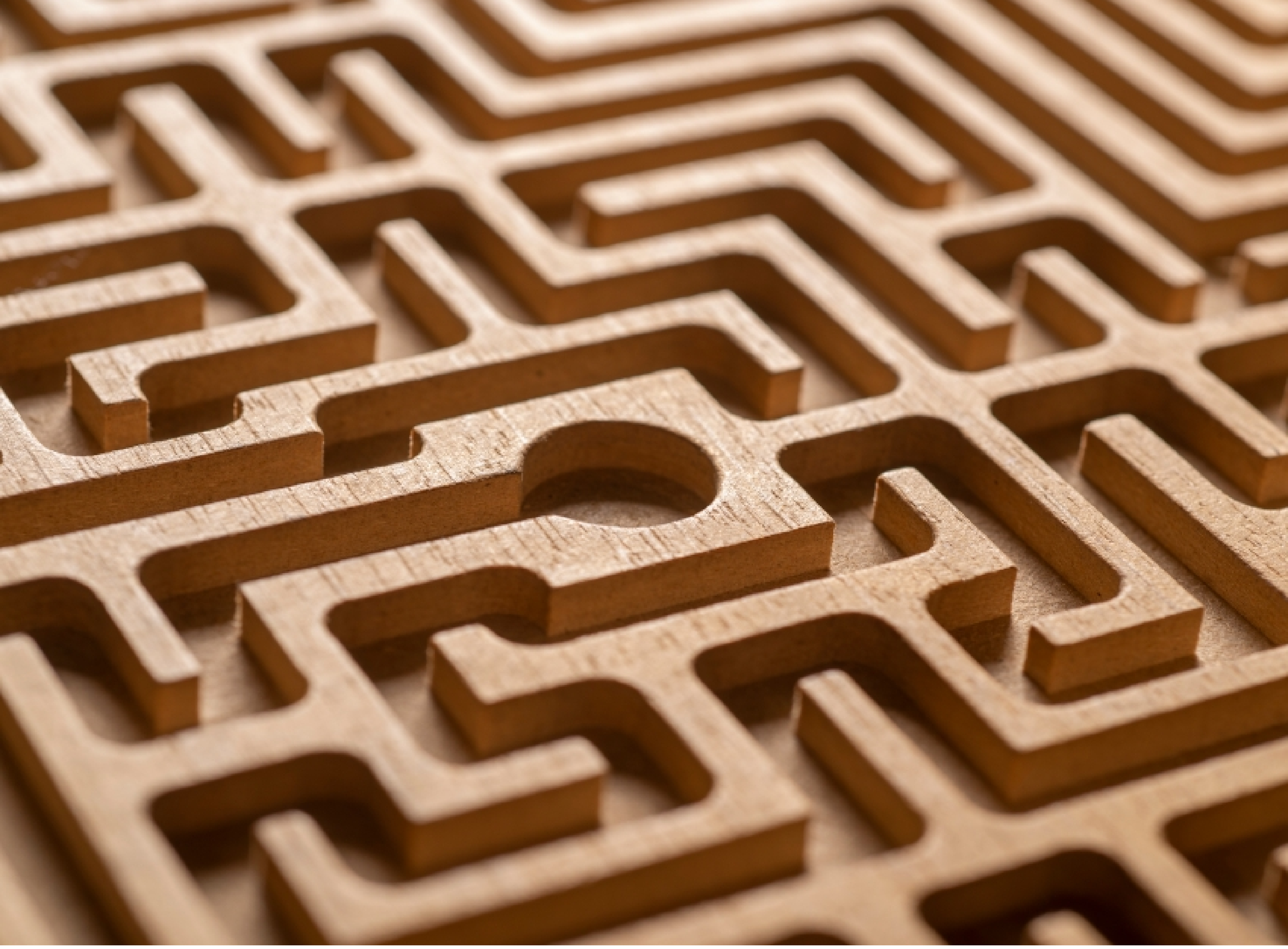


Language Testing and Assessment

Selected Papers from the 45th FAAPI Conference

Edited by

Mario López-Barrios, Melina Porto, María Elisa Romano and Fabiana Sacchi

LAALTA
Latin American Association for Language Testing and Assessment

FAAPI
Federación Argentina de Asociaciones de Profesores de Inglés

Language Testing and Assessment : Selected Papers from the 45th FAAPI Conference / Mario Luis López Barrios ...
[et al.]; compilación de Mario Luis López Barrios. - 1a ed. - Córdoba: Federación Argentina de Asociaciones de Profesores de Inglés, 2022.
Libro digital, PDF

Archivo Digital: descarga y online
ISBN 978-987-98045-2-0

1. Sistemas de Evaluación. 2. Autoevaluación. 3. Enseñanza de Lenguas Extranjeras. I. López Barrios, Mario Luis, comp.
CDD 428.0071

ISBN 978-987-98045-2-0
9 789879 804520

Language Testing and Assessment

Selected Papers from the 45th FAAPI Conference

Edited by

Mario López-Barrios, Melina Porto, María Elisa Romano and Fabiana Sacchi

Contents

Foreword

As the saying goes, there is always a first time. In the 2021 edition of the annual FAAPI Conference, two things were new. This 45th conference was held between September 17th and 19th, 2021, as the first online event of this kind since the creation of the Federation back in 1971. Moreover, for the first time, the Federation held the conference in collaboration with a partner: the Latin American Association for Language Testing and Assessment (LAALTA). A sign of the times: flexibility, innovation and a desire to renew the mission of connecting and supporting English Language Teaching professionals in Argentina and beyond were the driving forces to meet new challenges prompted by the COVID-19 pandemic.

This volume includes a selection of 11 papers presented at the conference that were recommended for publication in the double-blind review procedure. The resulting selection was grouped into four categories:

A Assessment of specific areas in different educational contexts

B Perceptions and emotions about assessment practices

C Involving learners in assessment

D A retrospective view of assessment practices

The first part comprises five papers dealing with the assessment of different language areas in three educational contexts in Argentina: secondary school, university and continuing teacher development (CTD). Several research orientations are also noticeable, including empirical applied research, some following qualitative (Montemayor Borsinger et al., Zinkgraf & Fernández), others quantitative (Garay & Giménez) designs, action research (Pallero) and research and development (Caielli et al.). The paper by Pallero focuses on the assessment of intercultural competence in a secondary school context. To this end, the author carried out an action research project in which learners were introduced to three teaching sequences with an explicit intercultural focus. The impact of the intervention was assessed through a combination of techniques including different students' productions and their reflections on the experience, as well as the teacher's field notes of the classroom activities and of focus-groups interviews. Questions were developed based on Byram's (1997) model to search for evidence of intercultural learning and data from the reflections and interviews were coded in search for recurring themes. Different dimensions of interculturality were revealed suggesting development of this competence to different degrees.

In the second paper, Caielli, Regueira and Williams describe the development and implementation of a large-scale EFL reading comprehension test. Their aim was to measure the extent of this skill development in learners at the end of secondary school (year 12) in state and private schools in the city of Mar del Plata. The authors base the principles and design of their test on national and international assessment instruments that were implemented in Argentina. The results, though limited in scope, offer useful insights into learners' achievement in a particular context compared to the curriculum aims, as well as as empirical data that could be used in CTD sessions.

The two following papers focus on EFL majors. Montemayor Borsinger, Coria and Porto studied students' development of pluriliteracies and how these were assessed by getting students to produce an alternative book cover for a novel. To do so, they deployed different semiotic modes (pictures and texts) whose effectiveness was assessed by making use of qualitative assessment techniques. Drawing on categories from Systemic Functional Linguistics, the authors looked into the use of processes and lexical chains that made the students' productions more or less effective. The results show how students resort to different ways of meaning-making, thus centering not exclusively on linguistic systems that are the usual focus of assessment in written production.

In the field of pronunciation assessment, Garay and Giménez set out to explore the differential effects of using two frameworks of reference, namely Jenkins' (2000) English as a Lingua Franca and the Lingua Franca Core (LFC) and Cruttenden's (2008) International English sets of core features, as compared to the current assessment procedures used in a pronunciation course of a university EFL English programme. Differences were found in the three measures, with the two international sets of core features displaying lower degrees of penalization of errors than the local rating scheme. Although the baseline for the assessment of intelligibility is similar in the three criteria, the standard used in the university course was deemed to be suitable in view of the aim of training future teachers and translators.

The last paper in this section reports the assessment of a CPD course for practising EFL teachers on the pedagogical uses of Formulaic Sequences (FS). In it, Zinkgraf and Fernández discuss instances of formative and summative assessment with reference to the tasks that course participants engaged in. Excerpts from the participants' reflections allowed insights into their understanding of the concepts introduced and of their application in materials design.

The papers in the second section of this volume deal with perceptions and emotions about assessment practices in different higher education settings. The paper by Carrera, Cad and Salinas reports on a study conducted in a second-year writing course for English majors at Universidad Nacional de Córdoba, Argentina. The authors investigated how the use of an analytic scoring scale together with teacher written comments influenced the academic writing skills of the students and their perceptions about the type of formative feedback received. In order to do this, the researchers administered a questionnaire to students, conducted focus group interviews, and analysed the students' first and second drafts of their written productions. Results indicate that the participants considered that the analytic scoring scale provided important information about what was expected of them in the written assignments of the course, that they could serve as a self- assessment tool and that they were a good complement to the comments that the teachers provided to their work. In terms of academic writing development, the analysis of the written productions shows that they improved in terms of content, organization and layout.

The paper by Aguilera, Flores Murgo and Morales Morales describes the creation of a placement test designed to determine what mandatory

English courses students should take at University of O'Higgins, Chile. The placement instrument tests reading comprehension, listening comprehension and use of English (grammatical and lexical understanding). The authors conducted a study to learn about students and teachers' perceptions of the test and the test design process. Data were collected using different surveys administered to university students (after taking the placement test as part of the study), teachers in the English program and teachers who had participated in the design of the placement instrument. In relation to the test, the findings show that the students highlighted the quality of written and oral texts as a strength of the test and the format of the items as one of its weaknesses. In addition, teachers in the English program also identified layout and format as an aspect that needed to be improved. These results informed the subsequent revision of test items. In terms of the process of test design, the results show that collaborating in the creation of the placement test allowed teachers to revisit their beliefs about teaching and learning.

The last paper in this section focuses on the relation between test anxiety and test results in the context of a private university in Guadalajara, Mexico, where it is required for undergraduate students to take an international standardized test such as the TOEFL, IELTS or TOEIC in order to certify their level of English proficiency. In the paper, Santana describes a quantitative study conducted to establish correlations between test anxiety and results in the TOEIC (Test of English for International Communication). To this end, the researcher administered an anxiety questionnaire to students before they took the TOEIC. Results indicate that the mean score obtained by participants in the TOEIC was lower than what the university required and that participants showed little evidence of test anxiety before taking the test. However, there is a moderate negative effect of anxiety on test results that indicates that the greater the anxiety, the lower the TOEIC test score.

In the third section of the volume, authors present ways in which language learners can be involved in the process of assessment. In his article, Moya discusses the benefits of a participatory, dialogic approach to assessment and presents the case of an English for Specific Purposes course in which students and teacher redesigned the final test in the course to accommodate the learners' needs during the COVID-19 pandemic. The test designed was understood as a learning task divided into different stages: revision, compilation of test items, definition of assessment criteria, taking the test and responses to feedback. The author highlights how this process allowed students to exercise their agency and promoted a sense of ownership of the learning experience.

The article by Ortega describes two formative assessment tools: exit tickets (usually issued at the end of a lesson for students to reflect on its content) and choice boards (which provide students with choices of activities that they can/would like to do). The author provides examples of tasks used in an online course for teenage students at a private language school that were designed considering different levels of cognitive development and explains some of their benefits for students and teachers.

The last section looks into assessment practices of the past. In his paper. López Barrios discusses approaches to EFL assessment and their relation with pupil promotion practices at three points in the history of secondary school education in Argentina. The author concludes that innovative practices that placed learning at the forefront were short lived and followed by evaluative practices and promotion policies that aimed at penalizing rather than rewarding performance.

The papers in this volume address the complexities of assessment in language education by describing and analyzing a variety of assessment practices in different educational contexts. In so doing, they invite readers to reflect about assessment possibilities in their own contexts and imagine ways in which assessment could contribute to richer learning and teaching experiences in language education.

Note:

The editors have worked with the authors in the review process by mostly addressing paper clarity in terms of content and organization. We have also focused on issues in relation to word choice and sentence structure as well as other proofreading aspects such as misspellings and punctuation. Despite all the efforts, we acknowledge that readers may find aspects that need improvement.

About the paper contributors and editors

Ana Cecilia Cad is a lecturer in English Language and Phonetics and Phonology at Facultad de Lenguas, Universidad Nacional de Córdoba. She also works in EFL methodology in a tertiary institution. She holds an MA degree in Applied Linguistics and an EFL Teaching degree. Her research interests include teacher education, technology, action research and multimodality. anaceciliacad@unc.edu.ar

Elisabet Caielli (Mg.) is a researcher and associate professor at the EFL Teacher Education Programme, UNMdP. She has participated in academic experiences as visiting professor at the University of Bath. Her research interests include reading, EFL teaching and learning in disadvantaged contexts and the legitimization of local pedagogies. She is also a school teacher and member of the Teacher Classification Board at the municipal system of education, Mar del Plata. ecaielli@hotmail.com

Paul Carrera is a university professor who is doing his Master's Degree in English with a major in Applied Linguistics. He is currently working at Facultad de Lenguas, UNC, as teacher assistant in the courses English Language I and II. He also teaches reading comprehension courses in English to undergraduate students of Engineering, Chemistry and Economy. His main areas of interest are SFL genre pedagogy, corpus linguistics and technology applied to teaching EFL. paul.carrera@unc.edu.ar

Rosa Cerda Aguilera holds a degree in English Linguistics and Literature, and English Language Education from Universidad de Chile, and an MA in Applied Linguistics for TESOL from Durham University. She coordinates the English Language Program at Universidad de O'Higgins. She is interested in teacher development and material adaptation and creation.
rosa.cerda@uoh.cl

Ana María Coria is Teacher/Translator of English (UNLP). She is Head Teacher of Capacitación en Inglés and Associate Teacher of English Language II (UNLP). She holds a Postgraduate Degree in Reading Comprehension in L2

(UBA) and a Master's Degree (UNLP). Her research interests focus on Systemic Functional Linguistics. amfuseicab@gmail.com

Gabriela Fernández (Mg.) is the head of English 4 and a teacher assistant for English 1 for the Teacher and Translator Training Course at Facultad de Lenguas, in Universidad Nacional del Comahue. She has a Master's degree in Teaching English as a Foreign Language (Jaén University, Spain). She investigates the acquisition of formulaic sequences by EFL primary school students as a member of research project J033. g_v_fernandez@hotmail.com

Ingrid Flores Murgo holds a degree in English Language Education from Universidad Católica del Maule and is currently studying an MA in Neuroscience and education at Universidad Mayor. She is an adjunct teacher at the English Language Program at Universidad de O'Higgins. Her interests are neurolinguistics and pedagogical support. ingrid.flores@uoh.cl

María Garay is an assistant professor of English phonetics and phonology in the teacher and translator training courses at the School of Languages, Universidad Nacional de Córdoba, Argentina. She holds a postgraduate degree in Foreign Language Didactics. Her main research interests revolve around second language pronunciation, discourse analysis with a focus on discourse intonation, and materials design. maria.garay@unc.edu.ar

Florencia Giménez is a professor of English phonetics and phonology in the graduate and undergraduate programmes at the School of Languages, Universidad Nacional de Córdoba, Argentina. She holds an MA degree in English and Applied Linguistics. Her main research interests revolve around second language pronunciation, and discourse analysis with a focus on discourse intonation. fgimenez@unc.edu.ar

Mario Luis López-Barrios holds a Dr. phil. (Universität Kassel, Germany) and a degree in Language Education in English and German (Universidad Nacional de Córdoba, Argentina). Professor of Foreign Language Teaching (School of Languages, UNC). Areas of interest: Foreign Language Teaching, Materials Development, Research Methods in Applied Linguistics and Second Language Acquisition. mario.lopez.barrios@unc.edu.ar

Ann Montemayor-Borsinger holds a Master degree in Applied Linguistics (Bristol) and a PhD in Linguistics (Glasgow). She is Professor of Linguistics at UNRN, and Invited Professor on different postgraduate programs. Interests: Discourse Analysis in English, Spanish and French, drawing on Systemic Functional Linguistics. aborsinger@unrn.edu.ar

Gabriel Morales Morales holds a degree in English Linguistics and Literature from Universidad de Chile, and MA in TEFL from Universidad Metropolitana de Ciencias de la Educación. He is an adjunct teacher at the English Language Program at Universidad de O'Higgins. His research interests are lexical and textual aspects in TEFL materials. gabriel.morales@uoh.cl

Mario Moya is a Doctor of Education (Applied Linguistics) leading the Master's courses in English Language Teaching and Master in Education at the University of East London. His specialism is language education and teacher training. Mario works as an educational consultant for the British Council in academic literacies and English as a medium of instruction. mario.moya@uel.ac.uk

Luciana Ortega is a teacher of English and Licenciada en Lengua y Literatura Inglesas graduated from IES "Olga Cossettini" and Universidad del Centro Educativo Latinoamericano (UCEL). She is currently a student in the Master's programme in Linguistic Theory and Language Acquisition at Universidad Nacional de Rosario. Luciana works at "Asociación Rosarina de Cultura Inglesa" and Instituto Superior "San Bartolomé". luaortega23@gmail.com

María Pía Pallero is a teacher of English with ten years' experience at secondary and tertiary levels, currently working as a high-school language instructor and teacher trainer. She holds an MA in Foreign Language Teaching (Michigan State University), and a post-graduate Specialization in Education (UNICEN). She was awarded the Fulbright Foreign Language Teaching Assistant Scholarship (FLTA) 2015-2016. mpiapallero@gmail.com, pallerom@msu.edu

Melina Porto holds a Master's degree in ELT (Essex University), a PhD in Sciences of Education (UNLP) and a postdoctoral degree in Humanities and

Social Sciences (UBA). She is a researcher at CONICET, Professor at UNLP and Honorary Research Fellow at the University of East Anglia. Interests: intercultural citizenship education. melinaporto@conicet.gov.ar

Ana Lía Regueira, M.A is Director of research and holds a Chair at the EFL Teacher Education Programme, UNMdP. She has directed projects on teacher education, and language teaching and learning. She has taught postgraduate courses at UNMdP and at the off-site MA in ELT, King's College, University of London. She has participated in academic and research experiences as a visiting professor at the University of Bath and in secondary schools in the UK. aregueir@hotmail.com

María Elisa Romano holds a PhD degree in Language Sciences and is currently Head Professor of English Language II at Facultad de Lenguas, Universidad Nacional de Córdoba. She has conducted research on academic writing for 20 years and is currently coordinating a project on writing assessment. maria.elisa.romano@unc.edu.ar

Fabiana Sacchi (PhD in Foreign Language Education, The University of Texas, USA) is a teacher educator and researcher in the EFL teacher education program at Universidad Nacional de Río Cuarto, Argentina. Her research focuses on affective factors in language learning/teaching and intercultural, peace, and social justice education. fsacchi@hum.unrc.edu.ar

Julieta Salinas is a lecturer in English Language at Facultad de Lenguas, Universidad Nacional de Córdoba. She is currently working in her MA Thesis on Language and Intercultural Studies. Her research interests include writing, assessment and technology. jusalinas@unc.edu.ar

Josefina C. Santana has been involved in English language teaching, teacher training, course administration and materials development for over 35 years. She holds a PhD in Education from ITESO University in Guadalajara, Mexico. She is a full professor at Universidad Panamericana, located in the same city. She is a member of the Mexican National System of Researchers, and her research interests include cognitive and affective processes of language learning. jsantana@up.edu.mx

Jenifer Williams (Esp.) is a researcher and adjunct professor at the Teacher Education Area, Profesorado de Inglés, UNMdP. She has participated in academic and research experiences as a visiting professor at Purdue University (USA) and University of Bath (UK). Her research interests include cognition and teacher education, affective variables that influence SLA, and reading comprehension development. Currently, she is studying the legitimization of local pedagogies and teacher agency. jwilliam@mdp.edu.ar

Magdalena Zinkgraf (Mg.) teaches English 3 and the Writing workshop at the Teacher-training course in Universidad Nacional del Comahue She holds an MA. in ELT and Applied Linguistics, University of London. At Facultad de Lenguas, she leads the J033 research team, which investigates formulaic sequence acquisition. She has participated in numerous conferences and published some articles in national and international journals. maguizinkgraf@gmail.com

Assessing Intercultural Competence at Secondary Level: An Action Research Study

María Pía Pallero

Escuela Nacional "Adolfo Pérez Esquivel" (UNICEN)

This article reports an action research study on the assessment of intercultural competence at secondary school level which was carried out at an English as a Foreign language (EFL) class in Argentina. The aim of the project is to analyse the impact of applying a combination of assessment tools to measure intercultural development. The questions that inspired this work are:

- Can a combination of techniques favour the assessment of ICC skills in this teaching context?
- Are reflection tasks in L1 a useful tool for ICC assessment?

The following sections describe the investigation process and evaluate the effectiveness of the selected assessment tools. Finally, pedagogical implications are discussed.

Theoretical Framework

Intercultural Communicative Competence

Cultures constitute complex situated processes within which individuals establish their identities; they are "dynamic and emergent" (Liddicoat and Scarino, 2013, p. 21). This means that who we are is a product of our interaction with other members of our own cultural community as well as with subjects from other communities. In this regard, Ryan (2003) claims that the "paths of experience" (p. 131) we go through as we develop our identities lead us to get in contact with otherness in different ways. Hence, having a receptive and open-minded attitude toward cultural differences is a prerequisite to the development of the necessary skills to cross cultural boundaries. This is what interculturally competent individuals do: they embrace cultural diversity as a source of personal growth. They have the "ability to ensure a shared understanding by people of different social identities" (Byram, Gribkova & Starkey, 2002, p. 10) and interact with others acknowledging their complexity and multiple identities as well as their own. In other words, the intercultural person seeks shared universal cultural processes or *threads* and links them to their personal trajectories or *blocks* (Holliday, 2016).

In view of these concepts, there is a set of attitudes, values, and skills that intercultural speakers need to develop, which are described in Byram's (1997) model of *intercultural communicative competence* (ICC), recently revisited (Byram, 2021). Knowledge of self and others along with attitudes of curiosity and openness (*savoir être*) are considered preconditions for successful intercultural interaction. They also, and most importantly, entail the ability to *decenter*, i.e. to relativize one's own views and suspend disbelief. Similarly, ICC involves skills of interpreting and relating (*savoir comprendre*) which enable individuals to explain events from another culture by relating them to those from their own, and skills of discovery and interaction (*savoir apprendre/faire*) which are defined as the capacity to acquire knowledge of a culture and operate it in actual communication. Last but not least, it is important that the ability to critically evaluate the products, practices and perspectives of a given community, one's and others' values, known as critical cultural awareness (*savoir s'enganger*), be the ultimate goal of intercultural communication (Byram, 1997, 2021; Byram, Gribkova & Starkey, 2002).

Intercultural Education and Assessment

In second and foreign language teaching (FLT) pedagogy, there has been increasing interest in how interculturality is achieved and can be stimulated in the language class. Byram (1997) has discussed the central role of FLT in the "experience of otherness" (p. 3) as he argues that learners engage with another language as a medium to cope with familiar and unfamiliar experiences. Likewise, scholars like Furstenberg (2010) have claimed that intercultural competence should be the main objective of the language class. Language learners are expected to become *intercultural speakers* or *mediators* (Byram, Gribkova & Starkey, 2002) who are not only able to participate in interactions in the target language (TL) but are also capable of engaging in processes of interpretation. Thus, intercultural speakers act as mediators for other people by providing the means to understand diverse others to those who are not able to make such interpretations (Liddicoat & Scarino, 2013). It is worth pointing out that the intercultural dimension in language teaching does not disregard the development of the linguistic competence learners need to become skilled users of the TL.

In order for language learners to develop the expected attitudes and skills, instruction needs to be planned accordingly. The pedagogy towards the implementation of an intercultural approach in formal educational settings involves determining learning outcomes in light of the components of intercultural competence (Council of Europe, 2014). In this regard, Corbett (2003) claims that "the intercultural language course involves the designing and implementing of tasks which encourage the learner actively and systematically to seek cultural information, which then impacts upon his or her language behaviour" (p. 33). These pedagogical practices are expected to be *communicative tasks* (Nunan, 2004) - i.e. integrative units that bring together participants, purposes, contexts, and interaction in the pursuit of particular goals - which can be adapted by adding an *experiential dimension* - i.e. learning through experience, in either real or imagined situations stimulated in class (Liddicoat & Scarino, 2013). That is, teachers need to implement tasks that provide learners with opportunities for raising intercultural awareness as they interact, interpret and reflect upon culturally specific patterns of behavior through their developing frame of reference.

Liddicoat and Scarino (2013) argue that the assessment of intercultural competence is an interpretative process of enquiry into the meanings learners assign to their experiences. Therefore, it is necessary to employ a variety of techniques that can account for advances that go beyond observable performance. Deardorff (2006, 2011; Hua, 2016) suggests that it is to be measured in its separate components through a variety of methods.

One of her studies (2006) indicates that some of the qualitative methods scholars in the field have expressed a preference for are student interviews and judgement by self and others, followed by other widely accepted assessment tools such as analysis of narrative diaries, self-report instruments, focus groups, dialogues, and workshops. Similarly, Fantini (2009) points out that a combination of assessment types, formats and strategies should be used in order to determine progress in intercultural growth. He also contributes that the learner should be involved in the evaluative process through self-evaluation, reflection, and feedback.

State of the Art

Recent studies - most of which have been conducted in higher education settings - confirm that foreign language learners can make progress in their intercultural competence when instruction is intentionally designed and implemented to that end. An example of this is a project carried out by Toyoda (2016) in which advanced students of Japanese in Australia evidenced growing tolerance and flexibility in their thinking after taking a blended course in which intercultural awareness was fostered through the materials and the activities employed. The investigation that Gómez (2018) conducted with a group of undergraduate students of English at a Colombian university also suggests that materials selection plays an important part in the development of intercultural skills and attitudes. By participating in conversation clubs in which authentic international news was shared and discussed, these students were encouraged to acquire knowledge, enhance critical skills, and generate attitudes like empathy for other human beings.

Even though similar research is scarce in secondary education, there is a study undertaken by Moloney and Harbon (2010) in high school foreign language learning classes in Australia, whose structure and findings are considered valuable for the purposes of the present work. Results confirm that intercultural learning is achievable through task design and reveal that intercultural competence can be identified and measured if students are given the chance to demonstrate it as they complete intercultural tasks.

Finally, this study gives evidence that students benefit from connecting language instruction with their personal experience and reflecting on it.

The data collection methods employed in the investigations that were previously examined represent assessment tools of different kinds and illustrate the need for a multiplicity of techniques. In all the studies described, observation notes or field notes were used and then analyzed in search for patterns or themes. Whereas video diaries and end of semester surveys were administered by Toyoda (2016), artifacts (student worksheets) constituted a second data collection instrument in Gómez's (2018) research as well as a personal questionnaire administered once the conversation club had finished. Moloney and Harbon (2010) adopted focus group interviews instead. Taking these options into consideration, a variety of assessment tools are adopted in this project as well.

Participants

The participants in this study were a group of 17 Spanish speaking students aged 14-16 who were learning English at a state-run secondary school in Argentina. They had already received three years of language instruction in lower secondary, and they were taking the 'level 2 class'. According to the Common European Framework of Reference for Languages (see Council of Europe, 2001, 2018), their level of performance was elementary, A2.

Methods

The present investigation constitutes a qualitative study in which *action research* (AR) is the approach adopted. A reason for this methodological decision is that, as Burns (2010) states, AR involves teachers in a process of evaluation and reflection of their own teaching in which they identify problematic issues worthy of investigation and plan their intervention deliberately to bring about "critically informed changes in practice" (p. 5). It is a systematic and self-reflective type of inquiry. It should also be pointed out that AR constitutes a small-scale, contextualized type of inquiry which does not seek to generalize the obtained results; a characteristic that conforms with the ends of teacher-led research.

Following Kemmis and McTaggart (1992), the four broad phases of planning, action, observation, and reflection were followed. The initial phase consisted of identifying the problem or area of enquiry, generating possible research questions, planning instruction and data collection methods, and obtaining necessary permissions. During the action phase, planned activities were implemented over a term as data were collected in the form of observation, field notes, artifacts and reflection papers. The field notes were used to register the researcher's observations along the process. Reflection took place at the time the information collected was analyzed and interpreted. Eventually, the whole cycle was evaluated with the intention of identifying strengths and weaknesses that could shed light on future instruction.

Planning and Implementing Instruction

Three units of work were designed with the aim of producing gains in the participants' development of intercultural competence within the framework of *Task Based Language Teaching* (TBL) (Nunan, 2004). All three units were taught during the first term of the school year. Each of these units included the contents indicated in the curriculum created by the Foreign Language Department in the institution. In all of these units, there were several activities that were culturally oriented besides three 'intercultural lessons' in which we will be focusing on in particular.

Intercultural lesson 1 (IL1)

This lesson was inspired by an activity proposed by Corbett (2010) named 'Television viewing habits' (p. 61), but the topic was changed into teens' habits in their free time. During this lesson, learners explored the different ways in which people their age use their leisure time. They shared their own choices and compared them to those of the rest of the class. They listened to and reflected upon what a group of American students choose to do in their spare time as well. Learners finally interviewed classmates to find out about their use of free time.

Intercultural lesson 2 (IL 2)

This lesson was based on Corbett's (2010) activity called 'National dishes: cultural associations' (p. 197-199). In it, learners described Argentina's national dish, a task that required their reading about its origins and the development of some specific vocabulary. They then investigated the characteristics of one national dish from an English-speaking country as well as practices associated with it. They finally established comparisons between dishes and shared their conclusions through posters. As a communicative task, they created cooking videos using recipes of pastries traditionally prepared in English-speaking countries.

Intercultural lesson 3 (IL 3)

In this lesson, learners worked in groups to analyse a local legend associated to a tourist destination. One of them was, for example, the legend of the origin of the Iguazú Falls. They told the story to their peers, took down notes and compared all the legends. They were finally asked to create a flyer for tourists to learn about the legend related to the tourist attraction they were visiting.

Data Collection

Direct and indirect assessment of intercultural competence (Deardorff, 2006, 2011; Hua, 2016) was applied to account for both students' production as a form of evidence of their learning, and also their perception of learning. To analyze the former, data were collected through artifacts produced in L2 (worksheets, posters, and videos), and teacher's field notes. Learners' reflections in L1 and focus group interviews, also conducted in Spanish, are considered evidence of the latter. In the sub-sections that follow, each of the data collection methods is further discussed in terms of how evidence was obtained and what it comprises.

Artifacts

Evidence of students' work in the TL was obtained through the worksheets that learners completed during the intercultural lessons, as well as the posters and videos they created afterwards. One sample worksheet per group was collected for analysis. The participants also produced short reflections in Spanish (about 50 words long) in which they expressed what they had learned and their opinion about the activities.

Observation

Field notes were collected every class so as to gather evidence of learners' work and reactions during the whole process, and also in order for the teacher-researcher to keep a record of the difficulties the participants and herself faced during the project. Anecdotal information was included when considered eloquent as representative of intercultural development or as indicating strengths and weaknesses in lesson design.

The notes were kept in a digital format. They consist of a chart that contains the description of the activities that were done (date, activity type, teaching point), and a comments section right beside each activity. These comments involve descriptions of classwork from the teacher's standpoint in which the participants were often quoted.

Focus group interviews

Two focus group interviews were conducted by the teacher at the beginning of the second term. The language used was Spanish. Each of the focus groups consisted of 5 learners who were randomly selected to participate. The interviews were carried out in the classroom, during class time. As a result, 13 minutes of recorded material were obtained.

Data Analysis

The data obtained were analysed separately in search for emerging patterns. Then, the procedure of *triangulation* (Richards, in Dean Brown & Commbe, 2015) was used to compare resulting interpretations. It served to look at participants' progress in each of the areas of ICC identified in Byram's model (1997). To this end, the following rubric was used (table 1). A detailed account of the procedures adopted for the analysis of the data is established in the next segments.

Table 1

Rubric for the identification of ICC development

Skills	Evidence
Knowledge	Is there evidence of learners' increasing knowledge about other cultures?
Attitudes	Do participants show willingness to learn/adjust?
Reflection	Is there evidence of impact on self/own identity?
Comparisons/ Connections	Are learners able to make comparisons/connections across individuals, cultures, history, etc.?
Interaction skills	Are learners able to respond differently based on context and interlocutor? Is there evidence of awareness of the need to adjust according to the audience?
Investigation skills	Do learners ask the right questions? Are they able to find an informant / Do they know where to look?
Critical cultural awareness	Is there evidence of learners' ability to critically evaluate the products, practices and perspectives of a given culture, one's and others' values?

Coding of Qualitative Data

A total of 48 reflections in Spanish were transcribed and a document was created for each student containing the three reflections. Similarly, recorded material obtained from the focus group interviews was transcribed for further analysis. All transcripts were coded by means of qualitative analysis software, but the codes used were decided by the researcher in the process that is described as follows.

After several readings, a first coding stage consisted of defining descriptive codes (Taylor-Powell & Renner, 2003). Those were classified into unit, activity type, reactions, and culture words. A complete record is shown in tables 2 and 3 below. It must be noted that the reason for codes under 'reactions' and 'culture words' to be in Spanish is that participants' words were used when deciding these categories of analysis.

Then, following Saldaña (2012), pattern coding was employed to identify similarly coded data and give it a meaningful organization in a second round of analysis in terms of the research questions. Finally, similar patterns were grouped based on the rubric for the identification of ICC development (table 1). At this point, the two reflections that were written in English were analyzed separately.

Table 2

Codes

Levels 1 and 2	
Unit	Activity type
Free time	sharing pastimes
	video US students
Food	investigate
	cooking show
Legends	Argentine legends
	whole unit

Table 3

Codes

Level 3 codes	
Reactions	*saber/ conocer* (to know) *aprender* (to learn) *pensar* (to think) *comparar* (to compare) *similitudes/ diferencias* (similarities/ differences) *relacionar* (to relate) *bueno/ divertido* (good/ fun) *interesante* (interesting) *compartir* (to share)
Culture words	*otras personas* (other people) *otros países* (other countries) *tradiciones/costumbres* (traditions/customs) *la cultura* (the culture) *las/otras culturas* (cultures/ other cultures) *nuestra cultura* (our culture)

Analysis of Artifacts and Observation Data

Participants' productions in the TL were examined to account for gains in knowledge, the ability to establish comparisons and adjust according to the audience, and the impact on the learners' own identity. The teacher-researcher isolated examples from the data manually and grouped them based on the rubric for the identification of ICC development (table 1).

Likewise, the field notes were examined in search for meaningful information or events that could reveal growing ICC. The analysis focused on finding examples of learners' investigation skills and their attitudes while completing the activities. Instances of the impact of the tasks on their identity were also examined as well as examples of opportunities in which they critically evaluated cultural products, practices, and perspectives. Difficulties in the learning process and teacher's interventions were highlighted when considered relevant to account for intercultural development or to be worth mentioning in the pedagogical implications section.

Findings and discussion

After analysing the data obtained through direct and indirect forms of assessment, the teacher-researcher determined areas of growth in intercultural understandings. The results of the assessment process are reported below.

The participants in the study were able to develop the basic intercultural skills of knowledge and attitudes of openness and curiosity, also known as *savoir être*. Learners gained cultural knowledge about others: their peers, people from other countries, and communities in their country. At the same time, they deepened their understanding of products in their culture and the practices that go along with them, such as customs around *asado* (a type of barbecue), Argentina's most traditional food. Most important of all, they all showed a receptive attitude in their attempts to understand diverse cultural perspectives in every activity that was brought to them.

The use of the words *saber/conocer* [to know] and *aprender* [to learn] in reflections in the L1 could be considered evidence of students' perceived gains in knowledge. In reference to the intercultural activities about free time, 16 out of 17 participants used either of these words. For instance, one of the students pointed out, "*aprendí qué hacía cada uno de mis compañeros en su tiempo libre* [I learned what each of (my) classmates did in their free time]". Besides, 11 students used the terms (to learn/to know) as they reflected on the lesson related to food, and all of the students mentioned at least one of them when referring to their work on legends. For example, "a*prendí a ver cómo los pueblos originarios explicaban las cosas que no entendían* [I learned how indigenous populations explained what they could not understand]".

Not only did students display the ability to decenter as they tried to comprehend how indigenous populations explained the world surrounding them through stories, but they could also connect the legends with their own cultural identity. The words "*identidad* [identity]", "*raices* [roots]", and "*antepasados* [ancestors]", which appear in learners' reflections on unit 3, illustrate this point. A student claimed, *"con ellas pudimos aprender más sobre nuestras raíces, culturas argentinas, y a su vez la de otros sitios* [with them (the legends) we could learn more about our roots, Argentine cultures, and at the same time those of other places]". Another participant considered the study of the legends an opportunity to reflect upon the self. Reading these stories helped her "understand who we are today, why we are this or that way (*entender lo que somos hoy en día, por qué somos de tal manera)*".

Data show that learners were able to relativize their own views in the first intercultural lesson as well. One of the participants indicated, "*aprendí que no todos hacemos las mismas cosas en todos los países* [I learned that not everyone does the same things]". They gained awareness of the fact that what teenagers do in their free time depends on their personality and interest, the context where they live and their religion, among other factors. In like manner, there is evidence that demonstrates how stereotypes were deconstructed when learners expressed, for instance, that not every American teenager uses their leisure time the same way, or that what the movies show in this respect is not true for every teenager in the U.S. "[*N*]*o todo es como nos muestran en las películas* [not everything is like they show us in movies]", concluded one of the students.Even though all the participants established cultural comparisons in class, only some of them could relate cultural content to their own experience and identity. In their worksheets learners expressed, for example, "*Asado* and haggies [Scottish traditional dish] are accompanied by potatoes and spices" or "After eating these dishes, people watch TV"; they established differences such as "People in Argentina eat *asado* on special occasions, but people in Australia eat meat pie daily" or "*Asado* is a meat dish but cawl stew is a vegetable dish". It could be noted, however, that those who were able to link the local legends to trips taken in the past or a student who found out about her favorite writer's meal, made connections at a deeper level, thus showing a greater impact on their identity. Yet, the fact that all of the learners connected familiar cultural artifacts to unfamiliar ones in the TL constitutes a valuable step forward in what Byram (1997) calls the 'experience of otherness'.

It cannot be stated that learners highly developed the ability to search for cultural information and operate it in actual communication (*savoir apprendre/ faire)*. The main reason for this being their lack of experience in these practices, they needed the teacher's assistance to learn where to find reliable material and what to say or do with it in particular communicative contexts. However, it is possible to state that the students could look for cultural information in the TL and analyse sources with the aid of the teacher, as well as decide how to include this information in their own production with full awareness of audience. Besides, they were able to apply the linguistic exponents studied in class, which are socially and culturally significant, into their speech in communicative tasks. It is worth noting, though, that these individuals did not need to display the ability to "manage dysfunctions which arise in the course of interaction, drawing upon knowledge and skills" (Byram, 1997, p. 38) since the interactions they participated in were simple and they could prepare for them in advance.

With regard to interaction skills, it could be asserted that the way learners decided to design their leaflets about Argentine legends shows cultural awareness. It is evident that they were conscious of the need to provide the reader with background information before retelling the stories. For instance, one of the flyers included the following sentences "Prior to the arrival of the Spanish, the Guarani population explained the origin of the Iguazú Falls with a legend". Images were added to aid comprehension, these included a photo of the phenomenon or place the legends referred to and the setting of the story was mentioned. Figures 1 and 2 show examples of student artifacts.

In Figure 1, there is a map and the flag that represents indigenous communities in Argentina, the *Whipala* which serve as visual aids. In addition, the introductory lines indicate the setting of the story: "In the Tandil mountains, there is a rock that is upside down. The native peoples explained this with the legend of the Sentinel".

Figure 1

Flyer for tourists based on the Legend of the Sentinel

In the second image (Figures 2.1, 2.2), the students' intention to make their flyer clear and attractive to the audience is manifest in the choice of pictures and words. In Figure 2.1 there is a map and a real picture of the rivers whose origins the legend describes, and there is a question "Black River. Why black?" to arouse curiosity about the content of the story. Figure 2.2 contains pictures that illustrate key vocabulary which facilitates comprehension.

Figure 2.1.

Flyer based on the legend of rivers Neuquén and Limay (front)

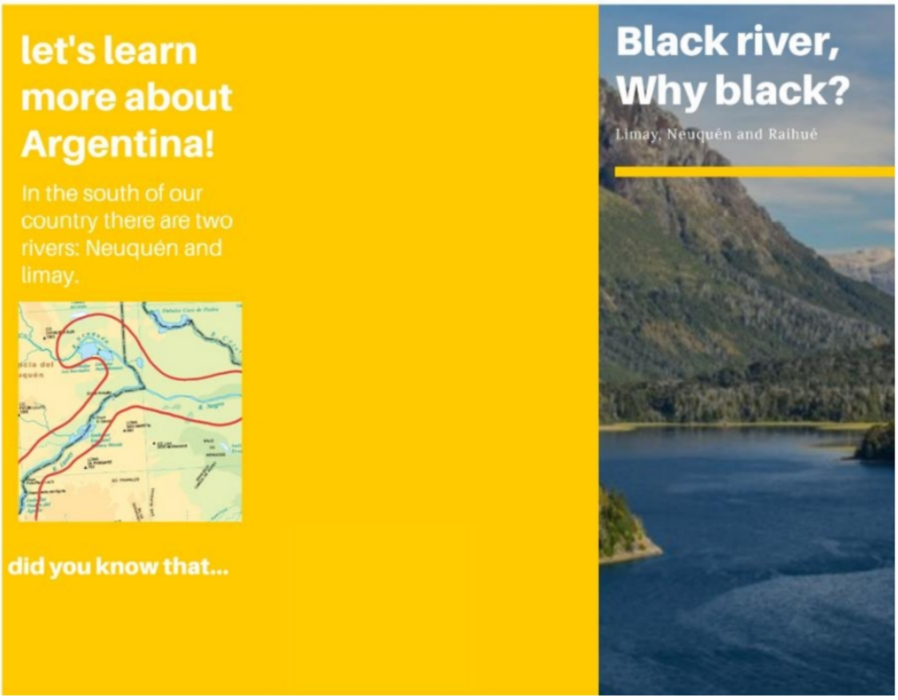
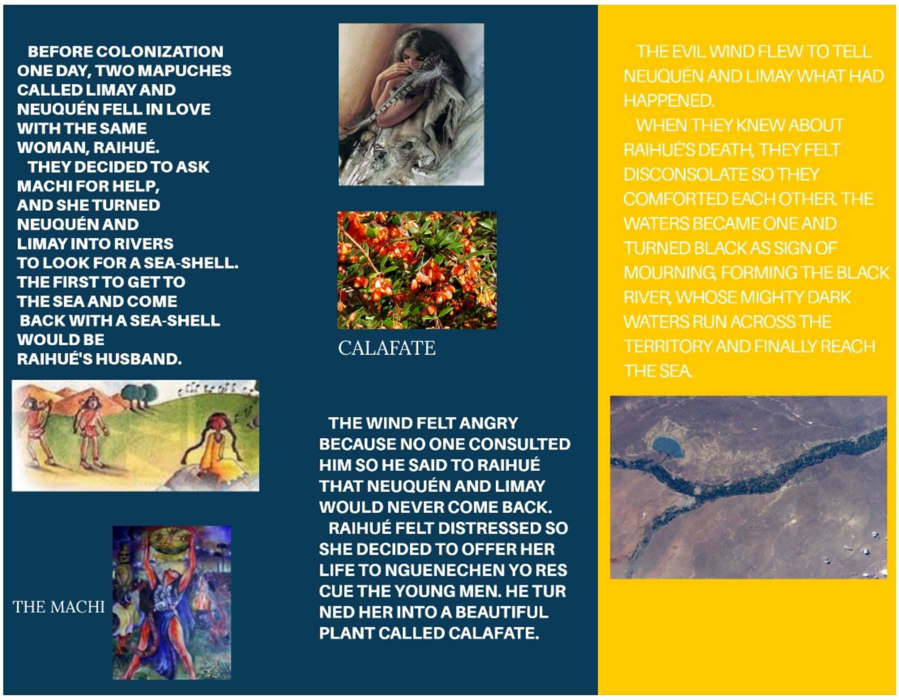


Figure 2.2.

Flyer based on the legend of rivers Neuquén and Limay (back)



No evidence of growth in learners' ability to critically evaluate cultural artifacts *(savoir s'engager)* was found, probably due to the fact that it was not required in the tasks. These activities focused on providing learners with opportunities to gain knowledge, decenter and compare products and practices. Thus, students did not get to compare perspectives which they did identify and discuss in the L1. Had the students been exposed to the type of intercultural instruction proposed in this project before, the teacher would have included tasks that invited them to become better aware of their own values and their influence on how they view other people's standards in the TL.

That said, it must be mentioned that participants' ICC development differs from individual to individual. The generalizations described above are not applicable to every learner in the class in the same way. In spite of this, it could be argued that this group as a whole gained cultural knowledge, compared cultural artifacts, displayed an open attitude towards cultural differences, applied the cultural knowledge they gained from investigating their productions in the target language, and reflected about the whole experience.

A final aspect to point out is that the participants in the study improved their language performance and their intercultural understandings, but these two processes cannot be considered to have progressed in the same degree. This may imply that when we refer to intercultural communicative competence development at lower levels, we may need to analyse both competences separately. We could see in this study that learners were able to make progress in intercultural understandings that they could not fully manifest in linguistic behavior in the TL.

Given their proficiency level, learners were considered to lack the linguistic tools to express exactly what they were thinking about in the L2. This is why they were not asked to write reflection papers in English. Instead, the use of the L1 made it possible for them to "fully articulate their thoughts and ideas" (Furstenberg, 2010, p. 331) as they explained what they had learned after engaging in the activities.

When writing their reflections in Spanish, they could use their own words and they were encouraged to speak their minds. Still, it took them some thinking time to put those thoughts on paper. It could be observed that it was not easy for them to find the words to express what they could better understand or had discovered, possibly because they had little or no experience in engaging in cultural reflection.

Nonetheless, it is believed that after some training in the process of intercultural reflection, learners could be guided as to express their thoughts in the TL. The fact that a high achieving student decided to write her last two reflection papers in English could be interpreted in this sense, that is, as a question of readiness in terms of language development. Therefore, it could be discussed whether there is a possibility of scaffolding reflection in similar way comparisons were, to progressively enable learners to do so in the TL.

Conclusion and Pedagogical Implications

Measuring learner gains in intercultural competence in its separate components (Deardorff 2006, 2011; Hua, 2016) and the use of a combination of assessment methods (Fantini, 2009) in the process was a favorable choice. Thanks to the former methodological decision, the teacher-researcher was able to detect the exact areas in which learners made progress and in which they did not. By means of the latter, it was possible to gather data from students' performance and perceptions that she would not otherwise have had access to.

In this regard, the role of reflection in the L1 was found to be of key importance in the assessment process. Giving learners the chance to reflect in their first language was highly beneficial for them to communicate their thoughts to the full, devoid of the constraints of doing so in the target language. It favored the work of the teacher in that reflections provided further evidence of the students' evolving mindsets which were not evident in their performance in English.

Future research may focus on what Byram, Houghton and Porto call *intercultural citizenship* (2018). The claim that "when (foreign/second/world) language teaching takes into consideration its educational function as well as its instrumental purposes, its aims begin to coincide with some of the aims of contemporary citizenship education" (p. 485). This combination of the intercultural approach with civic action has the potential to effect changes in the community surrounding students. An example of this is a binational project combining intercultural and ecological citizenship conducted in a primary school setting in Argentina (Porto, 2016).

References

Burns, A. (2010). *Doing action research in English language teaching: A guide for practitioners*. Routledge.

Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Multilingual Matters.

Byram, M. (2021). *Teaching and assessing intercultural communicative competence*. *Revisited* 2nd ed. Multilingual Matters.

Byram, M., Gribkova, B., & Starkey, H. (2002). *Developing the intercultural dimension in language teaching: A practical introduction for teachers*. Council of Europe.

Byram, M., Houghton, S. A. & Porto, M. (2018). Intercultural citizenship in the (foreign) language classroom. *Language Teaching Research*, *22*(5), 484–498.

Corbett, J. (2003). *An intercultural approach to English language teaching*. Multilingual Matters.

Corbett, J. (2010). *Intercultural language activities*. Cambridge University Press.

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

Council of Europe (2014). *Developing intercultural competence through education.* Strasbourg: Council of Europe.

Council of Europe (2018). *Common European Framework of Reference for Languages: Learning, teaching, assessment companion volume with new descriptors.* Council of Europe.

Dean Brown, J., & Coombe, C. (Eds.) (2015). *The Cambridge guide to research in language teaching and learning*. Cambridge University Press.

Deardorff, D. K. (2006). Identification and assessment of intercultural competence as a student outcome of internationalization. *Journal of Studies in International Education, 10*(3), 241-266.

Deardorff, D. K. (2011). Assessing intercultural competence. *New Directions for Institutional Research 149*, 65-79.

Fantini, A. E. (2009). Assessing intercultural competence: Issues and tools. In Deardorff, D. K. (Ed.), *The Sage handbook of intercultural competence* (pp. 456-476). Sage.

Furstenberg, G. (2010). Making culture the core of the language class: Can it be done? *The Modern Language Journal, 94*(2), 329-332.

Gómez, L. F. (2018). EFL learners' intercultural competence development through international news. *Gist Education and Learning Research Journal, 16*, 185-208.

Holliday, A. (2016). Differences and awareness in cultural travel: negotiating blocks and threads. *Language and Intercultural Communication, 16*(3), 318-331.

Hua, Z. (2016). *Research methods in intercultural communication. A practical guide*. Wiley Blackwell.

Kemmis, S. & McTaggart, R. (1992). *Cómo planificar la investigación-acción*. Ediciones Laertes, Serie Pedagogía.

Liddicoat, A. J., & Scarino, A. (2013). *Intercultural language teaching and learning*. Wiley.

Nunan, D. (2004). *Task based language teaching.* Cambridge University Press.

Maloney, R., & Harborn, L (2010), Making intercultural language learning visible and assessable. In B. Dupuy & L. Waugh (Eds.), *Proceedings of the Second International Conference on the Development and Assessment of Intercultural Competence* (pp. 281- 303). CERCLL The University of Arizona.

Porto, M. (2016). Ecological and intercultural citizenship in the primary English as a foreign language (EFL) classroom: an online project in Argentina. *Cambridge Journal of Education, 46*(4), 395-415.

Ryan, P. (2003). Searching for the intercultural person. In Alred, G., Byram M., & Fleming, M. (Eds.), *Intercultural experience and education.* (pp.131-154). Multilingual Matters.

Saldaña, J. (2012). *The coding manual for qualitative researchers*. Sage.

Taylor-Powel, E. & Renner, M. (2003). *Analyzing qualitative data*. University of Wisconsin System.

Toyoda, E. (2016). Intercultural knowledge, awareness and skills observed in a foreign language classroom. *Intercultural Education*, *27*(6), 505-516.

A Local Experience with Reading: Criterion-Referenced Test Design and Validation

Elisabet Caielli, Ana Lía Regueira and Jenifer Williams

Universidad Nacional de Mar del Plata

The context of the test

As school teachers of English we have often wondered about the extent to which our students can interpret texts on familiar topics and genres by the end of state secondary school after nine years of English as a compulsory subject in the school curriculum (from 4th year primary school, age 9, to the last year of secondary school, age 18). This gave rise to a specific question: How can we measure their reading ability? Our interest in this particular context has also been linked to our roles as teachers and researchers of the EFL Teacher Education Area at Universidad Nacional de Mar del Plata, since the reality of the local schools where our students will pursue their professional careers, has always been at the core of our practice in both roles. Consequently, we are concerned with what happens in the classrooms in order to inform, modify and strengthen the development of future teachers.

By 2014 there were no official records about Argentine secondary school students' reading abilities in English as a foreign language, but there was some information regarding levels of reading achievement in the mother tongue through the results of some standardized tests such as the Programme for International Student Assessment (PISA) (PISA, 2009) and the Operativo Nacional de Evaluación (ONE) of Argentina's Ministry of Education (Dirección Nacional de Información y Evaluación de la Calidad Educativa [DiNIECE], 2013). Having explored some of the results and having more questions than certainties we started the research project *Development and validation of instruments for the criterion-referenced evaluation of reading comprehension in Secondary Education. Curricular area: Foreign Language (English)* which succinctly consisted in the design and implementation of a census test of reading in English. The project followed two main goals; first, to provide information on the levels of reading comprehension performance in English of secondary school graduates and, second, to strengthen and consolidate the capacities of the research team in relation to criterion-referenced evaluation of learning.

In Mar del Plata, three kinds of state secondary education coexist: municipal and provincial schools and a secondary school affiliated to the National University of Mar del Plata. Taking into account the available resources of the research group and in order to limit the scope of our work, we decided to study the whole universe of 6th year secondary school students of the fifteen institutions of the municipal system of education in Mar del Plata. The main characteristic of municipal schools is that they are located in disadvantaged neighbourhoods. Because these students do not usually take private English lessons in addition to the formal school instruction, the population we tested was quite homogeneous regarding their proficiency level.

The development of the project consisted of the following stages: a) revision of the conceptual framework, b) design, validation, implementation and correction of the tests, and c) data analysis and conclusions.

In this contribution we are going to describe the process of the design of the instrument for assessing reading comprehension in English as a foreign language in students finishing secondary school.

Theoretical Framework

The Reading Model

Current reading models claim that comprehension of written texts is a process that implies examining the value of what is read, revising the text, reformulating and transforming it into a new text. We adhere to Rosenblatt's transactional theory (1988) that claims that “Every reading act is an event, a transaction involving a particular reader and a particular configuration of marks on a page, and occurring at a particular time in a particular context” (p. 4) and adds that every reading act is also influenced by the stance the reader takes during the transaction. Reading is considered as an activity in which various cognitive processes are put into play at the same time. In describing the process Rosenblatt (1988) claims that reading is a “complex, non-linear, self-correcting transaction between reader and text” (p. 4) that may end in satisfaction or frustration. We also agree with Goodman´s transactional sociopsycholinguistic view of reading that builds on Rosenblatt’s and adds that “the text [...] has no meaning in itself” and that the meaning “is represented by a writer in a text and constructed from a text by a reader” (Goodman, 1994, p. 1103). This view considers reading as a unitary process that includes the characteristics of the writer, the text and the reader's purpose and experience which ultimately influence the resultant meaning. In the same line of thought, Smith (1985) adds the concept of “The theory of the world in our heads” that guides predictions and inferences in the process of comprehension. Describing how this works from his position as a reader, he states that to interpret new data, “I can only make sense of the world in terms of what I know already. Anything I could not relate to what I know - to the theory of the world in my head - would not make sense to me”. (p. 74).

We also consider the text processing model proposed by Kintch and van Dijk (1978) that explains how readers construct their own situational model of reader interpretation, that is, the individual and unique comprehension of a piece of discourse. Grabe (2010) simplifies the model for the purpose of describing it. For him, fluent reading requires two types of cognitive processing: lower and higher processing mechanisms. Within lower processes he identifies word access, syntactic parsing and semantic proposition formation, i.e., the steps that need to be taken to arrive at the literal meaning at sentence level. When activating higher level processing, readers join the propositions to get to the text base, mainly the literal meaning of what is written and, from this basic comprehension, readers construct their own situational model of reader interpretation that is influenced by their purposes for reading, their background knowledge, experiences and attitudes. At the same time, readers have to make inferences and decide whether they are achieving their reading aim.

Understanding a written text, then, is a recursive process that involves both intertextuality (the relationship of the text with other oral or written texts), and intratextuality (the different ways in which the text is constructed including topic, vocabulary and grammar) (Williams, Regueira, & Caielli, 2016). To make sense of the written piece of discourse, during the reading act, readers make use of all the resources available to them like previous experiences with similar texts, knowledge of the topic, intentions and purposes, and everything that the text itself brings. Definitely, it is not easy for researchers to identify and isolate what is taking place in each step of the process. As regards reading in a foreign language, Velia (2011) claims that it is a conscious and self-regulated process in which the reader moves along two processing directions; from the parts to the whole and from the whole to the parts, without a fixed or preset order, characterized by a constant interchange of directions and by the execution of operations that vary according to text genre, reading purpose, available cultural clues and cooperation with other readers. Undoubtedly, there is an overlap of the different processes in a single reading situation and researchers and teachers find it difficult to describe and assess them. Considering these characteristics, for the purpose of this study, we adopted the definition of reading comprehension proposed by the DiNIECE for their tests:

Reading comprehension is the cognitive capacity to reconstruct the meaning of a written text through the process of reading. This process is recursive and involves tasks such as extraction, interpretation, reflection and evaluation of both local and global elements of the text. (DiNIECE, 2013, our translation).

This definition of reading makes reference to three mental processes that the reader puts into play when trying to understand a written piece of discourse: accessing and retrieving information, integrating information from different parts of the text and interpreting, and reflecting on the text and evaluating its characteristics and effectiveness. These three processes were also those assessed by PISA, ONE and APRENDER (Secretaría de Evaluación Educativa, 2016), the same year we implemented our instrument. We offer a brief explanation of what each of these processes entails.

- ***Accessing and retrieving*** involves revising, locating, selecting and, finally, contrasting the information in a relatively immediate or automatic way to solve the proposed task.

- ***Integrating and interpreting*** means reconstructing local and global meaning, identifying, comparing, contrasting, and integrating information to make sense of the text or to fill gaps of meaning. Integrating implies acknowledging and understanding text coherence, relating pieces of information from different sections of the text to give it sense. Interpreting involves extracting meaning of something that is not explicitly stated.

- ***Reflecting on the text and evaluating its characteristics and effectiveness*** implies relating the text to the reader's own conceptual and experiential frameworks, background knowledge and ideas; it entails the use of extra-textual knowledge to make judgements on the content, the form or text format, based on standards outside of the text. Throughout the reading process readers examine their own experiences to make comparisons, build hypotheses or make judgements about the content or the format of the text.

The Evaluation Model

According to Ravela (2006) "Evaluating basically consists in assessing a reality (an institution, the competencies of a group of individuals, an education system) by comparing systematic empirical evidence with a referent or conceptual definition about what is desirable for that particular reality" (p. 43, our translation). Following this definition, we decided to assess the EFL reading comprehension performance of 6th year secondary school students using as evidence the results of an ad hoc test based on the definition of reading comprehension proposed by the Ministry of Education (DiNIECE, 2013).

Learning achievements can be assessed in different ways depending on the context and purpose of the evaluation. In the case of this research project in particular, considering that the aim was to evaluate what students could do in relation to a specific domain, the reading skill, a criterion-referenced model was selected. Accordingly, we were not interested in comparing students among themselves as in norm-referenced evaluation, or in following their development along some time as is the case of progress evaluation (Ravela, 2006).

Our decision as regards the type of assessment was based on the fact that it was the most appropriate one for the data we wanted to collect. Most validated tests such as the national exams that measured L1 reading comprehension (ONE and APRENDER) follow a criterion-referenced type of evaluation. Moreover, the same model is adopted by international exams such as PISA, implemented by the Organization for Economic Co-operation and Development (OECD), and the Progress in International Reading Literacy Study (PIRLS) (Mullis et al., 2009), set by the International Association for the Evaluation of Educational Achievement (IEA). Another reason for the selection of this model was the manifest interest of the members of the research group to explore the ways in which standardized tests are designed.

Test Design

Indicators

Once the conceptual definition of the domain to be assessed has been defined for assessment purposes, the first step in test design is the operationalization of the construct in terms of indicators**.** According to Ferrer (2007), an indicator is described as a qualitative feature of the competence to be evaluated that provides evidence of what a learner can do. Indicators serve a two-fold purpose; on the one hand they guide the test designer in the formulation of tasks or questions, and on the other, they allow the evaluators to check if the subject can or cannot achieve what is described by the indicator. In order to make this clear, we present examples of indicators of the three processes.

An indicator that gives evidence of the learners´ability to *access and retrieve information* could be that the subjects are able to identify information that is salient in the text; or that they find information that is embedded in the text; or if they can find explicit information that appears in more than one part of the text. In reference to indicators that demonstrate that the subjects can *integrate and interpret* a piece of discourse we can mention their ability to understand the meaning of a word from context; to identify the main idea of a paragraph or to deduce an idea on the basis of the integration of explicit information, for instance. Some possible indicators of what the subjects can do with regard to *reflection and evaluation* are the ability to differentiate between fact and opinion; to identify the type of text; to identify the conflict of a story; or to understand the purpose of text features.

The texts

Both the foreign languages curriculum of the province of Buenos Aires (Subsecretaría de Educación, 2011) and the national guidelines for foreign language education (Consejo Federal de Educación, 2012) were the sources consulted for the formulation of indicators, the selection of text types used in the instrument and the design of the tasks the students had to solve. The selected texts were also endorsed by a consultation made to in-service teachers.

The corpus used represented three different genres: argumentative (article), informative (brochure) and narrative (fable) and each of them was associated with a different context of use. The argumentative semi-authentic text, *Are zoos a good thing*, could be typically found in an educational context as it had an instructional purpose, English language teaching. The authentic brochure *Visiting the Fashion Museum, Bath* represented a context of public use as a reader would resort to it to get to know about political, social, cultural and economic activities in a society. Last, the authentic narrative text, the fable *The Dog and his Shadow* could be identified as a text of private or recreational use, with the aim of recreating fictional worlds.

It could be speculated that each of these texts mainly required the implementation of a specific reading process if we consider its particular characteristics and functions. The argumentative text discussed advantages and disadvantages in relation to zoos. In this case, we could explore, among other aspects, the students' capacity to identify the structure and organization of the text, some of its characteristics, or to give an opinion on its content, that is, skills that relate to the process of reflection and evaluation of what is read. The informative text was a brochure that had the function of providing information about how to get to a tourist place, what the place offered, when you could visit the building and how you could get a ticket to visit it. When reading this type of text, readers generally locate and extract information with the purpose of finding specific data, such as how to get to the place or how to buy a ticket and there is usually no need to read the whole text to achieve these aims. The third text, the fable, presented more possibilities for assessing the extent to which the students could make correct inferences about the author's intentions, or if they could integrate information to explain a fact or the behavior of the character. In our instrument, the three processes were evaluated in each of the three texts.

The tasks

Formulating the questions for the instrument involved detailed reading of each text. Each piece of discourse was approached individually, and was thoroughly read trying to identify the main characteristics and the most salient aspects in terms of reading processes in general, and indicators of performance, in particular. In order to offer reliable samples of what the students were able to do, different question types were selected (multiple choice, open questions and underlining) since only one type might have given limited evidence of the reading processes. That is, the combination of methods made the available information more reliable and objective.

In order to formulate the items, the following criteria were considered:

- The questions were formulated in English and each one was related to a reading process and indicator in particular. Various levels of complexity were considered so as to assign a different grade of performance to each item.

- The closed multiple choice questions offered four alternatives with only one correct answer. The four options had a similar number of words, and each alternative was carefully formulated so that no alternative would look more salient than the others. The incorrect options were stated in such a way that the statement sounded plausible.

- Open or constructed questions, which required the students to produce an answer, sometimes focused on discourse analysis, or developing, explaining or supporting ideas. They could be answered in Spanish to make the students feel better predisposed to solve the test and to help the evaluators obtain more reliable data.

- The first questions were relatively easy to answer to encourage the students to feel comfortable, and they were sequenced according to the order in which the information could be found in the texts. The questions that evaluated higher order cognitive processing, such as the capacity to make inferences or evaluate the texts, were interspersed along the test so that the students who did not have enough time to solve all the items could still demonstrate this type of capacity.

The following chart illustrates the reading processes assessed, the indicators that made them observable and measurable and the corresponding test items.

Table 1

Reading processes, indicators and sample test questions

Reading process	Indicator	Sample test question
Access and retrieve	Identify information that is salient in the text.	What is the address of the Fashion Museum? _______________________
	Find information that is embedded in the	According to the text, how can you find out how

	text.	much a ticket costs? ___________________
	Locating information in the text and relating it to an illustration.	Underline the sentence in the text that describes the picture.
Integrate and interpret	Understand the meaning of a word from context.	What does the word "these" refer to in "A good zoo will enable these species to live and breed in a secure environment."? (paragraph 4) ___________________
	Identify the main idea of a paragraph.	Underline the paragraphs that describe disadvantages of zoos.
	Deduce an idea on the basis of the integration of explicit information.	In the end, the dog ate: a. one piece of meat b. two pieces of meat c. no pieces of meat d. half a piece of meat
Reflect on and evaluate	Differentiate between fact and opinion.	The text presents: a. contradictory facts b. contrasting opinions c. problems and solutions d. conflicting descriptions
	Identify the type of text.	The text is: a. A page of an atlas b. A page of a Geography

		book c. A page of a brochure d. A page of an encyclopedia
	Understand the purpose of text features.	31) Why is the map included? ________________________
	Identify the purpose of the text.	The purpose of the text is: a. to criticize the museum b. to provide information to visitors c. to narrate the history of the museum d. to show how to get to Bath
	Identify the conflict of a story.	The conflict of the story is introduced in: a. the 1st paragraph b. the 2nd paragraph c. the 3rd paragraph d. the title

Piloting

In the first stage of the project we designed and administered four pilot tests to a sample of approximately one hundred 6th year students in private and state schools a year before the actual test, i.e., in October 2015. This allowed us to study the application of the instrument with a population of similar characteristics in terms of age and length of English instruction, from state and private schools. Each student solved a test that included the three selected texts which were accompanied by approximately six questions each. In this way we could pilot an important number of questions to cover all the cognitive processes and indicators considered.

This direct contact with the classroom context through the pilot tests gave us the chance to evaluate different technical and methodological aspects to improve the test validity and reliability. On the basis of the results we could infer that the formulation of some of the questions needed improvement to avoid ambiguities and ensure comprehension. Some items had to be eliminated as they would not provide relevant data or were too easy or too difficult for all the students. In this step of the design process, a new analysis of the pertinence of the indicators was carried out and the reading comprehension construct was permanently debated within the research group, thus enhancing instrument validity.

With respect to practicality, i.e. the aspects related to the implementation of the instrument, the trial confirmed that the time allotted to the test was enough to solve it and that the instructions were clear and precise. We observed the need to modify the order of some questions and to reorganize the way in which they were distributed on the page to avoid possible confusion. We also decided to prepare three different versions of the instrument by changing the order of the texts so as to make sure each student worked individually without glimpsing at other pupils' answers.

The implementation and piloting of the test were carried out by the research group, and we were pleased to observe a high level of cooperation on the part of most of the students; however, a few showed lack of interest and did not make any effort to try to solve the tests. In informal conversations with the students they said that they found the texts similar to the ones they were used to reading in class in terms of genres and topics. As regards the questions, some students considered them relatively easy to answer while others found them difficult to understand.

Levels of performance

Three benchmarks were defined according to the number of correct answers:

- Level 3: a minimum of 60% of correct answers

- Level 2: between 40 and 59% of correct answers

- Level 1: less than 40% of correct answers

Each level was inclusive since it was built on the basis of the previous one. Thus, performance in level 3 assumes the mastery of the capacities and competencies expected for level 2; and level 2 presupposes the mastery of level 1 competencies. Grouping the evaluated population according to different levels of performance allowed us to understand the results better and to report on them more clearly.

In this way, we were able to identify the degree with which the students achieved the academic expectations set in the national foreign languages guidelines (NAP) (Consejo Federal de Educación, 2012) and in the reading comprehension targets stated in the EFL curriculum of the province of Buenos Aires (Subsecretaría de Educación, 2011).

General Results

The census covered the complete enrollment of 284 students in 6th year. From the 226 students who sat for the exam (79,60%), 14 students did not complete the test. Those who solved the test were assigned to the three groups described above; thus, 20 students (8,84%) exceeded mínimum proficiency, 30 students (13,27%) met mínimum proficiency and 162 students (71,68%) partially met the mínimum proficiency expected. The results per school and per course were presented to the local authorities, heads and teachers and the results according to different school contexts were reported in Caielli et al. (2018).

When studying the results according to the text types, we found that the candidates attained the highest number of correct answers (37,74%) in the narrative/fable, followed by the expository text/brochure and the argumentative text with the same amount of correct answers (24,24%).

As regards the reading processes, the instrument contained nine questions that tested reflecting and evaluating. This reading process showed the highest percentage of correct answers, 38,63%. Integrating and interpreting was assessed by means of twelve questions that attained 23,57% of correct responses. Finally, there were ten questions that involved accessing and retrieving information, which displayed 15,09% of correct answers.

The test was accompanied by a questionnaire devised with the aim of obtaining qualitative information that could be of interest for the analysis of the quantitative results obtained by the instrument. The questions were framed in Spanish and sought information about previous EFL learning experiences, considering both private tutoring and school instruction, text types read in the English lessons and purposes for learning English, among others. While going over the data collected there were some issues that called our attention, such as the apparent students' familiarity with expository texts that did not correlate with the test results.

Conclusion

In this work we have presented the steps taken in the design and validation of a criterion-referenced instrument to assess English reading comprehension abilities of students in their last year of secondary school in the municipal system of education of Mar del Plata. The process of test design was highly beneficial for the research group to deepen our understanding of what comprehension of written texts involved. It allowed us to reflect upon how such a cognitive demanding task is taken for granted among literate people, while it is, in fact, a creative, multifaceted process dependent upon many variables of the individual reader as well as of the text itself. From the methodological perspective the study was also useful to strengthen the capacities of the group in the field of criterion-referenced evaluation of learning and to generate reflection and questions on evaluation and assessment of learning in general.

The results provided information to the field of foreign language assessment, especially in reference to reading comprehension. Furthermore, the data gathered were useful to raise questions about the extent to which school ELT practices, particularly reading development and selected discourse types to work in class, were related to the approach and aims prescribed by the EFL curriculum of the province of Buenos Aires and the national foreign language teaching and learning guidelines (NAP). At first sight, taking into account the quantitative results of this evaluation experience, it can be assumed that the official documents have been scarcely considered when planning the teaching of reading since the students' performance in the test showed that they did not master the contents specified in the documents.

On second thoughts, from a qualitative perspective several issues deserve consideration when analyzing the results of a standard test such as this one. The first one was related to reliability: Would the results have been the same if the students had been tested in a different way, with a different instrument? Then, other questions arose: How had reading been treated in the English classroom? What reading processes had been favoured along the school trajectories? What kind of activities had been implemented for reading development? Though the analysis of these aspects are beyond the scope of our work, we can assert that, granting that standard census assessment is useful since it can provide a general view of reality at a given moment under some particular circumstances, a close inspection of the results should not be disregarded. We are very often surreptitiously forced to accept the results of these tests for the simple reason that they were implemented by official institutions such as international organizations and national ministries of education and that they portray a reality assumed to be unquestionable.

An interesting point to consider in the students' responses to the questionnaire was their interest in the inclusion of foreign languages in the curriculum. In spite of the fact that in the questionnaire about school trajectories most of the students claimed to have learned "very little" English, over 70% stated that they would have liked to study another foreign language apart from English. This was a really comforting figure that justifies further research on foreign language learning and teaching in our local contexts.

As teachers and researchers we are not only interested in the general result of students' abilities, but on how this information could be transformed into input for the improvement of teaching and learning. Thus, the project included a report for the Secretaría de Educación, Municipalidad de General Pueyrredon that accounted for the objectives of the project, the theoretical model, the process of test design and the results of the evaluation. Furthermore, a training course for teachers was carried out. The challenge was to provide useful information and guidance so that the focus of classroom practices in reference to reading development could be reconsidered. At the same time, theoretical principles on criterion-referenced evaluation were introduced with the intended aim that teachers would be able to make decisions on how to assess the skill and to design instruments for this purpose.

We understand that we need to become aware of the reality of our classrooms in order to overcome what is imposed on us through imported labels and methodologies in teaching and testing. From our role as teacher educators and researchers, the critical analysis of the process of standardized testing empowered our position regarding the need to strengthen local practices and the development of teacher agency.

References

Caielli, E., Williams, J., & Regueira, A. (2018). Experiencias lectoras en inglés como lengua extranjera en escuelas en situación de desfavorabilidad y su incidencia en una prueba de lectura. *Tejuelo. Didáctica de la Lengua y la Literatura. Educación*, 28, 67-90 https://mascvuex.unex.es/revistas/index.php/tejuelo/issue/view/251

Consejo Federal de Educación. (2012). *Núcleos de Aprendizaje Prioritarios. Educación Primaria y Secundaria. Lenguas Extranjeras*. Ministerio de

Cultura y Educación. https://www.educ.ar/recursos/132577/nap-lenguas-extranjeras-educacion-primaria-y-secundaria/download/inline

Dirección Nacional de Información y Evaluación de la Calidad Educativa (DiNIECE). (2006). *Evaluación criterial. Recorrido político y técnico-pedagógico en el proceso de elaboración, justificación y validación de los criterios de evaluación.* Ministerio de Educación, Ciencia y Tecnología de la República Argentina. http://www.bnm.me.gov.ar/giga1/documentos/EL001414.pdf

Dirección Nacional de Información y Evaluación de la Calidad Educativa (DiNIECE). (2013) *Criterios de Evaluación ONE 2013 para Lengua. 3ª y 6ª Educación Primaria*. Ministerio de Educación de la Nación. http://repositorio.educacion.gov.ar:8080/dspace/bitstream/handle/123456789/109983/Criterios%20de%20Lengua%203-%c2%a6%20y%206-%c2%a6%20Primaria.pdf?sequence=1

Ferrer, G. (2007). *Estudio comparado internacional sobre procesos de elaboración e implementación de estándares de currículum en América Latina*. Programa de Promoción de la Reforma Educativa en América Latina y el Caribe (PREAL). https://www.thedialogue.org/wp-content/uploads/2016/07/2008-Estudio-comparado-internacional-Ferrer.pdf

Goodman, K. (1994). Reading, Writing and Written Texts: A Transactional Sociopsycholinguistic View. In R. B. Ruddell, M. R. Ruddell & H. Singer (Eds.), *Theoretical Models and Processes of Reading* (pp. 1093-1130). International Reading Association.

Grabe, W. (2010). Reading in a second language. In R. Kaplan (Ed.), *The Oxford Handbook of Applied Linguistics* (pp. 88-99). Oxford University Press.

Kintch, W., & van Dijk, T. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363-394. https://www.academia.edu/35404615/Toward_a_Model_of_Text_Comprehension_and_Production

Mullis, I., Martin, M., Kennedy, A., Trong, K., & Sainsbury, M. (2009). *PIRLS 2011 Assessment Framework*. TIMSS & PIRLS International Study Center .

Programme for International Student Assessment (2009). *PISA 2009 Assessment Framework. Key Competencies in Reading, Mathematics and Science.* Organisation for Economic Co-operation and Development. https://www.oecd.org/pisa/pisaproducts/44455820.pdf

Ravela, P. (2006). *Para comprender las evaluaciones educativas. Fichas didácticas.* Programa de Promoción de la Reforma Educativa en América Latina y el Caribe (PREAL). https://blogfcbc.files.wordpress.com/2012/03/14-rabela-fichasdidacticas.pdf

Rosenblatt, L. M. (1988). Writing and Reading: The Transactional Theory. In *Center for the Study of Reading: A Reading Research and Education Center Report*. Technical Report N° 416. New York University. https://www.ideals.illinois.edu/bitstream/handle/2142/18044/ctrstreadtechrepv01988i00416_opt.pdf

Smith, F. (1985). *Reading without nonsense* (2nd ed.). Teachers College, Columbia University.

Secretaría de Evaluación Educativa. (2016) APRENDER. Ministerio de Educación, Cultura, Ciencia y Tecnología

Subsecretaría de Educación. (2011). *Diseños Curriculares para la Provincia de Buenos Aires. Lengua Extranjera. 6ª año. Educación Secundaria.* Dirección General de Cultura y Educación.

Velia, A., Nunez, M., & Lopez, L. (2011). La comprensión lectora en el contexto académico y su papel en el proceso de formación del profesional. *Cuadernos de Educación y Desarrollo*, *3*(26).

Williams, J., Regueira, A., & Caielli, E. (2016, August). How to measure English reading comprehension in secondary school: a proposal [Paper presentation]. Primeras Jornadas Nacionales del Profesorado de

Inglés. Facultad de Humanidades. Universidad Nacional de Mar del Plata.

Using a Pluriliteracies Task for Qualitative Assessment in Higher Education

Ann Montemayor-Borsinger[1], Ana María Coria[2] and Melina Porto[3]

[1] Universidad Nacional de Río Negro

[2] Universidad Nacional de La Plata

[3] Universidad Nacional de La Plata and CONICET

This article describes a pluriliteracies task used in an L2 classroom in a higher education setting and discusses its potential for qualitative assessment purposes. The theoretical framework is based on Meyer's et al. (2015) pluriliteracies model that considers the integration of content and language as fundamental to learning. The task under analysis was based on *Maus* (Spiegelman, 1996), a graphic novel which facilitated student awareness of the two central semiotic modes of pictures and language (Kress, 2010). Students were asked to design an alternative book cover for *Maus*, and to write about it. While the literature has stressed the affordances and constraints of each mode, here we focus on the interdependence characteristic of these 'modal ensembles' (Kress, 2010, p. 58), considered as relevant task products of a wider learning process that takes into account different realisations of meaning making. Four samples were selected of a corpus consisting of 46 alternative book covers and their texts. The selection was based on qualitative assessment in the Systemic Functional Linguistics tradition, taking into account the notion of more or less effective choices that realise meaning according to context (Berry, 1995, Christie & Derewianka, 2010).

We first describe the pluriliteracies task and the context in which we used it, and then discuss the assessment process with the notion of modal ensemble. Using the four covers of our choice with their accompanying written text, we illustrate our findings and highlight the potential of this task for qualitative assessment. We discuss implications for language teaching in all levels of education in light of current accountability and performative drives in assessment.

Background to the Study: Description of the Pluriliteracies Task and its Context

This case study (Yin, 2018) was undertaken in a regular language course in higher education in 2019. The course introduced students to literary works with intercultural citizenship and human rights content. As mentioned above, *Maus* by Art Spiegelman (1996) was chosen, and a CLIL approach (Banegas, 2013) was used with a historical focus (Ravelo, 2013). Participants were 46 second year undergraduate students, aged 18-22, with a B2+/C1 level of English, who engaged in work with the graphic novel in their practical classes for a period of three months. Classroom tasks were designed using a multiliteracies and translanguaging perspective (Bradley et al., 2015; Meyer et al., 2015; Wei, 2018), i.e. student responses to literature involved not only the written mode but also visual, digital, and arts-based elements (see Porto et al., 2019). Several classes were given on the implications of using cartoons for enhancing meaning-making while students were reading the story. They were then asked to design an alternative book cover for *Maus*, together with a written text describing their covers and their new choices. This alternative book cover with the accompanying text is our pluriliteracies task and our main data type.

The Relevance of the Notion of Modal Ensemble in Meaning-making and in the Assessment Process

The conceptualization of world language education in this course is plurilithic, which means it is not based on a monolithic view of language or "a single system of linguistic norms" (Hall, 2013, p. 214). Plurilithic perspectives facilitate and welcome student choice and flexibility in the outcomes of language learning. There is a shift from the traditional notions of competence, grammar and proficiency towards the concepts of repertoire and semiotic resources. Students are encouraged to make their meanings using all available resources (multilingual, artistic, semiotic, performative, digital) creatively (Bradley et al., 2018; Cope & Kalantzis 2015; The New London Group, 1996; Kress, 2010). The focus is not on language systems *per se* but on languages as emergent from contexts of interaction.

In particular, the graphic novel *Maus* facilitated student awareness of the two central semiotic modes of pictures and language (Kress, 2010). While the literature has stressed the affordances and constraints of each mode, the notion of modal ensemble (Kress, 2010) highlights their interdependence as different but complementary realisations of meaning making. Signs (in different modes) are organised as sign complexes that form a coherent and integrated communicational unit. Each sign in each mode makes its own contribution - distinct, potent, and strong – and the resultant meaning of the whole ensemble is more than the sum of its parts. For this reason, the combination of modes in a modal ensemble enhances students' meaning-making potential and becomes a suitable environment for learning (Kress, 2015).

At the same time, modal ensembles provide opportunities for qualitative assessment because they offer evidence or signs of the learning process. Kress (2015) states the rationale as follows:

> MM [multimodality] as the resource from which to select apt signifiers, requires that significance is accorded to the selections — as choices made — of the semiotic means, both the material and the conceptual means. This constitutes the semiotic work of design, by sign-makers. The assumption that the sign — as the motivated relation of form and meaning — is made on the basis of the sign-maker's interest, legitimates a 'reading back' from signs made in all and any modes to uncover the meanings and values attributed by the sign-maker to the cultural/semiotic resources used, and of course a reading back to her or his interest. (pp. 67-68)

The value of modal ensembles for qualitative assessment is that they disclose students' agency through the particular selection of choices they make. These choices also reveal their backgrounds, interests and personal meanings, and thus complement traditional assessment tools. There is "a shift from competent performance to apt design" (Kress, 2015, p. 49).

Our findings

Our analysis draws on Systemic Functional Linguistics and multimodality. Following ethical guidelines in qualitative research, student names are replaced by numbers to guarantee anonymity. Of the 46 alternative book covers produced by students, four samples were selected; two chosen as the most effective and two, as the least effective. We begin by analysing the original cover as an ensemble of signs to move on to the analysis of the signs on the four covers selected to finally show the complementarity between each cover and its corresponding statement.

The Original Cover as an Ensemble of Signs

The original cover blends language and images into a coherent semantic unit. The salient signs are multifunctional. The German word "*Maus*" for "mouse" that refers to the comic's main character is in red and on top of the cover –the letters dripping "blood" and the last "S" in Nazi lightning bolt typeface. – This combination of signs means much more than the actual word. The same multifunctional choices can be seen in the picture of Cat-as-Hitler: the predator is the cat and, at the same time, the predator is Hitler as cat –the typical moustache, the hair parted to one side. The Cat-as-Hitler picture has a swastika in the background and underneath are the victims, with clearly contrasting facial expressions and postures. On this note, the pictures on the cover work on two levels –as does the whole novel. It is the power of a scary looking cat above two helpless-looking mice and a menacing Hitler above two helpless Jews, "the little one wearing a greyish blue blanket, which Jews used to wear to protect themselves from the cold" (Ravelo de Goldzamd, 2010, p. 70).

The Four Redesigned Covers

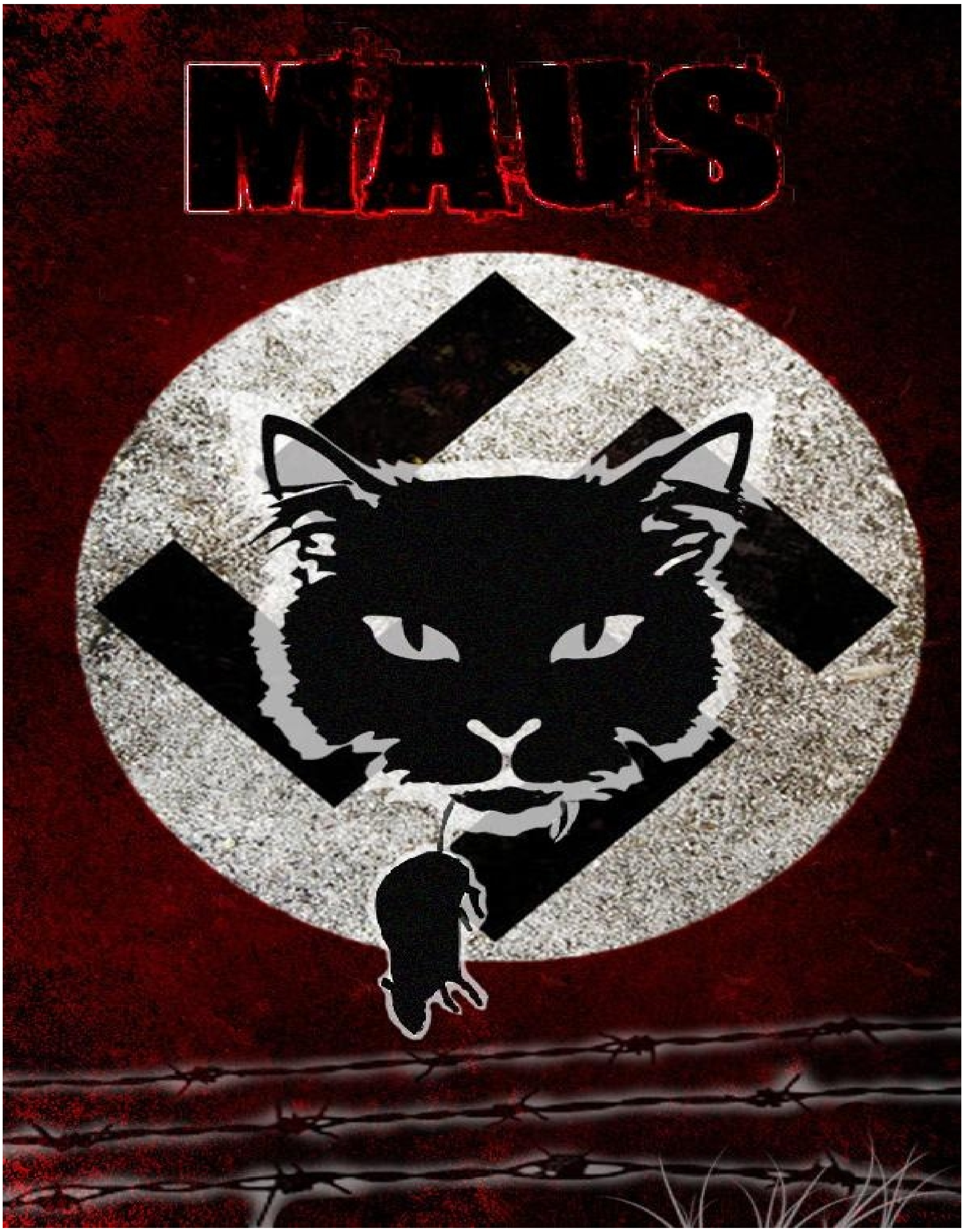
We include the covers here, first the two chosen as the most effective.

Figure 1

Student 1 (most effective) Cover

Figure 2

Student 2 (most effective) Cover



Then, the two chosen as the least effective.

Figure 3

Student 3 (least effective) Cover

Figure 4

Student 4 (least effective) Cover

Interestingly, none of the students saw the symbolism of the typeface of the letter "S" in *Maus* or the Cat-as-Hitler picture, both signs so strongly anchored in the cultural, historical and political context of the time. A possible hypothesis is that students were too young to give meaning to these signs that were so clearly anchored in the decades around the middle of the twentieth century. As Kress (2010) points out, the "makers of signs live in a world shaped by their histories (...) Inevitably, what has been and what is around and available (...) does shape the interest and the attention of the maker of the sign" (p. 74).

The colour red is on all four covers. St 1 uses red in a very creative way, both for the title and in three images: for the swastika-shaped red path along which the author's parents walk; for the wavy lines on the floor where is the image of the rat; for the cynical, impersonal motto "*Arbeit Macht Frei*" (Work Sets you Free) welcoming prisoners at the gates of concentration camps. In the case of St 2, this colour is subtly used as part of a red and black mottled background, and is chosen to outline the letters "MAUS", giving the impression that the word is being licked by flames. In contrast, in the less effective covers it is used in highly conventional ways, basically for the word "*Maus*" and to represent blood.

The two most effective tasks took up the swastika in unconventional ways, but none took up the metaphorical combination of the swastika with the Cat-as-Hitler picture so firmly anchored in the cultural context of the times. One of the effective covers (St 1) makes two sophisticated choices, using the shape of the swastika to represent a red path along which a frightened couple has to walk, and using it on one of the vehicles of a "Holocaust train". The other effective cover uses the swastika as background to the face of a cat. As for the two less effective covers, the centrality given differs. In the case of St 3, the swastika is absent. In the case of St 4, the swastika is just a copy-paste of a conventional image that appears on Wikipedia: the Eagle perched atop a big swastika.

The use of anthropomorphic creatures with an animal head and a human body is a recurrent characteristic of the comic and of course of the cover. This was taken up again only in one of the effective covers.

Interestingly, the students added signs that were not in the original cover.

The Holocaust train is a choice made by St 1, who, as well as St 2, portrays symbols that intend to bring the cruel reality of the regime to the viewer. The Holocaust train with the swastika symbol on a car, plus guards holding truncheons, a gate from a camp and the infamous phrase "*Arbeit Macht Frei*" constitute an ensemble which is a very effective composition of the context of the story.

The effective St 2 cover adds barbed wire at the bottom of the cover. Considering Kress's (2010/2015) idea of the sign as the basis of the sign-maker's interest, we can assume that for this student the electrified barbed wire all around concentration camps was a distinct symbol of the genocide. On this cover, the sign is made more dramatic by the fact that a mouse, hanging from the fierce looking cat's mouth, is just about to touch the top wire.

The two less effective covers only add the very traditional Star of David. St 3 chooses to place it on the mouse's chest in its conventional yellow colour. As for St 4, as has been pointed out previously, the Stars are part of a copy-paste design typical of digital pages about Nazi Germany.

A strong flavor of complex intertextuality is present only in St'1's ensemble. The design blends three photo-shoot images –more than any of the other covers analysed- and two languages: German –not just for the title but for the phrase "*Arbeit macht Frei*" - and English – which she chooses for the four lines of Shakespeare's "The Merchant of Venice". This cover is a very effective way of showing language and image interacting and it shows interpretation beyond the boundaries of the graphic novel under analysis to a more general context.

On the covers where cat and mouse are present –one effective (St 2) and the other less effective (St 3) - size is seen as a representation of power. In both of them, the cat is much bigger than the mouse. However, while the most effective (St 2) picks up on the symbolic issue of cat against the background of a swastika, with barbed wire underneath, in the less effective the only symbol of the historic context is the Star of David on the chest of the mouse.

As to facial expressions in the cats' eyes, the fixed gaze in the cat on the effective St 2 cover conveys hostility and dominance. On the other hand, the less effective St 3 cover pictures a cat peering over a kind of wall. This peering cat is not successful at provoking fear. Moreover, the mouse –though much smaller- has a careless –even "haughty posture"- more appropriate for somebody that is in control of a situation than for somebody who is a victim.

The Redesigned Covers as Ensembles of Signs

If we consider the signs chosen and the spatial orientation in all the covers, the two most effective, Students 1 and 2, seem to be so for different reasons.

St 1's cover appears as the most intertextual. The red title is laid out against a grey background. Below the title, three images in black and white -but which also include the red of blood - remind us of the massacre. Of these three, two are right below the title: to the left, a couple walking along a swastika-shaped, red path. To the right, the picture of a rat (not a mouse); the floor is white but there are also some wavy strokes in red. Below these two images, a third one where the cover portrays anthropomorphic cats and mice - and even a pig. For the bottom part, St 1 chooses to include four lines from "The Merchant of Venice", a play that stereotypes the Jews negatively. St 1's representation is a very effective sample of hypertextuality as it connects the beholder with another text. It also triggers interest by including the legend "*Arbeit Macht Frei*" and finally, it can be said to be a multilingual cover as it includes two distinct quotes in their original languages.

St 2's mottled cover with dramatic flames licking the word "MAUS" picks up the swastika against the white circle and the fierce-looking cat against the swastika. The predatory relationship between cat and mouse is present in her choice of the mouse hanging from the cat's mouth. Her cover includes barbed wire, a symbol of oppression, unfreedom, dehumanisation and crimes against humanity. At the same time, the barbed wire is a reminder of Holocaust survivors. In some way, it might mean "a survivor's tale".

As for the spatial orientation and the choice of signs in the least effective covers, they do not seem to have grasped the core message for different reasons. On St 3's cover, there is red in the title and the effect of blood dripping. The cat is on the upper spatial section and the mouse on the lower. However, on the whole, this cover does not convey a message of fear on the one hand, and helplessness on the other. Though staring at the beholder, the cat is peering behind a wall, as if not in total control of its prey.

This was certainly not the relationship between cats (victimizers) and mice (victims) in the historical context the novel portrays. More strikingly, the mouse on this cover is not pictured as the helpless being the context of the novel portrays.

St 4's cover is just a combination of conventional symbols: the two Stars of David, each at one side of the title, the Nazi eagle combined with the swastika, and an image of blood dripping, all of them highly conventional and readily available on the internet.

What the Relation between Modes Tells us

In this section, we show whether the most effective visual cover choices were accompanied by well-developed written observations. Our students were asked to use two different modes to express their meanings as a way of responding to their interpretation of "Maus". Their redesign of the original cover plus their written reflections on the creative process and choices thus provided us with more tools to assess their learning.

The Two Most Effective Covers

Here we include the written statements corresponding to the most effective covers.

> I decided to include in the cover three pictures that, from my point of view, best summarize what the story is about. In the first one, Art Spiegelman's parents can be seen walking through a path which resembles the Nazi Swastika, as a metaphor of what their lives were like when they were being chased by the Nazis. As a way of reflecting all the consequences that this persecution brought into their lives, I painted the path red. The second one is a rat, which is the animal used by the author to represent Jews, and which perfectly portrays the way in which Nazis (cats in the story) treated them. And the last one, is a Nazi concentration camp in which the quote "Arbeit Mach Frei" can be seen. This

quote means “Work sets you free”, which seems to be a very cruel irony, as most of the Jews who entered those camps to “work”, were never be free again. I thinks this quote is also very representative because it’s written in German, as well as the title of the novel, and it’s a way of showing that in this story Germans will play a key role.

At the bottom, I included a fragment from the well know Shakespeare’s play “The merchant of Venice”. This is a play in which Jews are characterized very negatively and full of stereotypes.

The quote is said by a jew character, and it emphasizes the idea that Jews are treated differently just because of their religious beliefs, and therefore the character remarks the idea that Jews are just like the rest of the people. Furthermore, Hitler was said to be a fan of this play, as he supported its anti- Jewish themes.

I tried to maintain similar colors to the original cover, but made special emphasis on color red, as “Maus” is a very sad and bloody story.

Written statement. Student 1. Most effective cover.

In the process of making this alternative cover of the book "Maus" by Art Spiegelman, I had in mind that I wanted to transmit the context in which the story is situated, the trauma the Holocaust triggered and what I felt when I first read it. Since it is a graphic novel, drawings have the biggest role in representing what the author tries to convey. Although I personally truly liked the original cover, I wanted to portray a stronger and more impactful appearance.

In my creation, I chose to implement colours such as black, white and red for the font and the different images, since I felt they would help illustrate the events and their significance more

accurately as the story unfolds. These colours joint together along with each item gave me the idea of blood, death and suffering.
Also, symbols abound throughout the graphic novel in its sequential frames, which I think are an aspect of great importance. The Swastika behind the cat occupies a big portion of the picture considering that I wanted to display the enormous power Nazis had compared to Jews along with the main point of the cover: the predatory relationship between cats and mice. The first are the cunning, aggressive hunters who visualize the tiny mice as an inferior race. On the other hand, the latter are often seen as filthy creatures that anyone would try to "get rid of"; they remain hidden in holes or any other dirty place out of their predators' sight. Many times, cats are used as a mice deterrent and this case is not an exception.
Jews drawn as sad-eyed, terrified mice and Nazis as menacing, hostile cats is a simple but effective metaphor Spiegelman applied to represent the deshumanizing cruelty this mass slaughter brought up between 1941 and 1945.
Finally, I decided to add rows of barbed wire fence to represent those that were used to separate the camps into different sections, where people would be tortured, abused and later systematically murdered in the most horrendous ways.

Written statement. Student 2. Most effective cover.

Processes Chosen for Description of Redesign

Both students describe the process of creation by using basically mental processes or complex modulated ones. In the case of St 1: “I decided to include”, “As a way of reflecting”, “I think”, “I tried to maintain”. As for St 2: “I had in mind”, “I wanted to transmit”, “I personally truly liked”, “I wanted to portray”, “I felt”, “I think”, “I wanted to display”, “I decided to add”. Both students select processes that inform the reader on their processes of choice.

The Lexical Chains

It can be said that covers and statements taken together constitute a coherent and cohesive ensemble. St 1's lexical chain includes: Nazi swastika, Nazis, persecution, Jews, Nazis, Nazi concentration camp, Jews, camps, German, Germans, "The Merchant of Venice" (intertextual), Jews, Jew character, Hitler, anti-Jewish themes, red, sad and bloody story. St 2's lexical chain includes: Trauma of Holocaust, black, red, white, blood , death, suffering, swastika, power, Nazis, Jews, predatory relationship, cats and mice, cats, mice deterrent, Jews, Nazis, hostile cats, mass slaughter.

The lexical chains reflect the visual. Whereas St 1's cover chooses signs that portray the historic context and centres on the anthropomorphic figures, thus choosing mainly the words Nazi/Jews, St 2's cover merges cats/ Nazis vs. mice/Jews and the predatory relationship between the powerful cats and the powerless mice. The cruelty encoded in the image chosen for the cat is realised in the meanings: "hostile cats", "mass slaughter" of the statement. All in all, these two students realise their meanings by complementing coherently their linguistic and visual repertoires.

The Two Less Effective Covers

Here we include the written statements corresponding to the least effective covers.

> When you come across the book cover of Maus, you can easily get the idea of what is going to be about. The metaphor is understandable, but I believe there are some elements which are missing in any of the covers designed by Art Spiegelman.
> Firstly, in the original cover, it is shown a circle with the swastika, and on top of it there is the image of a cat looking very similar to Hitler; below that, the mice hugging. I think that this was

presented in two different levels, and what I wanted to show is that those levels are connected throughout the whole story Vladek tells. The chase made by the Nazis to the Jews and their attempt to hide from them, is fairly all what happens from the time Hitler gets to presidency. That is why I chose it as the central image of my new cover. Besides, the cat's size compared to the mouse's shows the unequal power relation between the two, and how Nazis thought the Jews to be so inferior, almost as if they weren't human.
Secondly, I believe the font choice is much more descriptive than the original one. Although I thought important to keep the red colour, I found the new font to be more explicit about the horrors the Nazis have made with Jewish people.

Written statement. Student 3. Least effective cover.

For the cover of the graphic novel "Maus", I made different choices of images and colors. First, the title is in red in representation of blood, for the blood of everyone who died because of the Holocaust. Then, there are two stars of David, which are the symbol that represents the Jews. These two stars are at the beginning and the end of the title, suggesting that the story that is about to be told is a story relating the Jews.
Beneath the title, we can find the image of an eagle holding a swastika, a symbol of the Nazis, dripping blood, in order to show metaphorically the blood spilled by them during World War II, the blood of innocents.

Written statement. Student 4. Least effective cover.

The designers of the two less effective covers also produced less effective statements. The lack of effectiveness of their visual signs was replicated in their linguistic repertoires. One key aspect in St 3's and St 4's covers is the lack of emotion, not portraying effectively enough the horror of the story. We will try to show which language choices contribute to this air of impersonality.

Processes Chosen for the Description of the Redesign

St 3 resorts mainly to passive constructions such as: “it is shown”, “was presented”, “are connected”, and relational and existential processes: “there is”, “is”, “is”, “gets”, “weren’t”, “is”, although the description also includes “I wanted to show”, “I chose”, “I believe”, “I thought”. On the whole, the choice of alternating between more impersonal processes and mental ones distances both the writer/designer and the reader/beholder of the context of a story of this particular multimodal ensemble. As for St 4, with the exception of “(I)made”, all the processes are relational and existential: “is”, “there are”, “are”, “represents”, “are”, “is”, “we can find” (meaning: “there is”). Once again, the statement replicates the air of lack of personal involvement, present in the cover by means of the ensemble chosen.

The lexical chains

St 3’s cover makes several references to the original cover. In this sense, the explanation starts with the noun: “chase”. Then, “Nazis”, “Jews”, “Hitler”, “presidency”, “central image”, “cat’s size”, “mouse’s”, “unequal power relation”, “Nazis”, “Jews”,” “red colour”, “horrors”, “Nazis”, “Jewish people”. However, though the student refers to the difference in size between cat and mouse to represent unequal power, there is no reference to the mouse´s gesture or posture. The surprisingly haughty looking mouse is ignored. In St 4, the lexical chain of the statement includes “red”, “representation of blood”, “blood”, “Holocaust”, “stars of David”, “symbol”, “Jews”, “stars”, “story”, “story relating to Jews”, “image of an eagle”, “Swastika”, “symbol of the Nazis”, “blood”, “blood”, “blood of innocents”. Although image and speech organize representation differently, the lexical chain in St 4’s statement reflects the same kind of conventional meanings of the signs on the cover. Basically, symbols which one can find on a Wikipedia entry.

Conclusion and Implications for Language Teaching

In this article, we have shown how modal ensembles offer opportunities for a qualitative assessment that values student choice and personal meanings. Drawing on multimodality, our analysis has shown that effective tasks make bolder visual choices that blend traditional symbols with less established ones, accompanied by well-developed written observations on the meaning-making potential of their choices. This type of assessment is different from current accountability and performative drives with a static focus on language systems and pre-eminent attention to competence standards. Following Kress (2010, 2015), we have argued that modal ensembles allow students to transcend the boundaries between named languages, language varieties and other semiotic systems. In other words, learners can make meaning by complementing their linguistic repertoires with the multimodalities that form part of their semiotic repertoires. This complementarity is valuable because "the use of modes in combination offers a fuller means for conveying meaning, richer than the comparatively sparse capacities of the linguistic modes of speech and writing" (Kress, 2015, p. 57) usually favoured by schools and universities.

The implications for language teaching are significant. First, the evidence of learning that students are able to provide under particular circumstances is limited not only by the modes they have access to at a given moment (i.e. their current multimodal repertoires) but also by those modes they are allowed to deploy in specific assessment situations. Restricting students to exclusively linguistic means of meaning making runs counter to social justice language education principles based on the provision of relevant and meaningful learning experiences (including assessment) (Okan, 2019). Privileging only the communicative modes favoured in schools and universities (listening, speaking, reading, and writing) limits students' expression to the linguistic dimension and disregards other multimodal means such as the creative and artistic.

Second, this multimodal and translanguaging approach is recognised as an urgent need for assessment in the field (Cenoz & Gorter, 2019). It requires the consideration of multiple forms of evidence of learning such as students' productions and artefacts (written texts and spoken interactions but also redesigned book covers in our case, drawings, responses to triggers, informal anecdotes and so on) and evidence of reflection and language awareness in self-referential assessment. In this way, teachers can focus on their students' progress instead of looking at predetermined outcomes set by a test or a syllabus.

Finally, to be enacted in classrooms, this conceptualisation of assessment needs to be supported and endorsed by teachers, schools, universities, policy documents and curricular developments. Moreover, even when such support exists, teachers need praxis development, gained through teacher education or professional development. To conclude, the following questions arise: What views of language and learning do teacher education programs in specific contexts project? How are the notions of language and learning theorised in these programs? What kind of educational philosophy supports qualitative assessment using modal ensembles? What teacher identities and roles are needed and how can they be developed and supported? What are the risks and challenges involved and how can they be addressed pedagogically?

References

Banegas, D. L. (2013). The integration of content and language as a driving force in the EFL lesson. In E. Ushioda (Ed.), *International perspectives on motivation: Language learning and professional challenges* (pp. 82-97). Palgrave Macmillan.

Berry, M. (1995). Thematic options and success in writing. In M. Ghadessy (Ed.), *Thematic development in English texts* (pp. 55-83). Pinter.

Bradley, J., Moore, E., Simpson, J., & Atkinson, L. (2018). Translanguaging space and creative activity: theorising collaborative arts-based learning. *Language and Intercultural Communication, 18*(1), 54-73.

Cenoz, J., & Gorter, D. (2019). Multilingualism, translanguaging, and minority languages in SLA. *The Modern Language Journal*, *103*, 130-135.

Christie, F., & Derewianka, B. (2010). *School discourse*. Continuum.

Cope, B., & Kalantiz, M. (Eds). (2015). *A pedagogy of multiliteracies: Learning by design*. Palgrave Macmillan.

Hall, C. (2013). Cognitive contributions to plurilithic views of English and other languages. *Applied Linguistics*, *34*(2), 211–231.

Kress, K. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

Kress, G. (2015). Semiotic work. Applied Linguistics and a social semiotic account of Multimodality. *AILA Review, 28*, 49–71.

Meyer, O., Coyle, D., Halbach, A., Schuck, K., & Ting, T. (2015). A pluriliteracies approach to content and language integrated learning – mapping learner progressions in knowledge construction and meaning-making. *Language, Culture and Curriculum*, *28*(1), 41-57.

Okan, Z. (2019). Language and social justice. In R. Papa (Ed.), *Handbook on promoting social justice in education* (pp. 1–14). Springer.

Porto, M., Zamuner, A., & Miguel, V. (2019). Using literature in a language course in Higher Education: An opportunity to embed citizenship goals in English teaching and translation degrees. In D. Banegas, M. Porto, M. López-Barrios, & F. Perduca (Eds.), *Literature in ELT. Selected Papers from the 44th FAAPI Conference* (pp. 1-10). ASPI.

Ravelo, L. C. (2013). The use of comic strips as a means of teaching history in the EFL class: Proposal of activities based on two historical comic strips adhering to the principles of CLIL. *Latin American Journal of Content and Language Integrated Learning*, *6*(1), 1-19.

Ravelo de Goldzamd, L. C. (2010). *Análisis de Maus de Art Spiegelman, historieta de guerra de final abierto.*[Tesis de Maestría de la Universidad de Buenos Aires]. Repositorio Institucional de la Facultad de Filosofía y Letras. U.B.A.

http://repositorio.filo.uba.ar:8080/xmlui/handle/filodigital/1827/browse?value=Ravelo+de+Goldzamd%2C+Livia+Carolina&type=author

Spiegelman, A. ([1986] 1996). *Maus. A Survivor's Tale. My Father Bleeds History*. Pantheon Books.

The New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review, 66,* 60–92.

Wei, L. (2018). Translanguaging as a practical theory of language. *Applied Linguistics, 39*(1), 9–30.

Yin, R.K. (2018). *Case study research. Design and methods* (sixth edition). Sage.

The Assessment of English Pronunciation Using Different Frameworks of Reference

María Garay and Florencia Giménez

Facultad de Lenguas, Universidad Nacional de Córdoba

The teaching and assessment of pronunciation has traditionally had the native speaker (NS) as a model of performance (Pennington & Rogerson-Revell, 2019). However, for several decades now, the nativeness principle has been challenged by many linguists (Davies, 2003; Pennycook, 1995, 2012; Ur, 2012) given the difficulty, if not impossibility, for any non-native speaker of English (NNS) to achieve such an aim. A new goal for the teaching of English as a second, foreign or additional language has thus emerged. Ur (2012) describes this new aim as that of a fully competent speaker who may or may not originally have been a native speaker. This competent speaker is able to communicate proficiently and to show enough flexibility to adapt to different models and styles of communication when interacting with NS and NNS of English. In the field of pronunciation, Pennington and Rogerson-Revell (2019) define pronunciation competence as a set of skills at the receptive and productive levels that include segmental and prosodic phonology specifically relevant in international or lingua franca communication. Above all, being a competent speaker in an international context implies being easily understood by others (Isaacs & Trofimovich, 2012; Munro et al., 2006). Under this light, different models for the teaching of English to be used as an international language and a set of pronunciation features have been described by experts in this area.

This paper presents the results of an exploratory case study in which the assessment of English pronunciation under the light of different frameworks of reference is compared. To achieve that aim, speech samples of students of English of the School of Languages, Universidad Nacional de Córdoba (UNC), sitting for the final exam of the subject Pronunciation Practice were used to compare the marks allotted through impressionistic assessment and using a scale designed to be used with two different frameworks of reference: English as a Lingua Franca and the Lingua Franca

Core (LFC) (Jenkins, 2000) and International English (Cruttenden, 2008). The results are interpreted considering the context of the study and the different frameworks used.

Theoretical Framework

This section presents the theoretical foundations of this study. As already stated, the main purpose of this work is to inform about the way in which different frameworks of reference impact on the assessment of pronunciation in English. To explore this, it is of paramount importance to refer to the Nativeness and Intelligibility Principles and the goals for the teaching, learning and assessment of pronunciation aimed at by pedagogies that align with one or the other. Next, this section will delve into the different frameworks from which different priorities in the field of pronunciation derive. Finally, there is a brief reference to current assessment practices and the principles that guide them. These also form part of the theoretical foundations that motivated this study.

The Nativeness and the Intelligibility Principle

In 2005, Levis introduced the Nativeness Principle (NP) and the Intelligibility Principle (IP) as two ends of a continuum along which different goals for the teaching of pronunciation are positioned. He described these two terms thoroughly and referred to the impact each has in the teaching, learning and assessment of pronunciation. Levis also added that the IP is gradually gaining popularity since its goals are more easily attained by NNS.

The NP is positioned on one end of the continuum. A pedagogy that aligns with the NP puts emphasis on theory-based instruction followed by controlled practice in which imitation of NS models plays a key role. The objective is to attain native-like performance at the phonological level and, to achieve that aim, it is believed that a considerable amount of class time should be devoted to teacher-centred activities that rely on the theoretical expertise and native-like performance of the teacher. As Levis (2005) points out, this principle says that learners have to completely match a native speaker's production of all pronunciation targets, segmental and suprasegmental, in order to be communicatively competent. When pronunciation courses are guided by the NP, there is practically no content selection since there are no criteria that determine the weight different pronunciation features have. Instead, nearly all segmental and suprasegmental contrasts are taught. NS pronunciation is flawless and NNS performance should move towards trying to achieve that goal, despite the fact that “nativeness is usually unattainable” (Levis, 2020, p. 4).

Due to the fact that at the School of Languages students of English are being trained to be teachers, translators or researchers, instruction aims at a high level of performance to help students become communicatively competent users of the target language. Besides, in the three annual courses on phonetics, they receive formal explicit instruction in all the segmental and suprasegmental features of the English language and are expected to identify and produce them. This characterization leads us to conclude that instruction at the School of Languages is guided by the NP.

On the other end of the continuum, the IP can be found. This does not have the native speaker in mind and seems to be more consistent with the diversity of English accents in the world at present. A pedagogy that aligns with this principle sets priorities and makes a careful selection of contents to include in a pronunciation syllabus. Also, it includes theoretical explanations so that learners can complete communicative awareness building tasks in which controlled and free-speech production tasks are combined and carried out either individually or collaboratively. This way of teaching and learning pronunciation integrates analysis and reflection, and fosters the development of self-monitoring strategies. Above all, having intelligibility as a guiding principle favours diversity and aims at helping learners to be easily understood, “no matter how much the speaker’s accent diverges from that of a native speaker” (Pennington & Rogerson-Revell, 2019, p. 32-33). This principle is based on research that has found that strongly accented English can also be intelligible (Derwing & Munro, 2015; Munro & Derwing, 1995).

As Levis (2018, 2020) comments, it is undeniable that L2 teaching and learning based on both principles consider pronunciation a key component of the L2 syllabus. However, they differ in the importance given to different pronunciation features. Consequently, the impact on the evaluation of student success is inevitably different.

Assessment strongly influenced by the NP considers that any feature that deviates from a native-speaker model is a mistake, no matter the impact this deviation has on communication. All the pronunciation features have the same weight and they should all be penalised, no matter whether they produce breakdowns in communication or not.

The IP is, by contrast, context sensitive. Certain features should be learned and practised first because "they are more likely to harm communication than are others" (Levis, 2018, p. 34). If it has been found that "different pronunciation errors have different effects on a speaker's comprehensibility and intelligibility" (Derwing & Munro, 2015, p.76), that should be the key principle that guides assessment in the field of pronunciation.

A description of the NP and the IP is key to understanding the rationale behind the different frameworks of reference for the assessment of pronunciation analysed in this study. Besides, it also gives a clear picture of how contexts have changed and paradigms have shifted in pronunciation teaching and assessment (Levis, 2005).

The Lingua Franca Core

One of the linguists that first called into question the NP was Jenkins (2000). Considering that NNS of English will interact in English more frequently with other NNS of that language than with NS, Jenkins carried out research in order to spot which pronunciation features are crucial for effective communication among speakers that do not share their L1. Empirical evidence reveals that there is a set of pronunciation features that have a greater weight in communication than others and allow NNS of English (NNES) to communicate without major breakdowns in communication in contexts in which English is used as a Lingua Franca (ELF). Those features make up the Lingua Franca Core (LFC) and, according to Jenkins, they should be present in any pronunciation syllabus. Table 1 presents a summary of the LFC.

Table 1

Summary of the LFC (Jenkins, 2000)

SEGMENTALS	SUPRASEGMENTALS
Consonants: Most consonants.	Production and placement of nuclear stress.
Some substitutions of voiceless dental fricative /θ/ and its voiced counterpart /ð/ are acceptable.	Appropriate use of contrastive stress to signal meaning
Rhotic /r/ is preferred.	
Intervocalic /t/ (British English) is preferred.	
Allophonic variations within phonemes are permissible.	
Aspiration following word initial stressed voiceless plosives /p, t, k/ is required. If unaspirated, they may be mistaken for /b, d, g/ respectively.	
Consonant clusters:	
No omission of sounds in initial consonant clusters.	
Omission of middle and final clusters only permissible according to L1 English rules.	

/nt/ between vowels as in British and not American English.

Epenthesis is more accepted than unwanted omissions.

Vowels:

Maintenance of contrast between long and short vowel sounds, e.g. the contrast between ‘rid’ and ‘read’.

Vowel length that depends on the phonetic environment.

According to Jenkins (2000), both segmental and suprasegmental features must be present in a pronunciation syllabus. Within segmentals, ELF speakers should be able to produce most consonant sounds in the English phonological inventory, especially aspirated voiceless plosives, due to their high functional load, i.e. the importance these sounds have in distinguishing meaning in English. Consonant clusters, mainly those in word-initial position should be produced accurately. As regards vowels, phonemic and allophonic differences of length should also be produced. At the suprasegmental level, ELF speakers should be able to produce nuclear and contrastive stress.

The LFC is one of the frameworks that has been used to assess the pronunciation of NNS of English in this study since it is guided by the IP, as opposed to the key orienting principles that seem to guide instruction and assessment at the School of Languages.

International English

Cruttenden (2008) presents different sets of priorities and tolerances based on the purposes for which L2 learners learn this language. Apart from the Native Speaker Targets in which, as we have already said, all the pronunciation features are equally important, Cruttenden also presents

Amalgam English (AE) and International English (IE) and defines them as "categories of pronunciational and productive competence" (p. 318). This author defines Amalgam English as a hybrid between American and British varieties which may also include characteristics that can be transferred from the user's L1. IE, on its part, is English learned to be used as a lingua franca for international communication and "involves simplifying the Amalgam English in ways to make it easier for learners from many different language backgrounds (…) but not to an extent to make these varieties mutually unintelligible" (Cruttenden, 2008, p. 330).

Cruttenden (2008), like Jenkins, also suggests a set of features to guide pronunciation teachers and learners of English for international purposes. Cruttenden's IE was chosen as a basis for comparison in this study as it was considered to yield more significant results than the comparison LFC vs AE. Table 2 presents a brief account of the core pronunciation features for IE.

Table 2

Summary of the core pronunciation features for IE (Cruttenden, 2008)

SEGMENTALS	SUPRASEGMENTALS
Consonants:	Lexical stress in polysyllabic words.
Allow voicing distinctions to be made using different features than those used by native speakers.	No attempt need be made to use the weak forms of English or the weak syllables in polysyllabic words.
All forms of /r/ and /l/ are allowed but distinction between the two to be given high priority.	No effort needs to be made to learn native intonation patterns of English
Distinction between /v/ and /w/ should be insisted on; use of /ʊ/ for either or both discouraged	
Vowels:	

A reduction in the vowel inventory to five short and five long vowels is allowable.	

In comparison to the LFC, a reduction of features that should receive explicit instruction is evident and more tolerances than demands are explicitly stated. That is the case, for example of voicing distinctions at the level of consonants, different forms of /l/ and /r/ and a reduction of the inventory of vowel sounds. IE users should not make the effort to reproduce the rhythm or intonation of English, either. Compared to the frameworks presented so far, this seems to be the most flexible and lenient of the three.

Both the LFC and IE are frameworks that mainly describe the spoken competence of learners. However, they also highlight the importance of listening, understanding and adapting to other interlocutors as part of the processes of accommodation and adaptation mentioned before. For the sake of the present work, only the spoken competence of speakers has been analysed. In spite of their differences in number and type of features, both frameworks aim at intelligibility and present clear aims for the teaching and assessment of pronunciation. These core features will be the models against which the samples selected will be assessed.

Pronunciation Assessment

Current assessment practices in the field of pronunciation include terms such as intelligibility and communicative effectiveness (Pennington and Rogerson-Revell, 2019). These authors recommend that the “what” and “how” of assessment reflect the target or model that pronunciation aims at in the classroom and provide a set of guidelines to help teachers assess the pronunciation of their learners. Such assessment needs to include:

- Different aspects of pronunciation (segmentals, prosodies, fluency, intelligibility, referential and social meaning)
- A range of tasks and contexts (open and closed tasks, read and spontaneous production, familiar and unrehearsed material)

- Variety in audience (monologues and interactions) to allow the student to give evidence of adjusting and accommodating to other interlocutors

Cruttenden adds that “real assessment must be based on the general intelligibility and acceptability of a learner's performance as judged by the type of listener the speaker seeks to communicate with” (2008, p. 351). Even if a teacher or external judge (novice or expert) may assess a sample in a holistic or analytic way, the use of rating scales is a highly recommended practice for raters. In order to design assessment scales to judge a speaker’s performance, Harding (2017) proposes the observance of some technical issues:

- Avoid abstract terminology
- Maintain consistency of terminology across the scale
- Keep scales brief (no more than 6 levels)
- Design descriptors that are no longer than 5 clauses each

Even if the type of tasks used here have not been designed for the sake of this study, many of the guidelines proposed by Pennington and Rogerson-Revell (2019) are reflected in the type of test conducted. Harding’s (2017) recommendations will be taken into account when designing the scales and descriptors to assess the samples selected under the light of ELF and IE.

Research Questions

For this study, the following research questions (RQ) will be addressed:

RQ1. How much variation is there if the samples are assessed following the models suggested by the LFC and IE and criteria adopted in the original test?

RQ2. How different will be the baseline for passing when comparing the results of the three methods of assessment?

RQ3. Is the evaluation of pronunciation at the School of Languages more closely aligned with the NP or the IP?

Accordingly, the aim of this study is to compare the assessment of pronunciation of first-year university students based on the models suggested by the LFC and IE and the criteria adopted in the original test.

The expected outcome is that the ratings yielded by the original test will be lower than those that derive from the assessment based on the models suggested by the LFC and IE. If our hypothesis is confirmed, we expect our findings to align with previous research that indicates that "ideologies of nativeness and near-nativeness are deeply entrenched within L2 pronunciation" (Levis, 2020, p 3).

Methodology

Context of the Study

At the School of Languages, students aiming to become teachers, translators or researchers take three obligatory annual courses in English pronunciation. The students in each course are, in turn, divided into different groups taught by different instructors. However, the contents, objectives and material are common.

In these courses, the evaluation of pronunciation addresses three aspects: theoretical knowledge, perception and production. As regards oral production, there is general consensus as to the segmental and suprasegmental features to consider when evaluating a student's oral performance. It is agreed that instruction and, in turn, assessment are oriented towards communication. On the basis of this, the pronunciation instructors evaluate and grade their students' oral performance through "guided judgement" (Common European Framework, 2002, p. 189), a kind of judgement in which individual assessor subjectivity is reduced by complementing impression with conscious assessment in relation to specific criteria. In other words, no explicit scale is used during assessment and general impressions prevail when assessing our students' oral performance.

It is in this context that instructors in the area struggle to incorporate new methods to measure our students' oral performance as precisely and as objectively as possible. This is what motivated this study.

Samples and Participants

The samples selected for this analysis were taken from the final exams of eight first-year university students of English Pronunciation Practice, the first of the three courses mentioned. Because of the changes brought about by the ongoing lockdown in our country, the July, November and December exam sessions were carried out online and were recorded. This resulted in a wide range of samples already rated by a board of expert teachers out of which the recordings used here were extracted.

The final exam consists of two tasks: the first one is a retelling activity in which the students are assigned a story 30 minutes before the start of their performance. This is a semi-spontaneous task in which the students have to narrate the events presented in that story making choices such as point of view and tense. The second task is a reading aloud activity for which students are assigned a passage from the coursebook. These two tasks reflect the guidelines presented by Pennington and Rogerson-Revell (2019) in that different task types are included and both familiar and unfamiliar material is used.

As said, students are assessed without using rating scales; however, the raters have been members of the chair for a long time, so that they are experienced and share the same criteria, which is fully described in the syllabus of the subject. The main objectives of the course are to help students incorporate the sounds and rhythm of English to sound intelligible.

Students are awarded a numerical mark (1 to 10) to comply with the institution's requirements, with 4 being the lowest passing mark and 10 the highest level of performance. Table 3 presents the samples that were selected for this study. The samples selected belong to students who showed different levels of performance: two samples represent a very high level (marks 9 and 10), two samples reflect a satisfactory performance (mark: 6), two samples show a narrow pass (mark: 4) and two samples correspond to a failing mark (mark: 2). For ease of identification, the samples were labelled as follows:

Table 3

Selected samples

SAMPLE	LEVEL OF PERFORMANCE	MARK RECEIVED IN THE ORIGINAL TEST	CODING
1	High	10	H1
2	High	9	H2
3	Satisfactory	6	S1
4	Satisfactory	6	S2
5	Narrow pass	4	N1
6	Narrow pass	4	N2
7	Fail	2	F1
8	Fail	2	F2

Scale and Descriptors

To be able to compare the different methods of assessment, one single scale was designed and two experienced raters in the field of pronunciation were trained in the use of the scale to assess all the samples in this study, each rater working with one framework: LFC or IE. A descriptor that refers to unintelligible extracts has also been added to this scale. The descriptors designed are short, following Harding's (2017) suggestions. The simplicity of these descriptors makes them easy to use when judging the samples under the light of the two cores selected for this work: LFC and IE. The scale and its descriptors are presented below in Table 4.

Table 4

Scale and descriptors

MARK	DESCRIPTORS
10	The core features are produced **accurately,** despite rare inconsistencies. The speaker does not impose strain on the listener.
9	Production shares features of 10 and 7. More features of 10 than 7.
8	Production shares features of 10 and 7. More features of 7 than 10.
7	**Several** core features are **produced accurately** (about 75%), despite some inconsistencies. Speech produced in the sample can be easily understood.
6	Production shares features of 7 and 4. More features of 7 than 4.
5	Production shares features of 7 and 4. More features of 4 than 7.
4	**Only some** core features are **produced accurately** (about 50%). There are several inconsistencies but the sample remains generally intelligible.
3	Pronunciation is unintelligible at times and recurrent mispronunciations put a strain on the listener.
2	Pronunciation is mostly unintelligible and inappropriate mispronunciations place a heavy strain on the listener.
1	There were no attempts at producing stretches of fluent speech.

It is worth pointing out that, apart from assigning a mark to each sample, the raters were also asked to write comments in which they could specify and exemplify the problems they had detected in each sample.

Results and Discussion

Table 5 presents the ratings of the samples after being assessed with different frameworks of reference.

Table 5

Marks the samples received after being assessed with different frameworks of reference.

SAMPLE	ORIGINAL TEST	LFC	IE
H1	10	9	10
H2	9	9	9
S1	6	7	9
S2	6	6	9
N1	4	5	8
N2	4	5	7
F1	2	3	3
F2	2	3	3

As anticipated, the core features proposed by LFC and IE are more lenient than the criterion used by the raters in the original test. It was expected that the LFC scale would be slightly more demanding than that of IE which aimed only at minimal intelligibility and that was generally the case. This partly confirms our hypothesis.

It is worth mentioning that both samples with a failing mark in the original test also received failing marks when assessed following the LFC and IE scales as well. This seems to reflect that the tolerance threshold is quite similar, no matter the framework against which the samples are assessed. These results do not lead us to confirm that the NP is deeply entrenched in L2 pronunciation teaching in our department, as was initially expected.

The first research question aimed at comparing the results obtained from using the LFC and IE scales. As expected, the grades assigned to students with the LFC were generally lower than those assigned using the IE scale with up to three points in difference. The case of H1 is worth mentioning. This sample was granted a full mark in the original test and when rated against the features of IE. This student is originally from California and, a close phonological analysis of his performance reveals that, due to the variety of English that prevails in his birthplace, he did not use the British [t], described in the set of core features (see Table 1). This may be the reason why he was penalized. This finding could show that even NS need to adapt and show flexibility in order to become fully competent speakers in international contexts.

After comparing the results shown by the original test to those on the basis of the LFC and IE, it can be concluded that the scores obtained in the original test are generally lower. As expected, all students were granted higher marks when assessment was based on core features in IE, especially those who received a narrow pass in the original test. These two samples were granted three and four points more when assessed using the IE core features; these being the highest differences observed in the results.

Samples H2 and S2 are interesting cases for an in-depth analysis since they received the same mark in the original test and when assessed on the basis of the LFC. The fact that the score did not change can probably be explained by the nature of the mistakes made by those speakers. As seen in Table 6, those mistakes were a target feature in both the course syllabus in the original test and in the LFC. According to the assessment criteria used in each case, these speakers were penalized for not producing a clear contrast between minimal pairs, for substituting consonant sounds and for eliding sounds in final position. The number of occasions in which this occurred, among other features assessed in the original test, such as allophonic variations and fluency, seems to account for the differences in score obtained by these students.

Table 6

Some mistakes detected in H2 and S2

NATURE OF MISTAKE		EXPECTED PRODUCTION LFC AND IE	ACTUAL PRODUCTION
Difference in vowel length and quality	Beat	[bi:t]	[bɪt]
	Turn	[tɜ:n]	[ten]
Consonant substitution	Day	[**d**eɪ]	[**ð**eɪ]
	Large	[lɑ:**dʒ**]	[lɑ:**ʃ**]
Final sound elision	Bought	[bɔ:**t**]	[bɔ:]

It was particularly interesting to notice how failing marks in the original test also implied a failing mark in the LFC and IE scales. This was due to many extracts being unintelligible more than to the inaccurate production of the features described in both the LFC and IE cores. As seen in Table 8, these students were also penalized for incorrect stress placement in polysyllabic words and for the inaccurate production of a great number of features from both cores (aspirated voiceless plosives, unwanted elision of final sounds, vowel length differences in minimal pairs, sound substitution).

Table 8

Some mistakes detected in the samples that received a failing mark.

NATURE OF MISTAKE		EXPECTED PRODUCTION LFC AND IE	ACTUAL PRODUCTION
Difference in vowel length and quality	Leaves	[li:vz]	/lɪvz/
	Easy	[i:zi]	[ɪsi]
	Signs	[sa:ɪnz]	[sa.ɪnz]
Consonant	Theatre	[θɪətə]	[tɪətə]

substitution	Journey	[dʒɜː.ni]	[ʃɜː.ni]
	Random	[rændəm]	[rændəs]
Lack of aspiration	Trouble	[tr̥ʌbl]	[trʌbl]
Sound elision	Borrowed	[bɒrəʊd]	[bɒrəʊ]
	Scared	[skeəd]	[skeə]
Unintelligible extracts	Lose, various, beginning, figure, heard, third, through	------	Pick up some ...? ... was awkward? Not has comfortable?
Inappropriate word stress	Entertainment	enterTAINment	ENtertainment
	Desperate	DESperate	despeRATE

The second research question aimed at finding out how different the baseline for approval would be when comparing the results of the three methods of assessment. A comparison of the ratings, as seen in Table 5, seems to reveal that the yardstick of minimal intelligibility set for international standards may also apply to the assessment of students on English programmes at UNC. The students who were not awarded a passing mark in the original test, also received a failing mark when using the LFC and IE scales. As explained before, the baseline for approval was not how close the sample was to a native standard, but whether students were able to make themselves understood while carrying out the two tasks proposed by this exam. The lack of a significant divergence between the results yielded by the original tests and those of the LFC and IE scales shows that intelligibility is at the heart of all three assessment methods, a clear advantage when giving feedback to students and shedding light into the assessment criteria of the examining board.

The ratings in Table 5, together with the accompanying comments, answer the third research question: Is the evaluation of pronunciation at the School of Languages, more closely aligned with the NP or the IP? It is a fact that the assessment standard at university will be higher since we are training future professionals of the language; however, even if we may believe that the NS speaker model is still present in our assessment practices, the IP also seems to be present as the baseline for approval. This seems to reveal that raters in the subject English Pronunciation Practice are attuned with current practices.

Conclusions

There are some limitations to this study. In the first place, the small number of samples analysed on this occasion may only derive preliminary results. The analysis of a larger corpus in future studies may provide more conclusive findings. Secondly, these samples only included monologic productions which, thus, lack evidence of accommodation strategies among interlocutors to negotiate and interact in a community of speakers (Pennington & Rogerson-Revell, 2019). A variety of individual and collaborative interventions will definitely allow for speakers to give evidence of the whole range of competences described by Ur (2012). A third limitation may derive from the characteristic of the raters that participated in this study. They share the same L1 as the speakers whose samples were analysed, they are all trained professionals in the field of phonology and they were familiar with the objectives and research questions that guided the study. The outcome of the study might have varied, though, had the raters been different. Despite these limitations, we may derive preliminary conclusions from the analysis that, from now on, can guide further research in the field of assessment and inform assessment practices in English Pronunciation Practice.

Knowing that the teaching and assessment of pronunciation go hand in hand in the English classroom, it is important to acknowledge how assessment impacts on previous and future teaching. This comparison of ratings has only been useful in as far as it allows for further research into this area.

The in-depth analysis of the selected samples under this new light has reinforced the grounds teachers rely on for offering feedback, especially to those students who do not receive passing marks. Besides, the fact that pronunciation teachers have been in contact with different frameworks of reference, as it has been the case in this study, has proved to be beneficial so as not to set unrealistic or unattainable goals for both teachers and students. Finally, it is clear that the standards of achievement should be higher at university level than, for example, when English is taught for general purposes. However, the findings suggest that the criteria followed in the assessment of English pronunciation at university also seems to be changing for the better and getting more attuned with the role English and English speakers have in the world at present.

References

Cruttenden, A. (2008). *Gimson's pronunciation of English* (7th ed.). Arnold.

Cruttenden, A. (2014). *Gimson's pronunciation of English* (8th ed.). Routledge.

Harding, L. (2017). What do raters need in a pronunciation scale? The user's view. In T. Isaacs & P. Trofimovich (Eds.), *Second language pronunciation assessment*, (pp. 12-34). Multilingual Matters.

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford University Press.

Jenkins, J. (2015). *Global Englishes. A resource book for students* (3rd ed.). Routledge.

Levis, J. M. (2005). Changing Contexts and Shifting Paradigms in Pronunciation Teaching. *TESOL Quarterly, 39*, 369-377.

Levis, J. M. (2018). Setting Priorities. What teachers and researchers. In J.M. Levis (Ed.), *Intelligibility, oral communication, and the teaching of pronunciation* (pp. 33-57). Cambridge University Press.

Levis, J. M. (2020). Revisiting the Intelligibility and Nativeness Principles. *Journal of Second Language Pronunciation*, 6(3), 310–328.

Munro, M.J. and Derwing, T.M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45,* 73-97.

Munro, M., & Derwing, T. (2015). *Pronunciation fundamentals. Evidence-based perspectives for L2 teaching and research*. John Benjamins Publishing Company.

Pennington, M., & Rogerson-Revell, P. (2019). *English pronunciation teaching and research. Contemporary perspectives.* Palgrave Macmillan.

Seidlhofer, B. (2011) *Understanding English as a lingua franca*. Oxford University Press.

Ur, P. (2012). English as an international language: Implications for classroom teaching and teaching materials [Conference session]. L'enseignement des langues étrangères dans une société multiculturelle: quelle place pour le français? Bar-Ilan University, Israel. https://www.youtube.com/watch?v=yTidAm0dRR0

In-service Training on Teaching and Learning Formulaic Sequences and its Assessment

Magdalena Zinkgraf and Gabriela Fernández

Facultad de Lenguas, Universidad Nacional del Comahue

This paper describes a free yearly course taught by research team J033 at Comahue University during 2019 for primary- and secondary-school teachers of English and some insights gained through its formative and summative assessment (Brown, 2004). The course aimed to equip participants with theoretical-practical knowledge on how to teach contextualised formulaic sequences (Wray, 2002) (FSs) through authentic songs, legends and myths, and motivate them to engage in a future pilot experience teaching formulaic language in their own EFL settings.

The paper first discusses some theoretical issues related to formulaic sequences and their teaching in EFL contexts. Then it describes the course in terms of its aims and content, while also considering the different tasks completed by participants for formative and summative assessment. As instances of formative assessment (Brown, 2004), there were face-to-face and virtual group discussions in which teachers discussed key concepts in the bibliography on FS instruction, and oral presentations in which participants described sets of tasks designed for the introduction of FSs in their teaching contexts. Formal summative assessment consisted of three real-life tasks involving the implementation of central content in the course, which were meant to provide an insight into teachers' grasp of the underlying principles of FS teaching (Lindstromberg & Boers, 2008) and their application in EFL classrooms and, by extension, into the extent to which the workshops' objectives may have been accomplished.

The paper analyses some common features in the practical assignments submitted by participants for the assessment of the course. Tasks like summarising the theory to justify decisions on the sequences to teach and relating this to the aims of the tasks they designed posed some challenges to trainees. These challenges are examined and some implications are also discussed in terms of teaching FSs in in-service teachers' contexts.

Theoretical Background

Formulaic Sequences and the Need for Explicit Instruction in EFL

An important field that has rapidly advanced in applied linguistics research concerns the role of larger units of meaning than single lexical items and their teaching. Throughout the last three decades, traditional views on the nature of the lexicon and its units have shifted and given way to a conception of language no longer made up of grammar rules according to which vocabulary fills the gaps. Through the accumulated findings in corpus linguistics in the last three decades, language has been proved to be composed to a large extent by strings of words that speakers of an L1 community share, are familiar with, understand, expect to hear and use. These chains have received a number of names and different terms have been coined. In this paper, we use the term 'formulaic sequences' defined by Wray (2002) as

> a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar (p. 9).

These strings convey specific meanings familiar to the L1 community and usually fulfil pragmatic or discourse functions. Their pervasiveness in language is such that they have been found to make up more than 50% of written language (Erman & Warren, 2000). Siyanova-Chanturia and Pellicer-Sanchez (2019) offer a more complete definition for formulaic language:

> strings of letters, words, sounds [...], contiguous or non-contiguous, of any length, size, frequency, degree of compositionality, literality/figurativeness, abstractness and complexity, [...] that necessarily enjoy a degree of conventionality or familiarity among (typical) speakers of a language community or group, and that hold a strong relationship in communicating meaning (p. 5).

Following these same authors, in this article, the term *formulaic sequences* will be adopted to refer to instances of formulaic language that meet the requirements in the definition above.

Raised awareness of the pervasiveness of these sequences in language has led to a body of research into the functions these units of meaning fulfil in discourse, which are central to speakers' fluency. They also aid comprehension within a language community, freeing up speakers' processing load to help them concentrate on the meanings to be conveyed (Lin, 2019).

Teaching Formulaic Sequences in Foreign Language Settings

Although formulaic language is ubiquitous in every language for the native speakers of that language community, to foreign language learners, FSs become an important obstacle to sounding natural and being understood. Research has shown that even proficient learners of English find it difficult to express ideas by means of the typical and frequent phrases used by native speakers of the language (Alhassan & Wood, 2015; Erman, Forsberg Lundell & Lewis, 2016). They may find it difficult to spot them in the input, and therefore, it may take some time for them to realize the communicative power they hold. While speakers of a language will automatically select a phrase to transmit a whole idea, EFL learners may struggle combining isolated words to convey meanings. This seems to be the case even in university settings with otherwise highly proficient learners (Jones & Haywood, 2004; Zinkgräf & Verdú, 2015). Consequently, researchers like Boers and Lindstromberg (2012) suggest that explicit instruction of FSs should take up part of our class time and that EFL learners' attention needs to be directed to these phrases, their form and meanings.

The obvious contextual differences between learning an L1 and a foreign language highlight certain conditions that foreign language teaching should contemplate to help learners master FSs and authors agree that the teaching of FSs is essential to guarantee effective, natural and appropriate EFL use. Boers and Lindstromberg (2012) and Wood (2015) advocate FS-focused instruction in the classroom to optimize learners' formulaic language acquisition. For this instruction to be fruitful, a number of conditions need to be met: learners' awareness needs to be raised with respect to the existence and pervasiveness of FSs. In order to do so, explicit focus on the selected sequences should target not only the form of the sequence and its meaning but also certain discourse and pragmatic requirements that call for the use of such a sequence in that particular context. FS instruction will demand a selection of those FSs that both course developers and teachers should judge to be useful and pedagogically relevant to the communicative needs of their learners and to the purpose of the course (Boers & Lindstromberg, 2009). Contextual features like these and learners' individual characteristics may need special consideration in this selection.

FS-focused instruction borrows much from the original Focus-on-Form approaches (Doughty & Williams, 1998; Long & Robinson, 1998) which were implemented in the 90s for the teaching of grammatical features in a foreign language, but in this case, the focus is placed on formulaic sequences. Teachers need to provide learners with input that is seeded with contextually appropriate uses of the selected FSs in the L2, guaranteeing multiple encounters with the target sequences. To aid perception of these phrases, it is advisable to modify the input typographically (through bold type, highlighting or underlining) and give them saliency (Alharthi, 2015). It is also imperative that FSs are presented in both written and oral contexts since prosodic features typical of language production in the oral mode will contribute to learners' perception of FS boundaries and make these phrases become more entrenched in memory (Lindstromberg & Boers, 2008; Pellicer-Sánchez & Boers, 2019).

Following Nation's (2001) processes for the learning of vocabulary - single lexical items -, Lindstromberg and Boers (2008) and Boers and Lindstromberg (2009) offer a set of activities and procedures grouped under three different stages: noticing, retrieval and generative use (Hatami, 2015). As mentioned above, for learners to perceive FSs in the input, their attention needs to be drawn to the occurrence of these sequences in context, and to their form and meaning (noticing). Later, opportunities for learners to retrieve both the FS's form and meaning from memory are necessary to help them strengthen its form-meaning link (retrieval). At a subsequent stage, tasks where learners can meet them in new, unfamiliar contexts or use these sequences creatively (generative use) are essential for them to "reconceptualize the knowledge" of the focused phrases (Nation, 2001, pp. 68-9).

Selecting which FSs to Teach

Lindstromberg and Boers (2008) stress that certain FSs are more classroom-worthy than others and specify some criteria for teachers to select which FSs to focus on in their lessons: frequency, and linguistic motivation or memorability (p. 15). Boers and Lindstromberg (2009) on their part consider that FSs' usefulness is also a determining factor: based on the courses' characteristics, some sequences may be more necessary for learners to master and useful to convey specific meanings. They recommend that medium-frequency sequences should be the target of explicit instruction and so should more fixed sequences, which offer little choice of alternative components and therefore are more likely to be successfully remembered.

These criteria need to be combined with what they call 'teachability' (Boers & Lindstromberg, 2009, p. 68): some FSs are more susceptible to explicit instruction and better recall rate due to some degree of figurative meaning within their component words. Some chunks are linguistically motivated in that they contain alliteration in their make-up, a certain type of rhythm, or assonance that make them very suitable candidates to become better recalled with the teacher's help. Some other sequences lend themselves to the type of interpretation work based on their (metaphorical) "imageability". This inherent characteristic of some FSs needs to be exploited by teachers through activities that demand certain cognitive elaboration of FSs and that "engage them [learners] in the kind of rich, deep, 'elaborative' mental processing which is known to enhance retention of vocabulary in memory" (Lindstromberg & Boers, 2008, p.13). Wood (2015, p.143) also outlines a number of strategies that can be developed in EFL learners to help these sequences become entrenched in memory and automatized for future use.

The course

This in-service training course was planned as part of the activities of Research Project J033 at Facultad de Lenguas, Universidad Nacional del Comahue. The project had several aims, of which one was to raise EFL teachers' awareness of the significance of formulaic language in foreign language education.

Characteristics, Goals and Participants

Eleven teachers and four teachers-to-be from the provinces of Río Negro and Neuquén (Argentina) completed this nine-month-long in-service training taught by members of the Research Project J033 at Facultad de Lenguas, Universidad Nacional del Comahue. The course was intended for EFL primary and secondary school teachers and was taught from March to December in 2019. Its central aims were to a) equip participants with theoretical-practical knowledge on teaching contextualised FSs through authentic songs, legends and myths and b) engage them in a future pilot experience teaching formulaic language in their own EFL settings. In this experience volunteering teachers who had completed both modules would implement a formulaic approach to the teaching of vocabulary through legends and songs in their own primary school courses. They would be teaching lessons based on the material the research team had especially designed based on the tenets of FS-focused teaching (Lindstromberg & Boers, 2008; Hatami, 2015; Boers & Lindstromberg, 2009). Unfortunately, due to the spread of Covid-19 and the subsequent lockdown, the pilot experience had to be postponed.

Participation in the course was free of charge since it aimed at offering as wide an audience as possible the chance to engage in training in formulaic language teaching. The course was divided into two modules: a) Theoretical Foundations and Planning: Teaching English through Chunks in Legends and Songs and b) Teaching, Assessing and Researching Chunks through Legends and Songs in English. Each 35-hour module consisted of five monthly face-to-face meetings running five and a half hours each and allotted an extra hour and a half per meeting to the reading of selected material and virtual participation through the university's Moodle platform. This two-module course followed Nation (2001)'s and Lindstromberg and Boers (2008)'s three-step approach to vocabulary learning, which consists of providing EFL learners with opportunities for a) identification, b) retrieval and c) creative use of FSs.

Assessment in the Course

In an attempt to determine the impact of the modules on participants' conceptual framework, both formative and summative assessment tools were adopted during this in-service training experience for EFL teachers on teaching formulaic sequences. Because each of the modules covered different content and had different aims, and because each one needed to be certified institutionally separately, a written summative assessment task was included in both modules.

Formative Assessment Tasks. Brown and Abbeywickrama (2018) define formative assessment as the process involved in "evaluating participants in the process of "forming" their competencies and skills with the goal of helping them to continue that growth" (p. 8). With this purpose in mind, participants in the course were asked to complete some tasks that served as tools to determine their gradual appropriation of the content of the modules. These tasks implied different degrees of difficulty and, as the course progressed, they involved more concepts presented in the meetings.

At different stages throughout both modules, participants analysed some FSs present in songs and in legends, assessed how relevant they might be to learners in their teaching contexts and compared their teachability in terms of the criteria mentioned above. The video-recorded presentations of these group tasks were later input for trainers to explore participants' understanding of some of the concepts presented in the literature. This was especially done through the discussion of the theory they had read for one of the workshops: they considered Lindstromberg and Boers (2008)'s recommendations (pp. 15-16) so as to draw implications that related to their EFL teaching in context.

Trainees were also invited to analyse the applicability of different types of activities to suit the different stages in Nation's (2001) vocabulary learning process and to justify their reasons for the tasks to promote either noticing, retrieval or creative use of the sequences. These sample tasks were the basis on which they were later to design their own. After the discussion during the face-to-face meeting of the theoretical issues presented, they set to redesign the activity to contemplate these concepts and recommendations. In every case they were required to describe a possible context for the tasks, to contribute to raising their awareness of how context-dependent these activities are.

Most of the formative assessment tasks took place in the monthly face-to-face meetings and video-recordings of these meetings offer evidence of participants' interactions when solving these activities, and of their conclusions and doubts. Their comments in those recordings point to key concepts that might have generated some misunderstandings or difficulties, which will be later described.

Summative Assessment Tasks. Another fundamental component was the incorporation of assignments that "aim to measure, or summarize, what [a course participant] has grasped," (Brown & Abbeywickrama, 2018, p. 6). These tasks were part of the certification requirements set by the Consejo Provincial de Educación. In this case, participants completed three real-life tasks (one after module 1 and two after module 2) which intended to assess whether and to what extent they had understood and could apply the core contents of learning and teaching FSs.

At the end of the first module (Theoretical Foundations and Planning: Teaching English through Chunks in Legends and Songs), two tasks were administered. Firstly, participants were asked to choose to work with a fragment of either the lyrics of a song or a legend. On the basis of their choice, they were required to select four FSs they would present to one of their groups of students and a) briefly describe this context so that trainers could get a clear picture of their teaching scenario and b) justify their selection of FSs by referring to the materials they had read throughout the different sessions and the principles in the literature for such a selection. They were also tasked with designing two activities based on the song/legend and the set of FSs selected and briefly explaining a) their learning objectives, b) the general vocabulary learning processes outlined in Nation, (2001) and Lindstromberg and Boers (2008) that these activities developed and c) their reasons for including these tasks in a teaching sequence.

Participants were required to complete Task 1 individually since it was believed that they might not be sharing the same school setting or context in which they would like to implement the teaching sequence they would design (See Appendix). These assignments were submitted by email. Trainers read them carefully and provided written feedback consisting of comments, suggestions and an overall, impressionistic mark. The main objectives mentioned for the module were the guiding criteria for the marking of these assignments. The submitted assignments first needed to comply with each of the subtasks required, and then show evidence of participants' uptake of the module's content and a personal reflection showing their learning gains and teaching challenges as to the implementation of a FS-based approach in their own contexts. If necessary, participants were invited to resubmit their tasks fifteen days after the feedback.

The second task (after Module 2) required them to teach a simulated lesson on FSs based on a song or a legend to A1 level children outlined in the lesson plan crafted by the trainers. This task was meant to exemplify the role some of these teachers would play in the pilot experience that was supposed to take place in 2020.

As to the final assessment of the second module (Teaching, Assessing and Researching Chunks through Legends and Songs in English), in Task 3 participants were asked to write a report on the key aspects covered by the literature in the whole course. They were expected to a) review the content explored in the module and how it related to those in the first module of this training, b) support their analysis by referring to the reading materials, c) describe the implications of adopting a FS-based approach in their own teaching contexts, and d) consider the likelihood of its implementation in their own schools as part of this experience.

Some Insights into Participants' Perceptions of the Teaching of FSs

In this section we discuss some comments made by participants that shed light on their understanding of the material presented in the formative assessment tasks. A more in-depth analysis of participants' summative Tasks 1 and 3 follows.

Formative Assessment

Formative assessment tasks were completed mainly in our face-to-face meetings. Based on trainers' observation of trainees' behaviour during the completion of the tasks described in the previous section, it was evident that participants were engaged in the course: they systematically attended sessions and participated in tasks, group and plenary discussions. Their participation was sustained also in forums on the platform and trainees were always up to date with the key readings set from one session to the next. In their active partaking of discussions, they voiced some of the difficulties they found when relating the theoretical concepts presented in the literature to practical aspects. For example, faced with the task to "Select which FSs from this legend they would choose to teach in their context" (second face-to-face meeting during Module 1 in the course), they expressed some doubts such as

(1) "I was not sure...We found many [FSs]" (Participant A)
(2) "The concept [chunk] is a bit broad. What is a chunk? Which are high priority?" (Participant B)

Although the concept had already been presented extensively in the first meeting and key readings introducing it had been analysed, the application of theory to a specific task made them confront these doubts, which triggered further discussion and analysis with trainers. On the other hand, in the same task, Participant C mentioned one strategy in particular that Lindstromberg and Boers (2008) suggest as contributing to the organisation of the vocabulary to be focused on and, therefore, to the ease with which learners can better recall these sequences: grouping sequences based on semantic fields. This is evinced in (3):

(3) "We selected many and grouped those ones related to 'talking'." (Participant C)

These comments show participants' gradual uptake of the concepts discussed in this second face-to-face meeting on FS selection and teachability. They show how, for some, the concept was so broad that they found it difficult to decide on which FSs to focus. For others, the reading materials had equipped them with practical tools to solve the task.

When trainees were asked to analyse specially designed sample tasks to focus on the formulaic sequences in the selected songs and legends, and, in turn, decide how these could be adapted and implemented in their own educational settings, they mentioned their own learners' characteristics and took measures to cater for some possible shortcomings in the tasks. This is reflected in (4) and (5):

(4) "I was thinking of presenting words in isolation so…" (Participant I)
(5) "I'm not sure if they know some vocabulary. Maybe I need to pre-teach some words." (Participant K)

These statements are taken as evidence of a re-elaboration process on the part of trainees, who were struggling to reconcile both views of language and teaching: focusing on individual words and on FSs.

Summative Assessment

A quick overview of the final assignments (Task 1) submitted by participants for the first module indicates that four trainees successfully completed and passed the task. The remaining nine were asked to take into account the comments, questions and suggestions included in the feedback and to resubmit two weeks later. Some advanced undergraduates taking the course handed in a third version of their work.

Some of the subtasks in Assignment 1 appear to have presented trainees with some degree of difficulty. Justifying their answers and decisions in the task designed while referring to the set readings posed a challenge and at times they had trouble deciding which concepts in the theory presented could help them account for task design. We have included excerpts of the assignments submitted by some of the trainees where reference to the literature is missing or its relevance to the decisions made, unclear.

In Activity 1 designed by Participant E for Task 1, there is evidence of the uptake of concepts like 'noticing' and its importance. However**,** little theoretical background is provided for this activity and no reference is made to Nation's approach to vocabulary teaching (2001). Also, although it is true that this first activity aims at noticing, there is evidence in the task of some form of 'retrieval' fostered as well, when the trainee mentions that its aim is to help learners "remember them by means of singing the song once". Participant E seems unaware of this other process involved in her activity.

> "The students will be given a copy with the song "Shine" by Jason Mraz with empty gaps to fill in with the four chunks in a box. They will listen to the song and complete it with the phrase they listen to. Ss will exchange the copy with his/her classmate so as to check the gaps they completed listening to the song one more time. Then, all together will sing the song. On the blackboard, the teacher will write four definitions for each chunk worked on the song. The students will be given sheets of paper with the phrases and they will stick them in the definitions. Then, they will check together.
>
> The main objective for these activities is to foster students´ abilities to talk about the topics friendship and support. The first activity is aimed at helping students identify the chunks and remember them by means of singing the song once**.** The **process involved in the learning of vocabulary in this task is noticing** since students encounter the chunks for the first time and they see them repeatedly throughout the song.

Besides, students will match the phrases with their definitions on the blackboard which makes them enhance the vocabulary." (Participant E)

In the following excerpt from Participant G's Task 1, it is evident that she understands the inherent interrelation between FS selection and the characteristics of the course where they are to be taught. However, little information is offered on the found frequency for the selected sequences and how she has used that information in the task designed.

The selection of these chunks has been made taking into account first of all the ages and interests of the students I work with and also the level of frequency these ones have in real life communication. (Participant G)

Contrarily, Participant I was very careful to include the results of her own analysis of the relative frequency of possible chunks to select as shown in Table 1 below.

Table 1.

FS Frequencies (Participant I)

CHUNK	WWW	DICTIONARY	BNC	COCA
stopped talking	16,700,000	YES	77	696
fell hopelessly in love with	42,100	YES (fall in love with sb)	1 (fell hopelessly in love with her) 151 (fall in love with)	5 (fell hopelessly in love with) 1191 (fall in love with sb/sth)
caught sight of	3,480,000	YES (catch sight of sb/sth)	263	776
leave me alone	31,900,000	YES (leave sb/	163	1107

		sth alone)		

She also refers to this criterion (frequency) in combination to Boers and Lindstromberg's (2009) utility, basing this connection on the readings for the course.

> As regards the second criterion, utility, Boers & Lindstromberg (2009) state that "which chunks are particularly useful for a given group of students depends ultimately on their level of proficiency and on the objectives of the course they are taking."[1] Taking into consideration the first aspect, the chunks selected contain words that might result familiar to the students in the 3rd year of secondary school who have an elementary-intermediate level of English proficiency; therefore, these chunks would be easy for learners to remember them. (Participant I)

However, some degree of misunderstanding and misapplication of concepts in the literature was also found in their justifications for their choice of activities (see Participant B's excerpt below).

> The teaching of these chunks will be useful since chunks facilitate fluent language production and achieve communicative purposes successfully. The mastery of them can help students sound more natural as they will sound linguistically accurate. They "appear to be stored in the long-term memory as single memorized units (i.e., prefabricated chunks or ready-made wholes) and can therefore be processed and accessed more quickly and easily than the same sequences of words when generated creatively.' (Sarvena Hatami, 2014). (Participant B)

Although a citation of the literature analysed has been included, the purpose of this quote is not clear, since no connection is established with the activities proposed or the justification behind them. The implications for the tasks designed that derive from this general quote are not made explicit.

Task 3 in the summative assessment of Module 2 asked participants to reflect on their learning experiences from both modules. Participant A submitted a very thorough reflection that connected the content of all the sessions in the training course, of which an excerpt is included below.

> "This new approach brings about Nation's (2001) three psychological processes needed to help students remember these sequences of words, i.e. Chunks: noticing, retrieval and generative use. As a start, we got a deep introduction to the study of words (Nation, 2001), the implications of learning vocabulary (Schmitt, 2010), last but least, the notion of chunks (Lindstromberg and Boers, 2008). In later meetings, we carried out activities designed on the basis of Nation's guidelines for effective vocabulary learning which were suitable for primary students concerning with the short time of the classes, students' interests and a variety of activities. Needless to say, the chosen songs and legends were presented in an appropriate way. Both included different instances of presentation of vocabulary, repetition of the selected chunks and matching activities for instance, to be carried out in three steps: before, during and after dealing with them. In addition to this, the content of the songs and legends are related with students' personal experiences and in this way, this methodology helps students foster their creativity through hypothesizing and giving personal opinions on certain situations, thus making the learning of the chunks memorable for the learners." (Participant A)

In this final task for Module 2 in the course, Participant A threads together the different reading materials explored and analysed in the course, highlighting their most important contributions to the teaching of formulaic sequences, and linking these to the tasks completed in the face-to-face meetings. In her assignment, she justifies task design purposes and FS selection in terms of the context where trainees work. This excerpt offers a wider perspective than Task 1 by one trainee in particular of the whole in-service training course in retrospective. The long-term learning process seems to have been positive and clarifying, even if in Task 1 difficulties and obstacles were encountered.

This section has offered some insights into participants' appropriation process of the content of the in-service training course taught during 2019 in justifying their decisions and designs in the light of the theory discussed. These challenges, however, reflect participants' first submission of the task. Through a process approach to their assignments, and guided by feedback offered by the course's trainers, they overcame these difficulties and achieved the goals set for the course. Their final versions and Task 3 solved many of the issues highlighted in the excerpts above, and show evidence of their intake of the main content and activities of this training experience on the teaching of formulaic sequences in EFL settings.

Implications for the Future

Extensive research into language and vocabulary learning and teaching has confirmed the ubiquity and usefulness of FSs. On these grounds, it is imperative that EFL teachers are informed about these findings and trained to help their students to incorporate these strings of words into their active repertoire. Because FSs are "just "big words" (Thornbury, 2019, p. 6), most principles of effective vocabulary teaching apply equally to the teaching of chunks as they do to the teaching of individual words (Lindstromberg & Boers, 2008). Therefore, teachers will need to make sure they a) teach chunks in context, b) clarify both their meaning and form, c) allow for opportunities to practise them in guided and meaningful contexts, d) review and recycle them systematically and e) equip learners with a set of strategies for the identification, recording and recycling of FSs (Lindstromberg & Boers, 2008). These methodological priorities in the teaching of vocabulary are captured by Nation's (2001) three-stage approach, based on EFL learners' a) recognition, b) controlled intensive practice and c) generative application in true-to-life scenarios of the target vocabulary items. As to the criteria for selecting and sequencing chunks, EFL teachers need to be sensible and practical to draw their students' attention to those set phrases that are a) relatively frequent, b) useful and relevant to their learners' communicative needs, c) memorable for their contexts and background knowledge and d) teachable in the EFL classroom. Their learners' age, level of proficiency and the purpose of the course should also guide teachers in this selection. As learning vocabulary is usually laborious with many learners, these informed decisions should be combined with small doses of FSs to focus on and with multiple natural opportunities for EFL learners to encounter them and identify their boundaries, patterns, and usage restrictions.

In sum, since FSs are intrinsic components of the lexicon, the greatest challenge is to raise EFL teachers' awareness of "their pervasiveness [...] and potential in promoting fluency (Granger & Meunier, 2008, p. 248)" so that learners can be aided in their own perception, practice and use of FSs. In this respect, in-service teacher training plays a pivotal role.

Notes

This paper is based on Research Project J033, Facultad de Lenguas, funded by Universidad Nacional del Comahue.

References

Alharthi, T. (2015). Adding more fuel to the fire: A study of attrition in formulaic sequences by adult learners. *Arab World English Journal* (*AWEJ*), 6(3), 230-243. https://dx.doi.org/10.24093/awej/vol6no3.15

Alhassan, L., & Wood, D. (2015). The effectiveness of focused instruction of formulaic sequences in augmenting l2 learners' academic writing skills: A quantitative research study. *Journal of English for Academic Purposes*, *17*, 51-62. https://doi.org/10.1016/j.jeap.2015.02.001

Boers, F., & Lindstromberg, S. (2009). *Optimizing a lexical approach to instructed second language acquisition*. Palgrave Macmillan.

Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics, 32*, 83-110.

https://doi.org/10.1017/S0267190512000050

Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Pearson Education ESL.

Brown, H. D., & Abbeywickrama, P. (2018). *Language assessment: Principles and classroom practices*. Pearson Education ESL.

Doughty, C., & Williams, J. (1998). *Focus on form in classroom second language acquisition*. Cambridge University Press.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text and Talk*, *20*(1), 29-62.

Erman, B., Forsberg Lundell, F., & Lewis, M. (2016). Formulaic language in advanced second language acquisition and use. In K. Hyltenstam (Ed.), *Advanced proficiency and exceptional ability in second languages* (pp. 111-148). Walter de Gruyter.

Granger S., & Meunier, F. (2008). Phraseology in language and teaching: Where to from here? In F. Meunier, & S. Granger (Eds.). *Phraseology in foreign language learning and teaching* (pp. 247-252). John Benjamins Publishing.

Hatami, S. (2015). Teaching formulaic sequences in the ESL classroom. *TESOL Journal*, *6*(1), 112-129. https://doi.org/10.1002/tesj.143

Jones, M., & Haywood, S. (2004). Facilitating the acquisition of formulaic sequences: An exploratory study in an EAP context. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 269-300). John Benjamins Publishing.

Lin, P. (2019). Formulaic language and speech prosody. In A. Siyanova-Chanturia, & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 79-94). Routledge.

Lindstromberg, S., & Boers, F. (2008). *Teaching chunks of language. From noticing to remembering*. Helbling Languages.

Long, M., & Robinson, P. (1998). Focus on form: Theory, research and practice. In C. Doughty, & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15-40). Cambridge University Press.

Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

Pellicer-Sánchez A., & Boers, F. (2019). Pedagogical approaches to the teaching and learning of formulaic language. In A. Siyanova-Chanturia, & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 153-173). Routledge.

Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (2019). Formulaic language: setting the scene. In A. Siyanova-Chanturia, & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 1-15). Routledge.

Thornbury, S. (2019). Learning language in chunks. *Cambridge Papers in ELT series*. [pdf] Cambridge University Press https://www.cambridge.org/elt/blog/wp-content/uploads/2019/10/Learning-Language-in-Chunks.pdf

Wood, D. (2015). *Fundamentals of formulaic language: An introduction*. Bloomsbury.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.

Zinkgräf, M., & Verdú, M. A. (2015). Formulaic sequences involving 'fact' in EAP production: A corpus study. In L. Anglada, N. L. Sapag, D. L. Banegas & M. A. Soto (Eds.), *EFL classrooms in the new millennium: Selected papers from the 40th FAAPI Conference* (pp. 58-69). https://www.faapi.org.ar/wp-content/uploads/2021/02/FAAPI2015.pdf

Appendix

Task 1

Ciclo de Formación 1: "Theoretical Foundations and Planning: Teaching English through Chunks in Legends and Songs" (transferencia del Proyecto 04/J033, UNCo)

FINAL ASSIGNMENT_

1. For the development of this task you must choose whether to work with part of the lyrics of a song or a legend (see the material below).

Part of the song "Shine" by Jason Mraz:

SHINE
Once upon a time there was a meeting of minds
The sun and the moon made a deal with the sky
One would take the morning and the other the night
Together they would blanket the world with light
But the moon had a shadow, he felt like a liar
The sun was the only one who carried the fire
The sun saw this, she kept on glowing
Bound to the moon, never saying, "you owe me"
She said I'll shine on you (shine on you)
I will shine on you (shine on you)
I will shine on you (shine on you)
I will shine on you
(...)

Legend: "Narcissus"

Echo was a wood nymph who was punished by Hera because she never stopped talking. She was cursed to only repeat what other people said to her. One day Echo saw Narcissus. She watched him. She had never seen such a handsome young man and she fell hopelessly in love with him. Every day she followed him, and he often caught sight of her. At first, he took no notice of her; he was very used to girls falling in love with him. Then he grew irritated that everywhere he went, Echo was there too.

“Go away. I don’t love you,” he shouted. “Love you, love you,” repeated Echo. “Leave me alone,” called Narcissus. “Alone, alone,” replied Echo.

She wandered sadly away through the trees and, as the weeks passed, she grew thinner and paler until she faded away altogether; nothing of her was left except her voice, which always repeated what anyone said.

The Impact of Electronic Indirect Feedback on the Writing Process of Undergraduate Students

Paul Alexis Carrera, Ana Cecilia Cad and Julieta Salinas

Universidad Nacional de Córdoba, Facultad de Lenguas

Academic writing is at the heart of the teaching and learning process in higher education as undergraduates are mainly assessed by what they write. Carlino (2005) claims that teaching learners academic writing is one of the most powerful methods to learn, so they must be guided in their process of learning how to write in the academy. Thus, providing them with quality feedback is essential to help them enhance their writing skills.

This qualitative study, which is part of a larger research project (Proyecto Consolidar 2018-2021, SECYT), aims to create an analytic scoring scale to provide students with highly effective and reliable feedback, so that they can develop strong writing skills. The scale employed in this research has been developed by the members of the English Language II chair at a university English programme. The analytic scale has been refined after being tested for validity and reliability (Dörnyei, 2007), and it is the first time that the scale is implemented with students using an embedded electronic rubric run by Moodle. Together with the scale, coded feedback and teachers' comments were used to provide students with further feedback.

The question that this research addresses is whether the combination of the analytic scoring scale and the teachers' comments has an impact on students' writing performance. Thus, this research aims to analyse the impact of electronic indirect feedback on undergraduate students' academic writing skills by means of the analysis of students' compositions and students' perceptions regarding the type of feedback received.

Theoretical Framework

Analytic Scale

Different researchers have explored the effectiveness of analytic scoring scales to assess second language writing. Reddy and Andrade (2010) state that these rubrics have three basic components: *evaluation criteria, quality descriptors* and a *scoring strategy*. The *evaluation criteria* are a set of factors that an instructor may consider when assessing their students' work. *Descriptors* are detailed explanations of expected behaviours to demonstrate a skill or proficiency. They introduce the criterion to achieve a certain level, such as excellent, good, fair or inadequate. Finally, scale *scores* contribute to the assessment of students' products and/or processes in a valid and reliable way.

An analytic scoring scale is a rubric that includes explicit performance expectation for each criterion and band. It allows instructors to assess individual aspects of a task and obtain one unified grade for the whole task. Analytic scoring rubrics allow students to have a clear understanding of the criteria used to assess them. This knowledge can help students set learning goals and develop strategies to improve language learning.

Reddy and Andrade (2010) also state that, when scales are used by students during instances of formative assessment, these tools fulfil the double purpose of teaching and evaluating at the same time. Some authors (Kroll, 1990; Weigle, 2002) state that these scales are particularly useful for second language writers as they can focus on different aspects involved in the writing process separately. Thus, by breaking down the assessment of writing into different components, teachers and students can measure particular aspects of the process. Besides, students can "understand the target for their learning and the standards of quality for a particular assignment", and also "make dependable judgments about their own work that can inform revision and improvement" (Reddy & Andrade, 2010, p. 437). Weigle (2002) also states that those students who are evaluated using an analytic scoring scale show more improvement than those assessed by other means.

Formative Assessment

Assessment is an essential component of the teaching-learning process. Authors such as Collins and O'Brien (2011) define assessment as "any method used to better understand the current knowledge that a student possesses" (p. 36). In general terms, assessment is mostly carried out by teachers through the use of different instruments: exams, homework, assignments, projects, etc. Hyland (2002, 2003) claims that this assessment as well as feedback contribute significantly to the development of writing skills and to an improvement in the learning process as such. Therefore, effective assessment procedures are necessary to ensure that students learn and develop the desired skills.

Formative assessment is performed along the course, and its main objective is to get feedback about the teaching-learning processes, their strengths and weaknesses (O'Malley & Pierce, 1996; Collins & O'Brien, 2003). According to Hyland and Hyland (2006), formative assessment in writing is "crucial for encouraging and consolidating learning" (p. 177). That is why formative assessment is a dynamic and continuous process in which teachers and students cooperate in the acquisition and development of skills as well as in the building of knowledge. It can be defined as an instance in which the process of learning is placed in focus to measure students' progress with the aim of making "decisions about the next steps in instruction that are . . . better founded '' (Black & Wiliam, 2008, p. 7). In view of Wiliam and Thompson's (2008) seminal work on assessment, formative assessment is made of five key strategies, namely explicitly announcing the teacher's assessment criteria, providing instances of practise that simulate assessment, providing constructive feedback, helping students assess one another and making students the protagonists of their own learning processes. Providing constructive feedback plays an important role in students' learning process as it orients them towards those aspects that need improvement while it reaffirms those aspects that have been successfully acquired.

Electronic Indirect Feedback

Electronic feedback can be defined as the means by which teacher feedback is provided through technology (Ware & Warschauer, 2006). It is possible to mention at least three benefits of electronic feedback. One is that this type of feedback provides the space to monitor student-teacher and student-student interactions (DiGiovanni & Nagaswami, 2001). The same authors also posit that it encourages students' participation. Hewett (2000) concludes that e-feedback provides the space to discuss more concrete and detailed issues as opposed to oral feedback. A type of feedback that could be used in combination with e-feedback is coded feedback.

Buckingham and Aktuğ-Ekinci (2017) define correction codes as "a form of guided indirect, metalinguistic feedback on learner writing" (p. 2). This type of feedback is provided with the aim of helping the learners review their knowledge about different linguistic aspects of their compositions that can be deemed as problematic by means of a coding system that both parties, students and teachers, are familiar with. According to these authors, one of the main benefits of this feedback is that it encourages the students to "develop a reflective, problem-solving approach to text revision" (p. 2). The different linguistic aspects that the learners may need to work on are related to syntax, punctuation, grammar, choice of vocabulary, among other categories. There are mixed results about the effectiveness of this type of feedback in the process of helping learners improve their writing skills. Although students deem receiving feedback necessary and they appreciate the linguistic clues received by means of written comments (e.g., Ferris & Roberts, 2001; Lee, 2008), students' ability to successfully solve linguistic problems seems to be dependent upon the degree of complexity of the error type.

Literature Review

In recent years, there has been an increasing interest in exploring the use of analytic scales in the field of assessment and particularly in the assessment of second language writing skills. Most of this research focuses on enhancing both validity and reliability of an analytic scale when different instructors are involved in the assessment process (e.g., Hattingh & van der Walt, 2013).

Other studies have focused on the impact that this analytic scale has on the development of students' writing tasks. In his study, Bolton (2006) states that undergraduate and graduate business students value the use of rubrics since they allowed them to become familiar with the course assessment criteria which, in turn, helped them to reduce their uncertainty, determine the work that a task required, self-assess their own performance, estimate grades prior to their assignment submission and improve their performance in coming assignments. Similar results were obtained in Andrade and Du's (2005) study with undergraduate students who claimed that using rubrics provided them with information to produce better quality work, achieve higher grades and reduce their anxiety about assignments. Some interesting findings from this study were that students did not read the rubric entirely and that some of them did not consider the assessment criteria as the set standards of a specific discipline but as the particular demand of a teacher.

Ahour and Mukundan (2009) explored the contribution of each writing component to the variance of their students' overall writing performance. The researchers assessed the writing samples of 128 TESOL undergraduate students in Malaysia using Wier's (1990) analytic scale. They found that the areas that exhibited the most variance were grammar, cohesion, vocabulary, spelling, punctuation and content while the component organization was the least impacted by the use of the scale. According to these findings, the use of the scale helped students become aware of the areas that they needed to improve the most.

Methodology

Participants and Context

This study was conducted with one group of second-year undergraduate students (B1+) from an English programme at the University of Cordoba, Argentina. They were students of an English course where academic essay writing is taught. Students need to master different types of essays: expository, cause and effect, and comparison and contrast. Instruction on writing and the coded feedback was provided online through virtual sessions and supported by materials on the Learning Management System Moodle.

Three teacher-researchers were in charge of assessing students' work using the analytic scoring scale for which they had received previous training. They also administered a questionnaire, moderated a focus group interview and conducted text analysis of essays. Other members of the larger research team analysed the data to ensure research validity and reliability.

Procedure

In the first semester of 2021, students wrote an expository essay on a word document following the instructions provided by the teachers. This task was uploaded to the virtual classroom. One week later, through this virtual environment, they received feedback on their work by means of teachers' comments using the comments feature in Microsoft Word, and the analytic scoring scale embedded in the feedback section of the assignment in the virtual classroom. Second year students in this teaching context are familiar with receiving feedback by means of teachers' comments; however, this was the first time they were introduced to the analytic scoring scale. After two weeks, students had to submit a second draft taking into account all the aspects that required their special attention. Students were not trained to use the analytic scoring scale.

The Analytic Scoring Scale and Teachers' Comments

The analytic scoring scale used in the study was adapted from Tribble (1996 in Kern 2000). Adaptations were necessary to make the scale suit the needs of our local teaching context (Canavosio et al, 2021). The version with which students were assessed presents four categories: *content appropriateness*, *organization*, *quality of expression and language accuracy* and *formal appropriateness and layout* (see Appendix). Each band was subdivided into four bands: Excellent/Very Good, Good, Fair, Inadequate and

Unacceptable. Each band has its descriptors that offer enough information for the students to understand why their work had been placed in a particular band. Each band accommodates a range of scores to ease the process of assessing students.

Teachers also provided students with indirect feedback using the comments feature on Word. This feedback focused on the same aspects as the band: *content appropriateness*, *organization*, *quality of expression and language accuracy* and *formal appropriateness and layout*. The main difference between these feedback tools is that, while the scale offered general feedback, teachers' comments provided students with more personalized feedback.

Data Collection Instruments and Analysis

Focus Group

A focus group is a data collection instrument that aims at exploring perceptions and ideas participants have about different topics. In keeping with Krueger (2002), we worked with carefully recruited participants in a comfortable environment. Thus, we decided to carry out the group interview in the participants' L1, so that language was not an obstacle in expressing their views. One of the teacher-researchers acted as moderator asking pre-determined questions that guided the interview and organized the interaction. The other two researchers recorded the session and engaged in note-taking, observing and noting any special behaviour that might be significant for the discussion. In this particular case, four students attended the focus group interview, which was carried out in a virtual environment, a Meet session. Some aspects that may have been observed in a face-to-face group interview, such as body language, were not that noticeable in this given context. However, there was an interesting exchange of ideas, feelings, perceptions and suggestions about receiving feedback with an analytic scoring scale.

The focus group interview was transcribed and the data were analysed using content analysis, which involved identifying salient patterns through a process of codification and categorization. The categories were then compared to the questionnaire results.

Questionnaire

A 12-item questionnaire was designed to inquire about students' perceptions on the use of an electronic analytic scale to receive feedback on their academic essays. We received responses from 14 out of 34 students.

The first question inquired about students' prior experience receiving feedback using an analytic scoring scale. Questions 2 and 3 checked students' understanding of the descriptors of the analytic scale. Questions 4 to 8 explored students' perceptions on the effectiveness of this tool to improve their writing skills. In question 9, students had to express whether they would like to continue receiving feedback using the scale. Question 10 invited students to state whether they would implement any changes to the scale in order to obtain highly informative feedback. In question 11, students were asked to indicate which tool, either the analytic scale or teachers' comments, was the most effective to provide them with feedback. Question 12 allowed students to add any information they considered necessary and it had not been addressed by any of the previous questions.

Close-ended questions were quickly processed by the Google form. Meanwhile, open-ended questions were analysed by each researcher separately and then the results of the analysis were compared and contrasted to minimize the subjectivity of any researcher affecting the results. Open-ended responses were analysed using content analysis, identifying salient categories in the data.

Text Analysis

Text-based studies aim at measuring or analysing certain features of the texts. Of the nine categories proposed by Polio (2001), this study only focused on *overall quality*, *linguistic accuracy*, *syntactic complexity*, *lexical features*, and *coherence and discourse features*.

Text assessment by means of holistic scales, analytic scales and ranking typically belongs to the realm of *overall quality*. The absence of errors, for example, mistakes at the level of word choice, spelling, or punctuation can be associated with *linguistic accuracy*. *Syntactic complexity* is defined as a combination of basic and complex structures that a language offers (Wolfe-Quintero, 1998). Lexical richness is measured in the category *lexical features*, which is defined by Laufer and Nation (1995) as quantifying ''the degree to which a writer is using a varied and large vocabulary" (p. 307). Measuring the overall coherence of a text by means of a scale, which provides a direct and objective way to measure the effectiveness of a given treatment, belongs to the category *coherence and discourse features*. These features can also be analysed by comparing texts written by native speakers and non-native speakers to understand how speakers of different languages organize their texts. Finally, *coherence and discourse features* can also be analysed by examining texts written by native speakers to draw conclusions in terms of how genres are written in a given field of study.

Based on these five categories, *overall quality*, *linguistic accuracy*, *syntactic complexity*, *lexical features*, and *coherence and discourse features,* an excel file was designed to process the data obtained from the text analysis. This file included code names for each student, the four categories included in the scale, and a space to include examples that show improvement in students' performance and a change in band within the rubric for each category. Given the qualitative nature of this study, examples play a key role to understand the impact of electronic assessment.

Results

The categories that arose from the analysis of the focus group discussion and questionnaire were the following: a) impact of the scale on students' writing skills; b) scale as a self-assessment tool and c) students' general perception of the scale.

Impact of the Scale on Students' Writing Skills

The focus group interview as well as the questionnaire provided information about the impact that the analytic scoring scale had on students' writing skills. Students stated that this was the first time they used an analytic scoring scale as an assessment tool.

In the focus group interview, students were asked if the scale provided them with necessary information to adopt strategies or improve their writing skills. All of them agreed that the scale provided a general understanding of what was expected from their essays and obtained a general overview of what needed to be improved. In terms of the acquisition of strategies, all the participants stated that the scale did not help them in those terms. In addition, all the participants highlighted the fact that the scale was useful only if it was combined with coded feedback. One participant pointed out that it could be useful as a post–writing tool because she used it only after receiving feedback and not as an instrument to help her when writing her draft. She added that it would be more useful if it were shorter as she only read the areas she needed to work on and not all the other bands.

Similarly, in the answers to the questionnaire, all participants agreed that the analytic scale helped them improve their writing skills. Even though none of them had received previous training in the use of the scoring scale, only two participants indicated that they would like to receive some training while 12 participants did not find such training necessary. When asked to assess the analytic scoring scale in terms of its usefulness, eight participants judged the tool to be effective and six stated that it was very useful. In addition, eight participants mentioned that the area in which the scale had helped them improve the most was *organization*, three of them mentioned *quality of expression* and *language accuracy*, two others referred to *content appropriateness*, and one student referred to *formal appropriateness and layout*. By contrast, when participants had to identify the areas for which the scale proved to be less useful, five of them made reference to *quality of expression* and *language accuracy*, three mentioned *content* *appropriateness*, two alluded to *content appropriateness* and two to *formal appropriateness and layout*.

Overall, the answers that the participating students provided in the focus group interview and the questionnaire evidence that the implementation of an analytic scoring scale has an impact on the development of students' writing skills, particularly if combined with teachers' comments. According to the results, the scale provides students with an overview of their performance.

Scale as a Self-Assessment Tool

Both the focus group and the questionnaire collected information about students` perceptions of using the analytic scale as a self-assessment tool and of its usefulness to clarify the assessment criteria of written tasks.

During the focus group session two participants stated that the scale provided them with overall feedback on their performance and pointed out that this type of assessment has a strong visual impact that clearly indicates the areas that need improvement. One student said that it helped her to get organized and identify the areas that needed improvement. At the same time, she stated that teachers' comments were more useful as she could detect mistakes faster.

Similar results were obtained from the questionnaire. Out of the 14 participants, 12 stated that they would like to receive feedback by means of an analytic scoring scale while two stated the opposite. These two students found that the scale offered general feedback on their performance while they felt the need for more specific types of feedback. As one student pointed out that "even when the scale assesses all students following the same criteria, some of us are still uncertain about the changes that need to be made to improve our work" (ST 13). Most students agreed that the scale offered information about general aspects of their writing. One student shared: "I believe that the scale has helped me gain a general understanding of my writing performance and check the aspects that I need to improve." (ST 1)

Overall, the participants believed that the scale could work as a complement to the indirect feedback received. All of them agreed that the scale was a good tool if it was used in tandem with teachers' comments, which offered them more specific information about the aspects that needed improvement.

Students' General Perception of the Scale

Tools that promote self-assessment need to be practical and simple to understand. Participants were asked if the analytic scoring scale was clear or if there were any concepts that needed further explanation. Most of the students established that the descriptors of the scale were simple to understand. One student mentioned that "each descriptor offers detailed information; thus, it is not difficult to understand the scale" (ST9). Only one participant stated that the descriptors were hard to follow. When inquired about which descriptor they found the hardest to understand, they signalled content appropriateness. According to the participants, this descriptor does not offer enough information to improve their work.

Participants were also asked if they felt the need to receive a training session in order to apply and use the analytic scoring scale. Most participants agreed that they would not need such a session. One of them stated that the scale was very clear and visually easy to follow. A second participant agreed with this statement, but added that the feedback offered by the scale should be supplemented by other types of feedback such as teachers' comments.

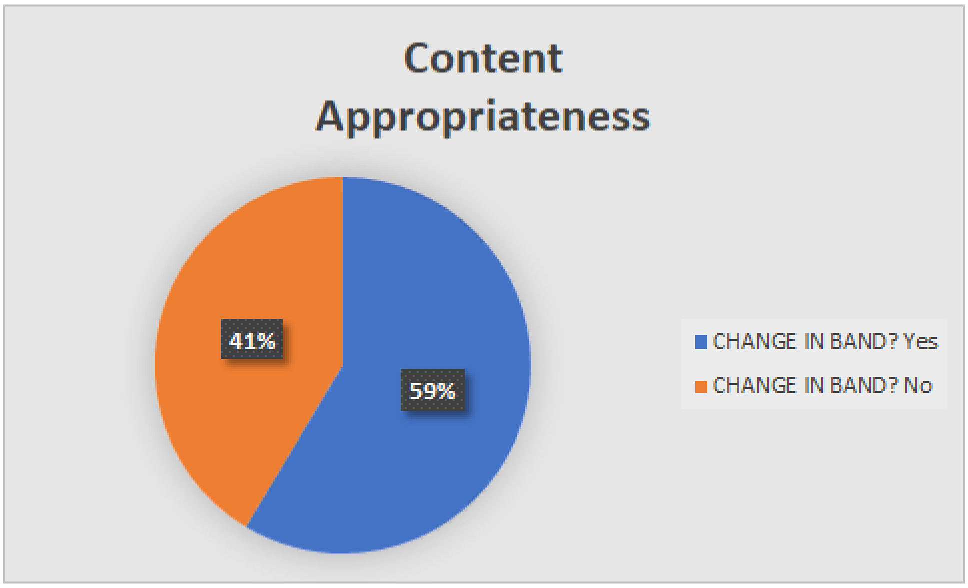
In conclusion, the analytic scoring scale was perceived as a useful self-assessment tool for students to be aware of the correction criteria of their work and to get a clear idea of the areas that needed improvement. Still, it is necessary to emphasize that to improve the effectiveness of the scale, teachers' comments should be part of the feedback that students receive.

Text-Based Analysis

The implementation of the analytic rubric together with the coded feedback yielded the following results after the text analysis was conducted on the first and second drafts of the essays. The number of students who submitted their first draft was 51 against 29 who submitted their second draft. The focus of this analysis will be on the 29 students who completed the whole writing process.

Figure 1

Content Appropriateness



As reflected in figure 1, 59% of students improved their compositions in the category of *content appropriateness*. The following is one example that shows an instance of improvement in this category. For the sake of brevity, examples from the other categories were not included.

Student 45

Draft 1: "First, taking up any activity for the sake of leisure enhances an individual's physical health. Even if it is not related to sports, it considerably prevents symptoms of low blood pressure. Such symptoms besides causing dizziness and the risk of fainting could also damage the heart and the brain. Furthermore, people with hobbies are less likely to harm their bodies. Seeing that it makes people active, it keeps them away from self-destructive recreation - such as drinking alcohol and taking drugs - caused by boredom."

Draft 2: "First, taking up any activity for the sake of leisure enhances an individual's physical health. Activities that include exercise like running or any other sports help the heart pump more blood through the human body preventing cardiovascular diseases. This is because frequent exercise increases muscle tone helping blood to circulate better; consequently, avoids clots to form inside the heart. Furthermore, physical activity reduces cholesterol in the blood by avoiding fats being deposited on blood vessels. Therefore, high levels of cholesterol are suppressed as well as the risk of developing symptoms of heart diseases, stroke, high blood pressure and diabetes."

Figure 2 shows that 55% of students improved their compositions in terms of *organization*, and figure 3 shows that 45% could improve their *quality of expression and language accuracy*.

Figure 2

Organization

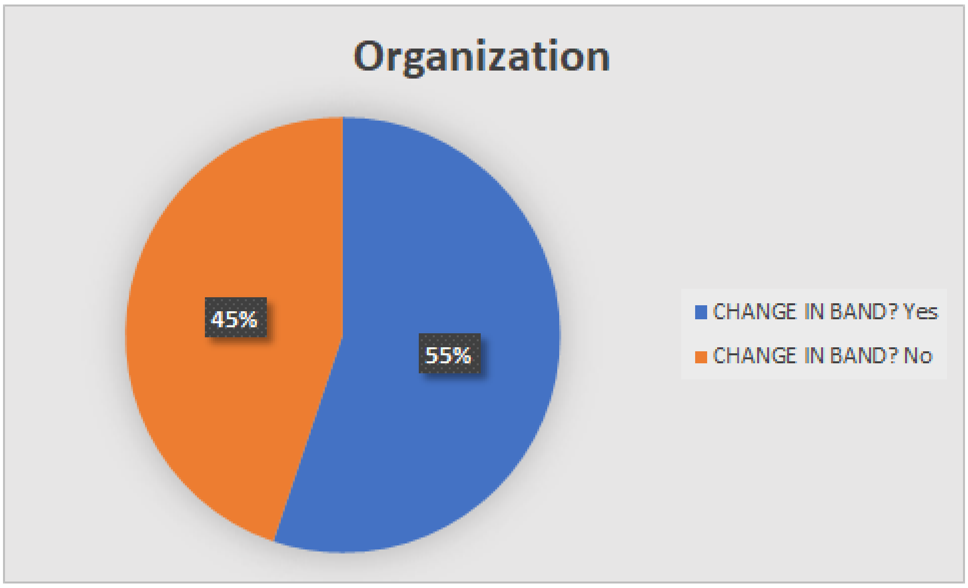


Figure 3

Quality of Expression and Language accuracy

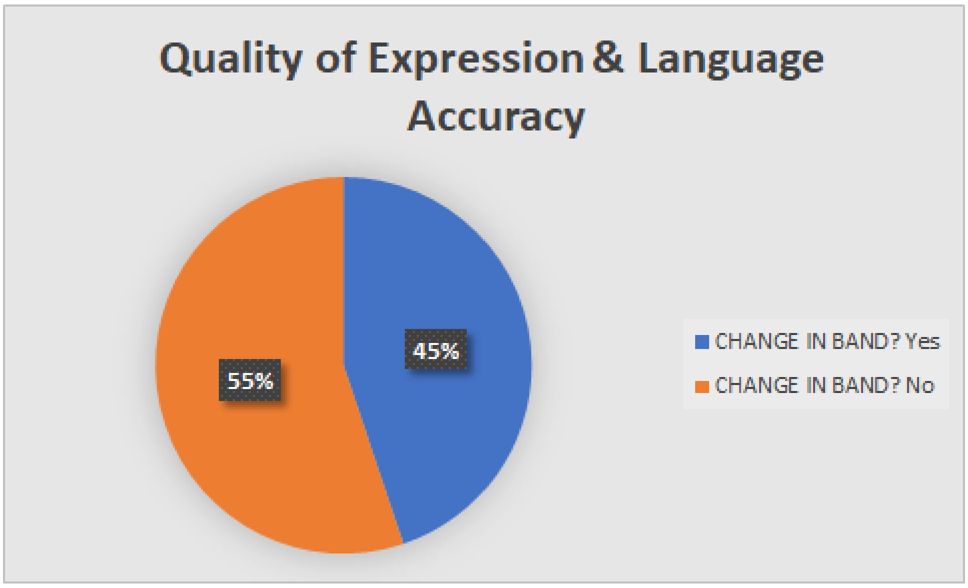
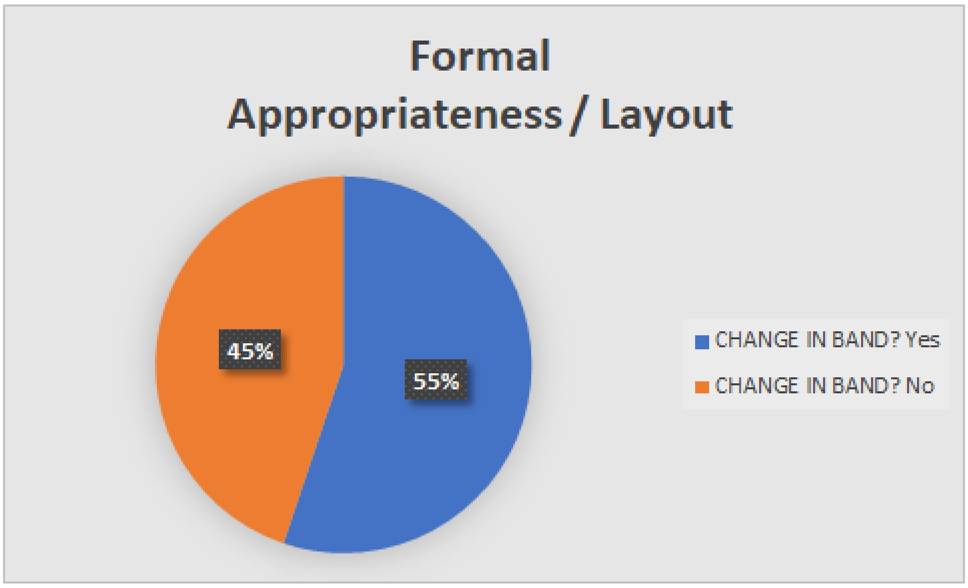


Figure 4 indicates that 55% of learners improve the *formal appropriateness and layout* of their compositions.

Figure 4

Formal Appropriateness/ Layout



Discussion and Conclusion

The results of the interview and questionnaire data reinforce Andrade and Du (2005), Bolton (2006), and Ahour and Mukundan´s (2009) findings since our students stated that the information provided by the analytic scoring scale made them aware of their general performance. At the same time, as information was organized by criteria and bands, students gained a better understanding of the evaluation process and they could identify their areas of improvement and strength. Students' conception of the scale as being a complement to teacher's comments was stressed by the participants when they explained that the information provided by this instrument made sense only after reading the comments and coded feedback provided by the instructor. Thus, this instrument seems to have served as a complementary assessment tool.

Most participants shared that the scale was clearly written and that they did not experience any difficulties trying to understand the bands. However, the participants believed that the scale did not help them develop strategies that could contribute to the development of their writing skills. By contrast, the instances of coded feedback and teacher's comments were highly valued and regarded as key to developing their writing skills.

In the questionnaire, students stated that the feedback provided by the scale helped them to improve their work. In their opinion, the areas of more salient improvements were organization (57,1%) and language (21,5%). Contrary to students' beliefs, the data obtained from the text analysis yielded that content was the area in which students showed more improvements (59%), followed by organization and layout respectively (55%). Students viewed content as the second least area of improvement whereas the text analysis showed that it was the area that almost 60% of students improved. While participants claimed that language development was the second most salient benefit of incorporating the rubric, the text analysis revealed that only 45% of students showed some improvement in this regard. Finally, the most noticeable difference was in layout since a few participants claimed that the rubric helped them improve in this regard while the text analysis revealed that 55% of students could make improvements.

Given the important role feedback may have in improving students' writing skills, it is necessary to understand students' perceptions of the feedback they receive in instances of formative assessment because this information is crucial to their language learning processes. Furthermore, analysing the students' perceptions and their written productions contributes to creating a much fuller picture of the impact of assessment in the learners' writing skills.

What seems to be a mismatch between students' perception of the benefits of the rubric in terms of the areas of improvement in their written production and the results that the text analysis yielded could be explained in terms of students' preferred choice: coded feedback and teacher comments. In the questionnaire, we asked students to refer to the benefits of the rubric, not coded feedback or teacher comments. The benefits stemming from coded feedback and teacher comments emerged as a result of the focus group interview. Thus, students' improvements in their second draft could be the consequence of implementing the rubric, coded feedback and teacher comments.

After conducting this investigation, we are still not certain whether the benefits of incorporating the rubric are known by the learners, or if its benefits are limited, as the participants stated. While the participants agreed that the rubric helped them to understand what was expected of them and to have a general picture of which areas they needed to improve, they were not aware of how the scale could help them develop their writing skills. However, understanding which areas one needs to improve and becoming familiar with teacher's expectations are essential strategies that can help learners develop their writing skills. Further studies need to be conducted to assess the benefits of the rubric by itself rather than in combination with the coded feedback and teacher comments.

References

Ahour, T., & Mukundan, J. (2009) Analytic assessment of writing: Diagnosing areas of strength and weakness in the writing of TESL undergraduate students. *Iranian Journal of Language Studies (IJLS), 3* (2), 195-208.

Andrade, H. & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation 10* (5), 1-11.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5-31. https://doi.org/10.1007/s11092-008-9068-5

Bolton, C.F. (2006). Rubrics and adult learners: Andragogy and assessment. *Assessment Update, 18* (3), 5-6.

Buckingham, L., & Aktuğ-Ekinci, D. (2017). Interpreting coded feedback on writing: Turkish EFL students' approaches to revision. *Journal of English for Academic Purposes*, *26*, 1–16. https://doi.org/10.1016/j.jeap.2017.01.001

Canavosio, A. A., Cad, A. C., & Salinas, J. (2021) Criterios de evaluación en escritura académica en inglés: Encuentros entre evaluadores y confiabilidad. *Academia peruana de la lengua, 69,* 256-295.

Collins, J., & O'Brien, N. (2003). *The Greenwood dictionary of education*. Greenwood Press.

DiGiovanni, E., & Nagaswami, G. (2001). Online peer review: An alternative to face-to-face? *ELT Journal, 55*(3), 263–272 https://doi.org/10.1093/elt/55.3.263

Dörnyei, Z. (2007). *Research methods in applied linguistics.* Oxford University Press.

Ferris, D. R., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing, 10*, 161-184. https://doi.org/10.1016/S1060-3743(01)00039-X

Hattingh, K., & van der Walt, J. (2013). The development and validation of a rating scale for ESL essay writing. *Journal for Language Teaching*, *47*(1). 73-105. https://doi.org/10.4314/jlt.v47i1.4

Hewett, B. (2000). Characteristics of interactive oral and computer-mediated peer group talk and its influence on revision. *Computers and Composition, 17*, 265–288. https://doi.org/10.1016/S8755-4615(00)00035-9

Hyland, K., & Hyland, F. (2006). Interpersonal aspects of response: Constructing and interpreting teacher written feedback. In K. Hyland & F. Hyland (eds.), 206–224.

Hyland, K. (2002). Activity and evaluation: Reporting practices in academic writing. In J. Flowerdew (Ed.), *Academic discourse* (pp. 115-130). Longman.

Hyland, K. (2003). *Second language writing*. Cambridge University Press.

Kern, R. (2000). *Literacy and Language Teaching*. Oxford University Press.

Kroll, B. (Ed.) (1990). *Second language writing: Research insights for the classroom.* Cambridge University Press.

Krueger, R (2002). Designing and conducting focus group interviews. Retrieved from https://www.eiu.edu/ihec/Krueger-FocusGroupInterviews.pdf

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*(3), 307–322 https://doi.org/10.1093/applin/16.3.307

Lee, I. (2008). Understanding teachers' written feedback practices in Hong Kong

secondary classrooms. *Journal of Second Language Writing*, *17*(2), 69-85.

Liu, Q., & Brown, D. (2015). Methodological synthesis of research on the effectiveness of corrective feedback in L2 writing. *Journal of Second Language Writing, 30*, 66-81. https://doi.org/10.1016/j.jslw.2015.08.011

O'Malley, J y Pierce, L. (1996). *Authentic Assessment for English Language Learners: Practical Approaches for Instructions.* Boston: Addison-Wesley Publishing Company.

Polio, C. (2001). Research methodology in second language writing: The case of text-based

studies. In T. Silva & P. K. Matsuda (Eds.), *On second language writing* (pp. 91-116). Erlbaum.

Reddy, Y. M., & Andrade, H. (2010) A review of rubric use in higher education. *Assessment & Evaluation in Higher Education, 35*(4), 435-448. https://doi.org/10.1080/02602930902862859

Sachs, R., & Polio, C. (2007). Learners' use of two types of written feedback on an L2 writing task. *Studies in Second Language Acquisition, 29*, 67-100. https://doi.org/10.1017/S0272263107070039

Storch, N., &Wigglesworth, G. (2010). Learners' processing, uptake, and retention of corrective feedback on writing: Case studies. *Studies in Second Language Acquisition, 32,* 303-334. https://doi.org/10.1017/S0272263109990532

Weigle, S.C. (2002). *Assessing writing*. Cambridge University Press.

Ware, P., & Warschauer, M. (2006). Electronic feedback and second language writing. In K. Hyland and F. Hyland (Eds.) Feedback and second language writing (105–122). Cambridge University Press. https://doi.org/10.1017/cbo9781139524742.008

Wiliam, D., & Thompson, M. (2008). Integrating assessment with learning: What will it take to make it work? In C.A. Dwyer (Ed.) The future of assessment: Shaping teaching and learning (53- 82). Routledge. https://doi.org/10.4324/9781315086545-3

Wolfe-Quintero, K. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity (Technical Report)*. National Foreign Language Resource Center.

Appendix

ANALYTIC SCORING SCALE

Criteria	Score	Descriptor
Content Appropriateness How well does the writing accomplish its purpose/ address the essay question?	35 -31	**Excellent/Very good:** Thorough and complete response to the essay question; highly focused on the issue; information is relevant and accurate and/or appropriate (no section within any of the paragraphs is inaccurate or off topic).
	30-25	**Good:** Complete response to the essay question; adequately focused on the issue; information is mostly relevant and accurate and/or appropriate (one section within any of the paragraphs may be inaccurate or off topic).
	24-21	**Fair:** Fairly complete response to the essay question (probably with some minor omissions); mostly focused on the issue, but some inaccurate, and/or inappropriate information (two sections within any of the paragraphs may be inaccurate or off topic).
	20-10	**Inadequate:** Quite incomplete response to the essay question (major omissions); slightly focused on the issue, but information is often irrelevant, inaccurate, and/or inappropriate (one out of two body paragraphs may be inaccurate or off topic).
	9-0	**Unacceptable:** Incomplete response to the essay question; totally out of focus on the issue; information is irrelevant, inaccurate, and/or inappropriate (more than one body paragraph may be inaccurate or off topic).
Organization How well does the organization of the text adapt to the type of essay and method of development required?	30-26	**Excellent/Very good:** Fully consistent with method of development/ organization of the particular type of essay required; topic and subtopics are appropriately organized and sequenced according to the thesis statement and topic sentences; relations among text sections are clear to the reader. Thesis statement and topic sentences are appropriate and accurately focused.
	25-21	**Good:** Generally consistent with method of development/ organization of the particular type of essay required; topic and subtopics are appropriately organized and sequenced according to the thesis statement and topic sentences; relations among text sections are mostly clear to the reader (a few transition signals may be absent or misused). Thesis statement and topic sentences are adequate.
	20-18	**Fair:** Mostly consistent with method of development/ organization of the particular type of essay required; relations among text sections are sometimes unclear (some transition signals may be absent or misused). Some topic sentences may be absent or incomplete and/or the thesis statement may display some problems in focus.
	17-9	**Inadequate:** Mostly inconsistent with method of development/ organization of the particular type of essay required (a different type of essay or an extremely short type of essay); relations among text sections are seldom clear. Some topic sentences may be absent or wrongly focused and/ or thesis statement may be incomplete or unsuitable.
	8-0	**Unacceptable:** Totally inconsistent with method of development/ organization of the particular type of essay required (the text is not an essay); relations among text sections are frequently unclear. Most topic sentences may be absent or wrongly focused and thesis statement is absent

Criteria	Score	Descriptor
Quality of expression & Language Accuracy How well does language use suit the context of communication?	25-22	**Excellent to very good:** Appropriate use of language for the context of communication. Great variety and specificity in the use of vocabulary. Accurate language/grammatical structures. No errors in spelling and punctuation.
	21-18	**Good:** Generally appropriate use of language for the context of communication. Some variety and specificity in the use of vocabulary. Mostly accurate language/grammatical structures. A few errors in spelling and punctuation.
	17-**15**	**Fair:** Barely appropriate use of language for the context of communication. Attempt at variety and specificity in the use of vocabulary. Some inaccurate grammatical structures. Some errors in spelling and punctuation.
	14-8	**Inadequate:** Inappropriate use of language for the context of communication. Little variety and some specificity in vocabulary. Several/Many inaccurate grammatical structures (some basic mistakes in grammar, e.g.: verb forms, agreement). Several errors in spelling or punctuation.
	7-0	**Unacceptable:** Totally inappropriate use of language for the context of communication. Lack of variety and specificity in the use of vocabulary. Too many inaccurate grammatical structures. Too many errors in spelling and punctuation.

Formal Appropriateness / Layout How well does the writing meet established norms for formatting, capitalization, neatness?	10-9	**Excellent to very good:** Appropriate use of English academic writing conventions.[1] Layout (title format, margins, spacing and indentation), and neatness fully meet established norms.
	8-7	**Good:** Generally appropriate use of academic English writing conventions. Layout and neatness may show some problems, but generally meet established norms.
	6	**Fair:** Acceptable use of academic English writing conventions. Layout and neatness partially meet established norms.
	5-3	**Inadequate:** Mostly inappropriate use of academic English writing conventions. Layout and neatness rarely meet established norms.
	2-0	**Unacceptable:** Totally inappropriate use of academic English writing conventions. Layout and neatness never meet established norms.

(Adapted from Tribble, 1996 in Kern, 2000, p. 287-289)

The Impact of Collaborative Test Design on a Higher Education EFL Program

Rosa Cerda Aguilera, Ingrid Flores Murgo and Gabriel Morales Morales

Institutional English Program, Universidad de O'Higgins

Assessment plays a crucial role in the language learning process since it provides educators with a method not only to monitor students' progress but also to group learners into levels at the beginning of a learning process. Considering the learners' context seems to be essential to place students into a suitable level, and in the case of most standardised instruments, this aspect has not been considered. In this vein, Giraldo (2019) has stated that language assessment is not an abstract process but needs purpose and context to function. Thus, context-bound topics and tasks are essential to improve the chances of obtaining accurate results.

Using placement tests to identify students' proficiency level comprehensively before assigning them to a particular course is essential to meet their language learning needs (Fulcher & Davidson, 2007). The majority of the existing placement tests for purchase are standardised, and although they fulfil their purpose, there is a lack of contextualised tests that assess students based on their background knowledge, as well as their previous learning processes. Not considering the learners' institutional context may have a negative impact on the results; consequently, the whole learning process might be unsuccessful for teachers and learners. An ill-placed learner might affect the whole learning experience, whether the level is easy or hard for them. Therefore, having access to a contextualised assessment instrument appears crucial to guarantee students and teachers the best conditions for classes to meet their objectives and needs.

The present study explores the process of creating a novel placement instrument for large-scale testing in a public university and its impact on teachers, students and authorities at the program and institution levels. The results indicate that the collaborative process of creating assessment tools has a direct impact upon teachers' beliefs about the goal of teaching and learning English, curriculum design and sequencing, and methodological approaches regarding ELT. Furthermore, from students' perspective, it is possible to observe a positive impact on task complexity perceptions and contextualised use of language.

Theoretical Framework

The theoretical basis for this work comprehends general assessment principles, and types of test materials and tasks, which are essential for test design experiences. Previous research in the design of placement instruments is presented as a means to explain the decisions made by the team in charge of this research work.

Assessment and Types of Tests

Assessment can play a series of different roles, and instruments take different forms and formats depending on the purpose they have. In this study, placement testing is addressed, which has the purpose of grouping students by matching them with the most suitable course for their needs (Green, 2013). These tests comprehend different skills (listening, reading, etc.), which are addressed at various levels of specificity (implicit, explicit comprehension), together with different levels of cognitive skills (applying, creating, etc.). In spite of the existence of highly valid and internationally recognised placement tests, authors like Giraldo (2019) and Fernández (2017) recommend the use of instruments aligned, as much as possible, with the institutional curriculum. The use of personalised (or contextualised) instruments provides educators with clear and straightforward evidence of students' performance, as they do not rely on scales that need to be re-interpreted in order to make placement decisions.

Regarding their format, most internationally recognized tests are adaptive, which means they provide learners with questions based on test-takers' success in previous questions. Opposite to adaptive, non-adaptive instruments provide the same proportion and type of questions to all takers, irrespective of their previous answers. Finally, regarding their length, placement tests are usually brief, as they are expected to provide results as soon as possible, thus most of them are built in online platforms. In some cases, the information collected through these instruments is complemented by the use of written or oral samples in case online placement test results are inaccurate or inconclusive (Fernández, 2017).

Based on the idea of personalising placement tests, language and tertiary education institutions around the world have worked in developing their own assessment instruments. A relevant experience on this regard is Papadima-Sophocleus' (2007), who mentions that in the context of a private college in Cyprus, it was necessary to carry out the design of placement instruments in collaboration with experts from areas such as information technology and mathematics, as well as to have a significant amount of time and monetary resources to carry out a project that proves itself to be sustainable and lasting. Similar to this experience, Ramírez (2020) claims that in the context of a public university in Colombia, the biggest challenge was to train teachers in assessment, since their Language Assessment Literacy (LAL) allows them to train and interpret the results of a test, but not develop their own. Moreover, Ramirez also highlights the importance of stakeholders' support for the design process, as placement instrument development is of paramount importance for the development of assessment policies within the institution, as these instruments provide the essential insight from which institutional decision-making processes occur.

Text Quality and Contextualization

An essential part of language assessment instruments is the texts that are selected and used for receptive skills tasks (Brown, 2010; Green, 2013). Authenticity, complexity and purpose are some of the key criteria to select texts. There has been considerable debate over what is authentic and what is not, and the overall agreement among researchers is that texts created for ELT purposes are non-authentic, while texts created for communicative purposes are authentic (Kumaravadivelu, 2006). A third type of material corresponds to semi-authentic texts, which were produced for communicative purposes, but have been adapted for teaching, learning and assessment (Ávila, 2020). These have a close relation with the concept of contextualization, as the closeness of these texts with real life communication allow for the construction of contextualised tasks, which have proved to be effective for enhancing students' motivation and competence development, as it has been observed in classroom research conducted by authors such as Dalla Costa, Spataro & Cad (2017), Ávila (2020), and Banegas et al (2020), among others.

Regarding purpose, Green (2013) claims that tasks should place students in real-life situations, and they must generate low levels of fatigue (tiredness) during and after the task realisation. Authors such as Hiebert (2012) and Williamson, Fitzgerald and Stenner (2013) claim that texts must achieve a level of complexity which is challenging yet addressable for students; this is achieved by measuring texts' features such as vocabulary frequency and diversity (type-token ratio), sentence complexity (t-units) and use of cohesive devices, among others. Together with reader and content characteristics, these factors must be considered to formulate tasks which suit the needs and abilities of students, reducing anxiety and achieving higher levels of competence measurement accuracy. Currently, these criteria are strongly considered for the formulation and sequence of materials in countries like the USA (Common Core Standards), and play a key role in reading comprehension research in Europe. Nevertheless, this topic and its implications for teaching, learning and assessment have not been explored in depth by Latin American researchers.

Study Background

Universidad de O'Higgins started its operations in March 2017, together with the Institutional English Program (IEP). IEP has as its primary objective to help students develop their language competencies in English for everyday, academic, and professional contexts. The English curriculum is organised into four levels of English for General Purposes (EGP) and two of English for Specific Purposes (ESP); all students at the University must complete four EGP and one ESP course to graduate.

As a means to place students in one of the levels of the programme, the tool that was traditionally used was the Oxford Placement Test (OPT), a web-based instrument focused on measuring the use of English and listening comprehension of individuals (Purpura, 2007). After one year of implementation, it was possible to identify as a drawback that information about the students' competencies was non-comprehensive, as it focused mainly on the use of English, defined as lexical, grammatical and pragmatic knowledge. This led to a considerable number of reconsideration requests from students, who felt they were not placed at the correct level. Based on this, the IEP staff developed a complementary oral evaluation that allowed reducing the number of requests. Nevertheless, the practicality of both instruments was affected by the high monetary costs of providing tests and examiners for the increasing number of students (currently 4,000 students approximately), exceeding the budget of the programme. Based on these issues, the Program decided to create a new placement instrument that could address those problems: The UOH Placement Test (UOHPT).

How does the New Test Work?

The test aims at measuring students' level of communicative competence, as addressed in Celce-Murcia et al, (1995), in their model of Communicative Competence. As a means to measure this accurately, the instrument comprehends three different aspects: Use of English, understood as grammatical and lexical understanding and use of the language; Reading Comprehension, defined as identifying general and specific information in written texts; and Listening Comprehension, which is similar to reading comprehension, but focused on oral texts. There are two language skills that were not considered for this instrument: speaking and writing, which, in order to be assessed thoroughly, required the collaboration of an examiner or the access to advanced natural language processing technology, and both required significant investment and specialised collaboration, not accessible at the moment.

In terms of the features mentioned, detailed in Table 1, some differences that need to be highlighted between the OPT and UOHPT are the prominence of language skills rather than language knowledge for the UOHPT. Moreover, while OPT deals with comprehension tasks through non-authentic texts, UOHPT considers semi-authentic texts. Finally, UOHPT is non-adaptive, which means that all takers must answer the same number of questions, distributed in a fixed number of questions per difficulty.

Table 1.

Contrast between OPT and UOHPT.

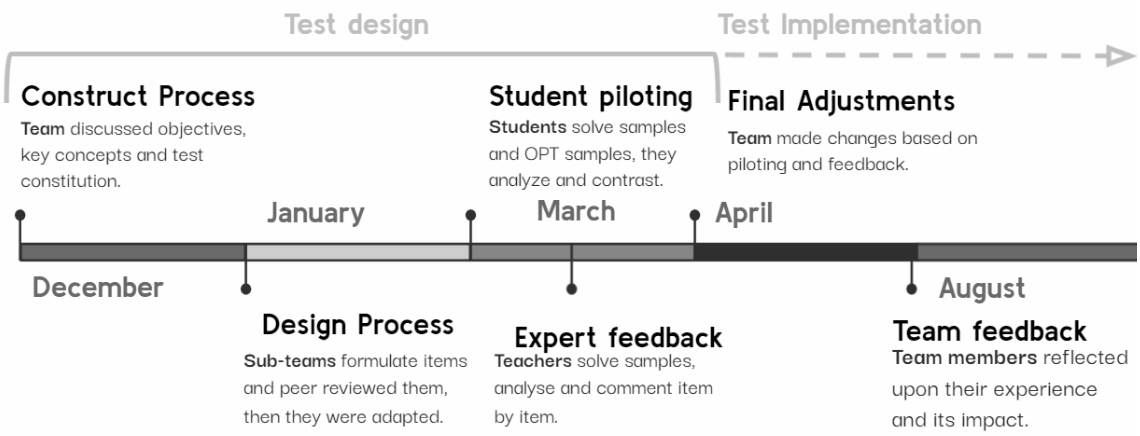
	UOH Placement Test (UOHPT)	Oxford Placement Test (OPT)
Components	Use of English Reading comprehension Listening comprehension	Use of English Reading comprehension
Type of test	Non-adaptive	Adaptive
Number of questions	8 Use of English 16 Listening 16 Reading	30 Use of English 15 Listening
CEFR level coverage	PreA1 to A2+	CEFR A1 to C2
Type of texts	Semi-authentic	Non-authentic

How was the Test Developed?

The design of the UOHPT was in charge of a team of 12 educators from different backgrounds, such as applied linguistics and international relations, among others. In order to cover four difficulty levels for the test, which would mirror the four levels in which students can be placed, the team was divided into four groups. As a means to engage in systematic work, the team met weekly for three and a half months, distributing tasks as can be observed in Figure 1. Each team was in charge of researching topics such as authenticity, text complexity, fatigue and language skills and subskills to train their peers and incorporate them in designing and analyzing sections and items.

Figure 1.

Instrument development timeline and key participants.



The first task carried out by the team was to complete a progression chart with key features to consider for texts and tasks for each type of item from their perspective. This distribution can be observed in Table 2. Secondly, teams developed a battery of items that fulfilled the characteristics previously established. Thirdly, teams cross-evaluated their peers' work in terms of linguistic and cognitive abilities. Fourthly, teams adjusted items, considering authenticity and the feedback received during the peer revision stage. Finally, each team selected representative items of their corresponding levels for the piloting process. After the piloting process, which considered both students' and teachers' feedback, major modifications were made before launching the definitive version of the test. Validation processes started after implementation and are still in progress.

Table 2.

Characteristics of items considered by teachers.

Type of item	Aspects to consider
Use of English	Lexical frequency
	Grammatical complexity
	Type of meaning
Reading and listening comprehension	Text type
	Length*
	Lexical density (Type-token ratio)
	Syntactic complexity (T-units per sentence)
	English variety**

*Measured in words for reading and second for listening comprehension.

** This element is exclusive for listening comprehension.

Research Questions

In view of the process previously described, the research questions that guided this study are the following:

1. What are ELT students' perceptions of the UOHPT in terms of strengths and weaknesses over OPT?
2. What are teachers' perceptions of the UOHPT and how does it relate to their understanding of cognitive and linguistic skills, as well as task difficulty?
3. How does the collaborative design of institutional instruments affect teachers' perceptions and classroom decision-making?

Methodology

Three different data collection instances were planned to obtain feedback about the instrument's components and quality, as well as the overall opinion of different actors of the institution. These instances allowed measuring the impact generated by the creation of UOHPT, and to identify relevant patterns that would allow answering the research questions:

Participants

For each instance, a different group of participants from the university community were selected, following specific criteria, which are listed below:

a) twelve students from the English Language Teaching (ELT) program from different cohorts (1 to 4); they were selected over students from other programs since they had been trained in teaching, English language and linguistics. Their feedback was expected to cover the test-taker perspective, enriched by their content-specific knowledge.

b) ten teachers from the IEP who did not participate in the test design process. It was expected that their feedback would shed light on the strengths and weaknesses of the instrument. Seven teachers who participated in the creation of the placement instrument. These individuals were asked to participate one semester after the implementation of the test. It was expected that their feedback would shed light on how the collaborative design experience affected their pedagogical decisions.

Procedure

Regarding the procedure for each of the three instances, each of them used different instruments and followed their own timeframes, represented in Figure 1. The detail for each design is the following:

a) group a) were tested using both, OPT and UOHPT within a time span of one month. After taking the tests, participants answered a 10-question survey (Appendix A) with two sections:

   I. perceptions on linguistic performance, test difficulty, fatigue, and test components.

   II. comparison between OHPT and UOHPT, focusing on what they measure, the difficulty they imply, and the fatigue they may produce.

b) group b) responded to different test forms, generated with 12 representative questions for each level from the question bank within a one-month time span. Based on this, they completed a matrix with question-by-question level of difficulty, aspect evaluated, and general comments on each item (Appendix B).

c) group c) answered a 20-question survey about the design process's impact on their pedagogical practices after one semester of the experience. The survey was organised into two sections (Appendix C):

   I. perceptions of professional impact after the collaborative design of a placement instrument

   II. general strengths and weaknesses of the design process

Quantitative and qualitative information obtained from these three instances was processed and analysed using Microsoft Excel and Atlas.ti software, respectively.

Results and Analysis

Data collected from the three feedback instances present different but complementary results, which aim in two directions: first, towards the test itself, and second, towards the context, students, and teachers.

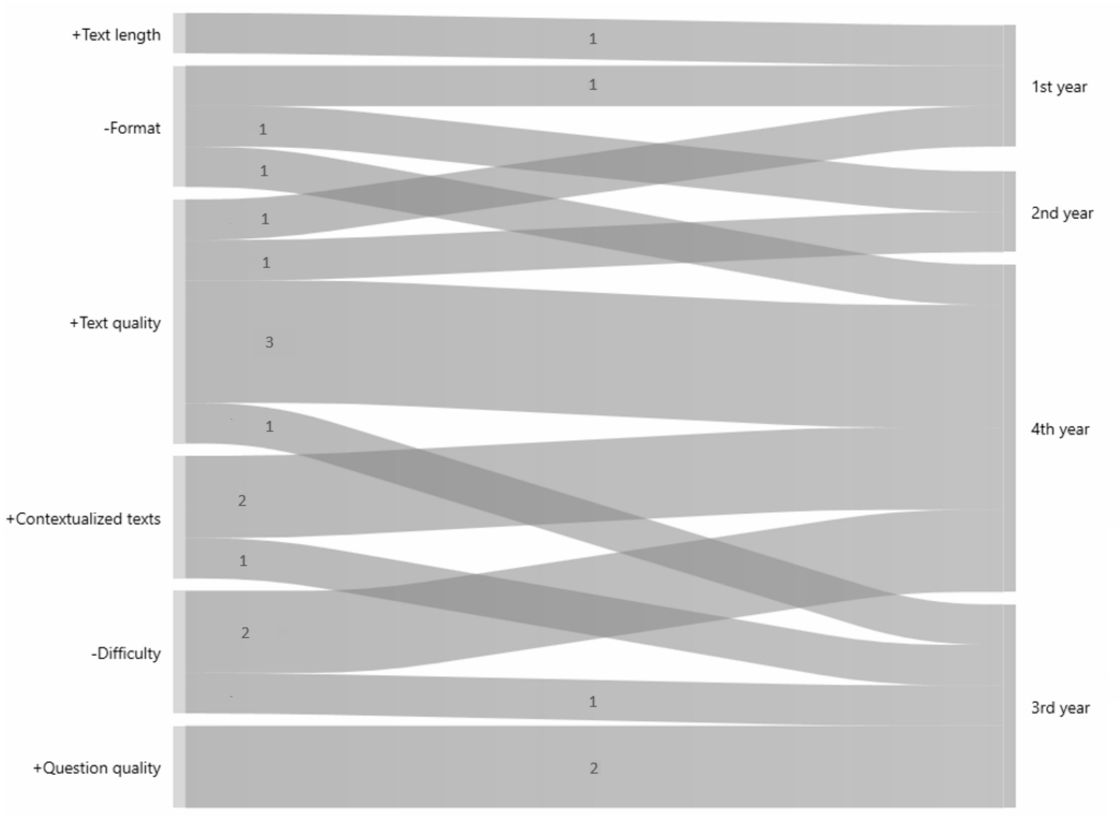
Student Feedback

Feedback obtained from students in the questionnaires for post-tests can be distributed into two different categories; the first is the overall strengths and weaknesses of the instrument compared to the former instrument, the second corresponds to the fatigue perceived by participants after taking each of these tests.

For the first category of results, strengths and challenges for the institutional placement test are represented in Figure 2, based on the comments made by students in the questionnaire.

Figure 2.

Test aspects detected by ELT students per cohort.



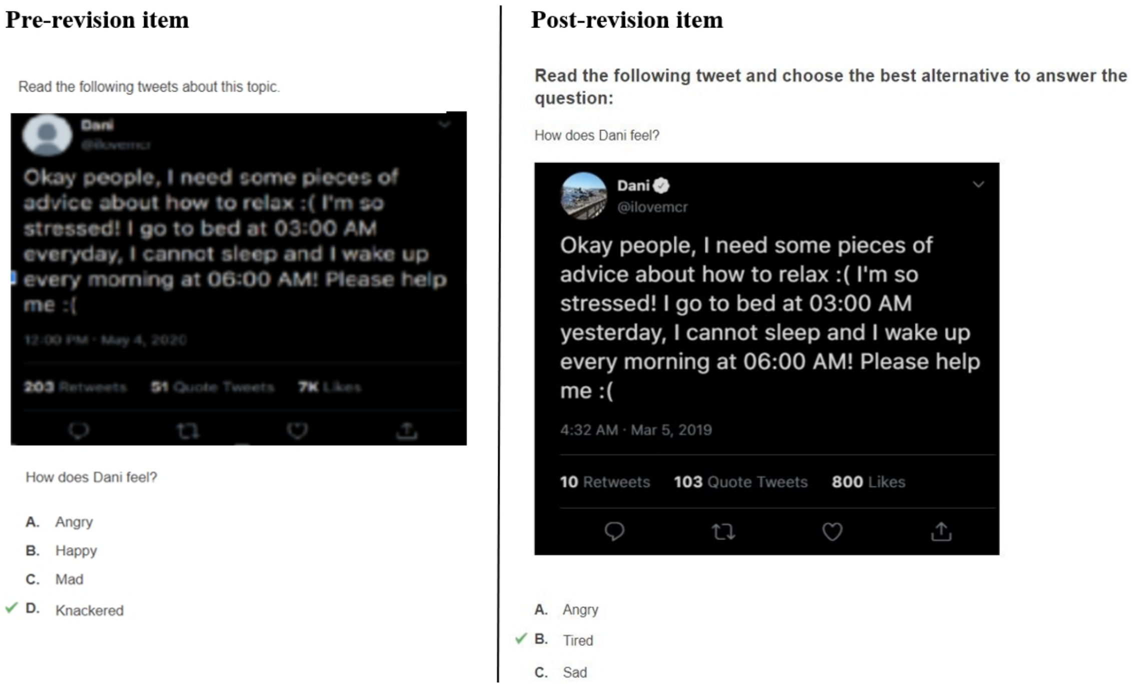
Note. Labels on the right correspond to students' cohorts, and labels on the left correspond to the major labels identified in comments. Lines indicate the presence of these labels per text, together with the number of occurrences.

On the side of strengths, the labels that had the highest frequency of occurrence were text quality, which was a general agreement among students from all cohorts. This is accompanied by high contextualization, observed by students from 3rd and 4th year, who believe this is an aspect that facilitates comprehension. According to students' opinions, texts (mainly internet posts and advertisements) are closer to students' everyday activities; thus, they are easier to understand. Other labels with minor occurrence were the quality of questions and instructions, as they were clear and accurate, as well as the length of texts, which was found adequate for the type of tasks and the nature of the instrument.

Regarding challenges, item format was a major aspect to revise according to all students; text layout needed to be corrected in order to be clearer for users to interact, as written texts were blurry due to format limitations, and some audios had excessive background noise. Regarding questions, some of them were ambiguous or had a format that would be difficult to address for students in an entry-level of competence, as it can be observed in Figure 3.

Figure 3.

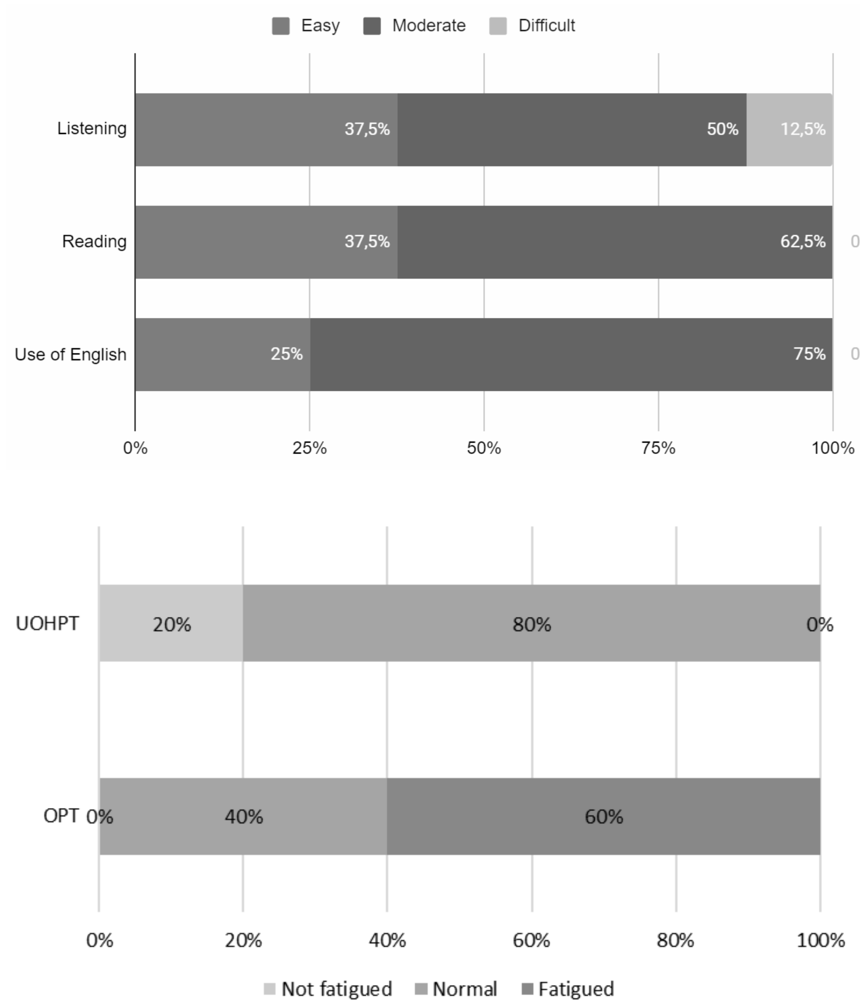
Pre-revision and post revision item sample.



In terms of test difficulty, the instrument was pointed out as ‘easy’ by students of 3rd and 4th years. One possible reason for this has to do with these students’ communicative competence, which goes beyond the planned levels for the placement and the programme. Regarding difficulty and fatigue perceptions, data indicates low levels of difficulty perceived by students who answered the UOHPT. As it can be observed in Figure 4, Reading and Listening items were found ‘easy’ in a similar distribution (37.5%). At the same time, there is a difference between the proportion of students who labelled these components as ‘moderate’ since 12,5% of the students labelled listening comprehension as ‘difficult’.

Figure 4.

Students' difficulty perception per test component and fatigue.



The high perception of the test as between 'easy' and 'moderate' is relatable to the fatigue measurement, which for the institutional instrument obtained an 80% 'normal' and a 20% 'not fatigued' perception measurements. These differ from the previous test (OPT), which obtained 60% as 'fatigued' and 40% as 'normal'. Based on this contrast, and the information previously analysed, it is possible to argue that UOHPT, due to features such as its high degree of contextualization, text quality and length, among other factors, constitutes an approachable instrument for students in this context, having low difficulty perception as well as low levels of fatigue.

Teacher Feedback

Feedback obtained from teachers who are part of the program and did not participate in the design process can be catalogued into two categories; first, the feedback about the test as well as teachers' beliefs about the TEFL process, and second, the correspondence among the test design features (linguistic skill, cognitive skill, and level). Regarding the first category, aspects to be improved were extracted from teachers' comments and represented in Figure 5.

Figure 5.

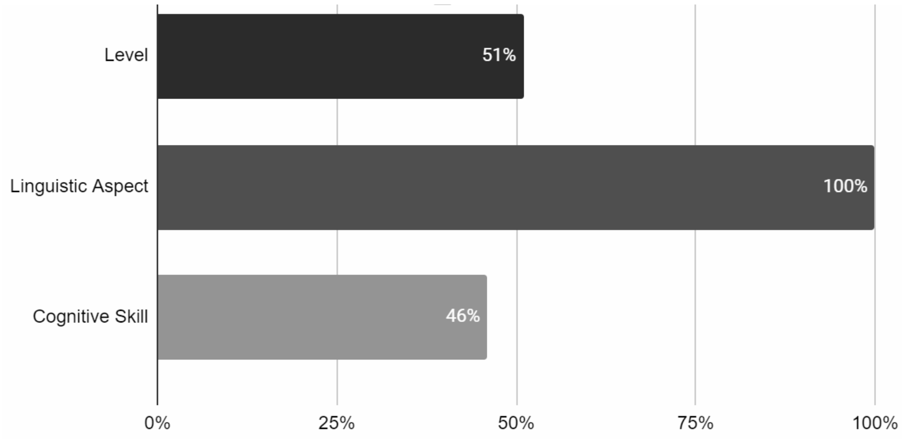
Teachers' observations and comments about the instrument.

Note. The occurrence of labels is organised increasingly; thus, bigger concepts had more repetitions than smaller ones.

The most significant test deficiencies detected by teachers corresponded to layout and format. It was necessary to unify and improve the quality questions and instructions to meet all students' needs and capacities, since they were formulated using obscure and/or complex language. Moreover, some of the answers were considered ambiguous or highly similar, as some questions appeared to have more than one possible answer. Finally, the quality of texts (both written and oral) required adaptation in terms of intelligibility. In turn, the labels detected with a minor number of occurrences corresponded to those of aim (cognitive skill), content and level, which are complementary with the information represented in Figure 6.

Figure 6.

Teacher accuracy at detecting test components' characteristics.



The overall observation that can be made about the placement instrument, when looking at the observations made by teachers, is that the linguistic aspect that is being measured in each item is clearer than the cognitive skill and difficulty level. From this perspective, it is possible to argue that the level of difficulty and the cognitive skill aim of exercises in the test needed to be revised so as to achieve a balance. Another possible explanation for this observation is that teachers in charge of this task showed difficulties at the time of detecting these aspects due to lack of theoretical knowledge on the matter.

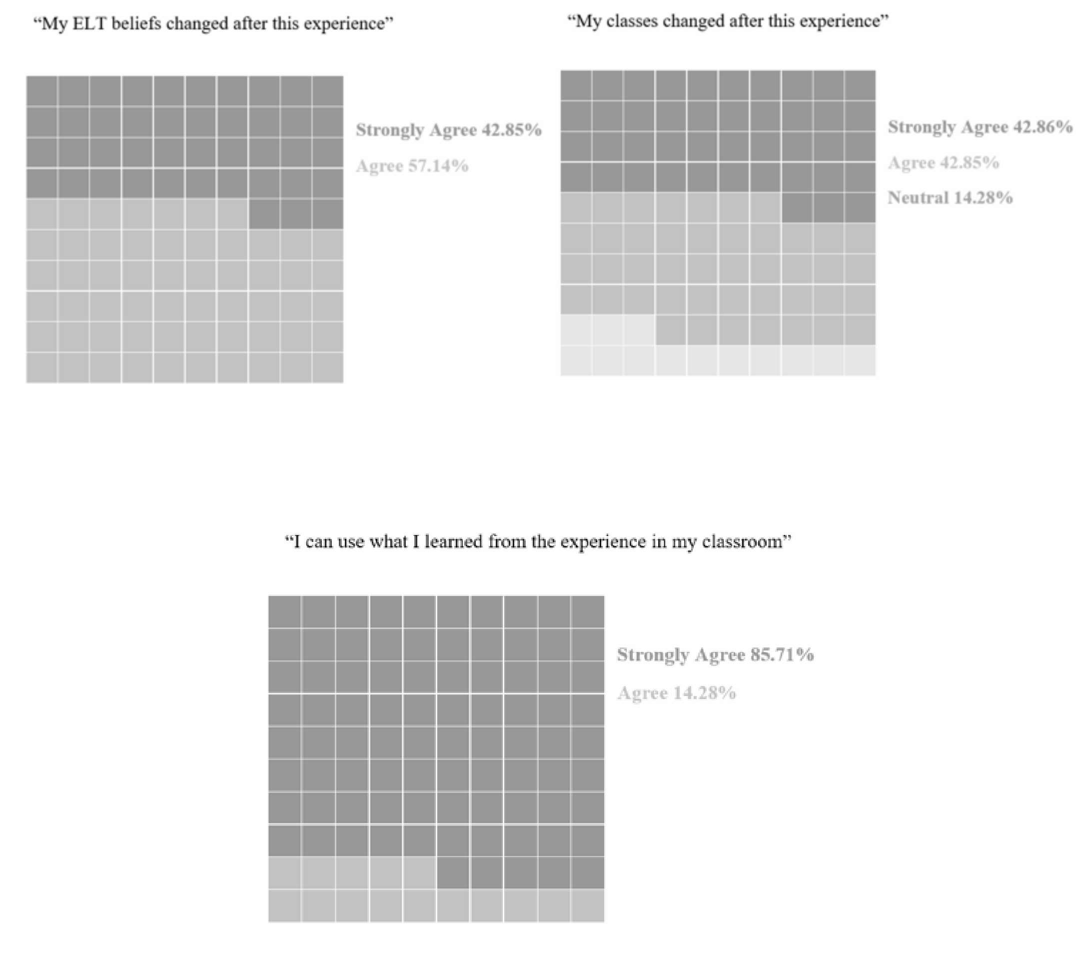
This data allowed for the revision of items before the test was released, making necessary changes based on teachers' feedback in order to suit students' needs. Moreover, it allowed identifying key concepts about language teaching, learning and assessment that needed a revision by the teachers in the program, thus leading to possible topics for teacher professional development, especially in topics concerning language assessment, as this is a major need of teachers, observed by López & Bernal (2009) and Herrera Mosquera & Macías (2015) and Giraldo (2018). Regarding the instrument's features and layout, major changes were made in item organisation, question structure, and accuracy of alternatives. Regarding texts, one of the major changes made was the adaptation of texts, which included the scripting and re-recording of authentic audios to avoid obscure samples.

Placement Team Feedback

The feedback obtained from the members of the team is not directly related to test features but to how the design process affected the beliefs and pedagogical practices of this group of teachers. As it can be observed in Figure 7, there is a major agreement regarding the impact of designing a placement test collaboratively, allowing teachers to revisit their beliefs about the teaching and learning processes.

Figure 7.

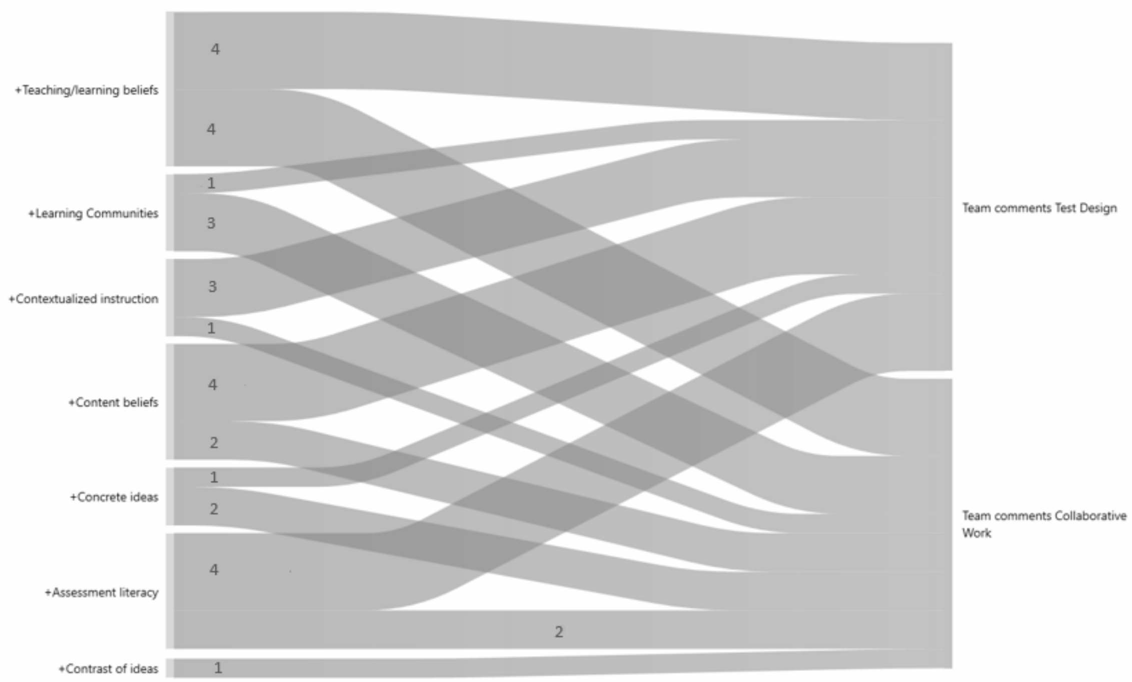
Perceived impact of the test design experience among team members.



The construct negotiation and formulation process allowed for the redefinition of some concepts that were conceived differently by each member of the team. The process of formulating exercises for the test itself allowed the team to transform the revisited constructs into exercises and tasks, which according to the participants, are potentially applicable in everyday classroom activities. At least 85.75% of participants have already implemented changes in their classes, with a mostly positive perception about student response, which is complementary with the information in Figure 8.

Figure 8.

Major aspects identified by team members for test design and collaborative work experiences.



As it can be observed in the figure, most of the beliefs that were revisited through the creation process were also supported by the collaborative nature of the design process. Some examples of this are teaching/learning beliefs that appear with the same weight for both, the evaluation of the creation process and the collaborative process. Similarly, the concept of learning communities appears for both instances, though in a different proportion. Finally, the only label that is exclusive for one of the instances is 'Exchange of ideas', which is exclusively related to the collaborative experience.

Conclusions

The previously reported results allow us to observe the high impact these instances of collaborative creation of assessment instruments have at all levels in an institution. Groups of both students and teachers actively engaged in the feedback process on the instrument, analysing it critically and providing their insight. This experience allowed them to identify the test as both closer to their reality and easier to administer and solve. Personalised, collective experiences of test creation seem to be, thus, an excellent alternative for achieving higher levels of placement accuracy and audience acceptance.

Moreover, apart from providing a more accurate measurement instrument for students, the work carried out by teachers generated space for continuous professional development. The collaborative revision and exchange of ideas and methodologies directly affected teachers' perceptions of their classroom practices and opened a space for critical reflection and innovation. This collaborative experience also allowed detecting significant differences and difficulties in a team of teachers, providing feedback instances with a two-fold purpose, since while the instrument was being reviewed, some flaws in terms of theoretical knowledge were detected. This insight allowed the team to engage in remedial actions, such as instances of training, and team organisation.

Some of the limitations for the implementation of this novel placement instrument were first the limited time to do research and design tasks due to teaching and administrative responsibilities. Second, not all the teachers who participated in the creative process managed similar knowledge about concepts like materials, methodology, etc., which required the team to devote extra time to address these topics.

There are essential validation processes that have to be carried out; in terms of the results obtained by the students in the test and their performance in the level they were assigned, test-retest instances and results across generations of students need to be analysed. The statistical analysis that must be carried out for these processes opens an opportunity to collaborate with other departments at the university. Furthermore, some alternatives to carry out further research could be to analyse whether specific elements, like social or school background, might influence their performance in the test. Moreover, the students who took the placement test will be asked to re-take it throughout the program, intending to analyse the extent and accuracy of the instrument from a longitudinal perspective.

While additional research should be done, there is no doubt that working collaboratively was the best way to conduct the creation of the test. Teachers agree that we can put into practice what we have learned in the creation process, which is important for us as professionals, as well as for the development of the Programme, since there is a solid team that after this process, are able to internalise and apply this information on assessment instances, class planning, and others. This process, as we were able to experience, brought us together and made us feel closer to the planning of the courses at a higher level, speaking more confidently when giving their opinions, and participating more actively when making decisions. We look forward to other instances to work collaboratively and make the Programme our own.

References

Ávila, J. I. (2020). *A self-study on the evolution of formative assessment practices.* [Undergraduate dissertation, Universidad Alberto Hurtado]. https://repositorio.uahurtado.cl/handle/11242/24917

Banegas, D. L., Roberts, G., Colucci, R., & Sarsa, B. A. (2020). Authenticity and motivation: a writing for publication experience. *ELT Journal*, *74*(1), 29-39.

Brown, H. D. (2010). *Language assessment: Principles and classroom practices*. Pearson Education.

Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied linguistics*, *6*(2), 5-35.

Dalla Costa, N. V., Spataro, C., & Cad, A. C. (2017). Designing authentic activities for a university course: An eclectic approach. *Authenticity in ELT*, 75.

Fernández, M. (2017). *Propuesta metodológica para la creación de un nuevo examen de inglés en las Pruebas de Acceso a la Universidad*. [Tesis Doctoral, Universidad de Granada]. Editorial de la Universidad de Granada. https://dialnet.unirioja.es/servlet/dctes?codigo=72377

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment. Advanced resource book*. Routledge.

Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile Issues in Teachers Professional Development*, *20*(1), 179-195.

Giraldo, F. (2019). Designing language assessments in context: Theoretical, technical, and institutional considerations. *How*, *26*(2), 123-143.

Green, A. (2013). *Exploring language assessment and testing: Language in action*. Routledge.

Herrera Mosquera, L., & Macías, V. D. F. (2015). A call for language assessment literacy in the education and development of teachers of English as a foreign language. *Colombian Applied Linguistics Journal*, *17*(2), 302-312.

Hiebert, E. H. (2012). The common core state standards and text complexity. *Teacher Librarian*, *39*(5), 13-19.

Kumaravadivelu, B. (2006). *Understanding language teaching: From method to postmethod*. Routledge.

López, A., & Bernal, R. (2009). Language testing in Colombia: A call for more teacher education and teacher training in language assessment. *Profile: Issues in Teachers' Professional Development, 11*(2), 55–70.

Papadima-Sophocleous, S. (2007). *Difficulties and constraints* involved *in developing a new English Placement Test online* [Conference Paper]. 10th Community Languages and English for Speakers of Other Languages Conference. Napier, New Zealand. http://ktisis.cut.ac.cy/handle/10488/3592

Purpura, J. (2007). The Oxford Placement Test: what does it measure and how? https://ataile.atauni.edu.tr/wp-content/uploads/2021/04/oxford-placement-test-the-meaning-of-OPT-scores.pdf

Ramírez, A. (2020). Challenges in the design and implementation of an English Placement Test for a Colombian Public University. *GIST – Education and Learning Research Journal, 21*, 191-208. https://doi.org/10.26817/16925777.825

Williamson, G. L., Fitzgerald, J., & Stenner, A. J. (2013). The Common Core State Standards' quantitative text complexity trajectory: Figuring out how much complexity is enough. *Educational Researcher*, *42*(2), 59-69.

Appendix A.

Student (group a) questionnaire

Encuesta Pilotaje UOHPT

Estimad@ estudiante,

Además de agradecer por tu participación en el proceso de pilotaje de la evaluación diagnóstica de Inglés del Programa de Inglés Institucional de la Universidad de O'Higgins, quisiéramos invitarte a cerrar el proceso contestando la siguiente encuesta. Los datos aportados por ti, nos ayudarán a calibrar el instrumento en función de tu mirada como docente en formación, por lo que agradecemos cualquier retroalimentación que puedas darnos con tal de mejorar el proceso de diagnóstico de tus compañer@s de la UOH.

¿Cómo catalogarías la dificultad del test?

- ◯ Muy Fácil
- ◯ Fácil
- ◯ Moderada
- ◯ Difícil
- ◯ Muy Difícil

¿Cómo etiquetarías los siguientes aspectos del test respecto de si dificultad?

Marca solo un óvalo por fila.

	Fácil	Moderado	Difícil
Listening	◯	◯	◯
Reading	◯	◯	◯
Use of English	◯	◯	◯

Contesta de acuerdo a tus percepciones

Marca solo un óvalo por fila.

	Bajo	Medio	Alto
¿Qué resultado creíste que obtendrías en el test antes de rendirlo?			
¿Qué resultado real obtuviste en el test?			
¿Cómo evaluarías tu nivel de inglés en términos generales?			

Respecto de la prueba diagnóstica ¿Qué tan de acuerdo estás con las siguientes afirmaciones? *

Marca solo un óvalo por fila.

	Totalmente de acuerdo	De acuerdo	Neutral	En desacuerdo	Totalmente en desacuerdo
El test presenta preguntas de distinta dificultad.					
El test evalúa comprensión lectora, comprensión auditiva y uso de inglés.					
El test tiene suficientes preguntas como para medir la competencia lingüística de un individuo.					
El tiempo entregado para contestarlo es proporcial a la cantidad de preguntas y su dificultad.					

Las instrucciones previas a rendir el test son claras y precisas.					
Las instrucciones de cada item son claras y precisas.					
Las preguntas de cada item no dan espacio a ambiguedades.					
Las alternativas de cada item están bien balanceadas.					
Los textos orales del test eran claros y adecuados para un test diagnóstico.					

¿Qué comentarios, sugerencias u observaciones te gustaría compartir respecto del test diagnóstico?

Oxford Placement Test y Test Diagnóstico UOH

Al haber rendido ambas evaluaciones, quisiéramos invitarte a reflexionar sobre tu experiencia en ambas, y contestar en función de los puntos en común y de separación entre ambas evaluaciones diagnósticas.

¿Cuáles de las siguientes características son propias de cada test y/o de ambos?

Selecciona todos los que correspondan.

	Oxford Placement Test	Test Diagnóstico UOH	Ninguno
Mide comprensión lectora	☐	☐	☐
Mide comprensión auditiva	☐	☐	☐
Mide gramática	☐	☐	☐
Mide vocabulario	☐	☐	☐
Mide aspectos pragmáticos	☐	☐	☐
Mide habilidades de nivel inferior (recordar, comprender, aplicar)	☐	☐	☐
Mide habilidades de nivel superior (analizar, evaluar, crear)	☐	☐	☐

¿Cómo evaluarías tu cansancio/desgaste luego de rendir ambos tests?

Marca solo un óvalo por fila.

	Para nada cansado/a	Normal	Cansado/a
Oxford Placement Test	◯	◯	◯
Test Diagnóstico UOH	◯	◯	◯

¿Cuáles son las fortalezas del Test Diagnóstico UOH frente al Oxford Placement Test?

____

____

____

____

¿Cuáles son las debilidades del Test Diagnóstico UOH frente al Oxford Placement Test?

____

____

____

____

Appendix B

Question analysis matrix (samples)

EJERCICIO	¿A qué nivel (1-4) corresponde la pregunta?				¿A qué componente (listening, reading, use of English) corresponde la pregunta?			¿Cómo la clasificarías de acuerdo a la Taxonomía de Bloom?	Sugerencias adicionales sobre el ejercicio
	1	2	3	4	Listening	Reading	Use of English		
Pregunta 1	☐	☐	☐	☑	☑	☐	☐	Analizar	Si lo pienso bien, creo que ambas son posibles por que un crucero es para eso, viajar por el mundo.
Pregunta 2	☑	☐	☐	☐	☐	☑	☐	Comprender	La pregunta después de la foto y no junto con el encabezado
Pregunta 3	☑	☐	☐	☐	☐	☐	☑	Aplicar	
Pregunta 4	☐	☑	☐	☐	☑	☐	☐	Recordar	adjetivos con ed o ing (nivel 3) quizás puede confundir un contenido que es más sobre "food" (nivel 2) creo que se debe decidir que aspecto de que nivel para poder posicionar a las personas
Pregunta 5	☐	☐	☑	☐	☐	☑	☐	Comprender	impaktada, no sabia que gorillaz no era real. Más allá de la construccion de la pregunta, el tópico es interesante
Pregunta 6	☑	☐	☐	☐	☐	☐	☑	Aplicar	
Pregunta 7	☐	☐	☐	☑	☑	☐	☐	Analizar	
Pregunta 8	☐	☑	☐	☐	☐	☑	☐	Analizar	
Pregunta 9	☐	☐	☑	☐	☐	☐	☑	Aplicar	
Pregunta 10	☐	☐	☑	☐	☑	☐	☐	Comprender	No se si importa pero me costo ponerla en algún nivel
Pregunta 11	☐	☐	☑	☐	☐	☑	☐	Analizar	El encabezado está redactado al reves
Pregunta 12	☐	☐	☐	☑	☐	☐	☑	Aplicar	

EJERCICIO	¿A qué nivel (1-4) corresponde la pregunta?				¿A qué componente (listening, reading, use of English) corresponde la pregunta?			¿Cómo la clasificarías de acuerdo a la Taxonomía de Bloom?	Sugerencias adicionales sobre el ejercicio
	1	2	3	4	Listening	Reading	Use of English		
Pregunta 1	☐	☑	☐	☐	☑	☐	☐	Comprender	
Pregunta 2	☑	☐	☐	☐	☐	☑	☐	Analizar	La instrucción podría ser un poco más clara, o que incluya ambos enunciados juntos.
Pregunta 3	☑	☐	☐	☐	☐	☐	☑	Aplicar	Quizá enunciar los días arriba, hay que abrir todas las pestañas para ver qué días hay
Pregunta 4	☐	☐	☑	☐	☑	☐	☐	Analizar	Cuesta un poco distinguir quién es quien en el audio
Pregunta 5	☐	☐	☐	☑	☐	☑	☐	Comprender	La pregunta podría ir más arriba, junto al enunciado, para saber qué buscar en el texto.
Pregunta 6	☐	☐	☑	☐	☐	☐	☑	Crear	
Pregunta 7	☐	☐	☑	☐	☑	☐	☐	Comprender	
Pregunta 8	☐	☑	☐	☐	☐	☑	☐	Comprender	
Pregunta 9	☐	☐	☑	☐	☐	☐	☑	Analizar	
Pregunta 10	☐	☐	☐	☑	☑	☐	☐	Evaluar	
Pregunta 11	☐	☐	☑	☐	☐	☑	☐	Comprender	
Pregunta 12	☐	☐	☑	☐	☐	☐	☑	Aplicar	

EJERCICIO	¿A qué nivel (1-4) corresponde la pregunta?				¿A qué componente (listening, reading, use of English) corresponde la pregunta?			¿Cómo la clasificarías de acuerdo a la Taxonomía de Bloom?	Sugerencias adicionales sobre el ejercicio
	1	2	3	4	Listening	Reading	Use of English		
Pregunta 1	☑	☐	☐	☐	☑	☐	☐	Comprender	
Pregunta 2	☐	☐	☑	☐	☐	☑	☐	Analizar	
Pregunta 3	☑	☐	☐	☐	☐	☐	☑	Recordar	
Pregunta 4	☐	☑	☐	☐	☑	☐	☐	Comprender	
Pregunta 5	☑	☐	☐	☐	☐	☑	☐	Comprender	
Pregunta 6	☐	☑	☐	☐	☐	☐	☑	Recordar	
Pregunta 7	☐	☐	☑	☐	☑	☐	☐	Analizar	Aunque yo le pondría "interpretar"
Pregunta 8	☐	☐	☐	☑	☐	☑	☐	Analizar	Aunque yo le pondría "interpretar"
Pregunta 9	☐	☐	☐	☑	☐	☐	☑	Recordar	
Pregunta 10	☐	☐	☐	☑	☑	☐	☐	Analizar	Pero sería realmente inferir. La pregunta lo dice.Las alternativas son confusas, ya que el audio se concentra más en las condiciones que en la temperatura, por lo que, tal vez, la respuesta correcta podría apuntar más al hecho que el tiempo va a estar cambiante.
Pregunta 11	☐	☐	☑	☑	☐	☑	☐	Analizar	Está muy buena la idea de la pregunta. Pero, tal, vez, se le podría cambiar el enunciado a escoger el lugar que él visitaría en vez de cuál sería el más apropiado. La mención a dibujar o tomar fotos en las opciones se podría, ya que no aparece nada referente a eso en la descripción del estudiante.
Pregunta 12	☐	☐	☑	☑	☐	☐	☑	Recordar	

Appendix C

Teacher (group c) questionnaire

Collaborative test design impact in a higher education EFL programme.

A continuación encontrarás una serie de enunciados , los que están relacionados con tu proceso de participación en la creación de nuestra herramienta de diagnóstico institucional de inglés. Esta encuesta es totalmente anónima, y está diseñada para que puedas compartir tus percepciones y evaluación de la experiencia.

Área de desarrollo personal/profesional en base a la experiencia de diseño.

En esta sección debes reflexionar sobre la experiencia de creación de un instrumento diagnóstico, tomando en cuenta las tareas que se hicieron y los aprendizajes tanto disciplinares como pedagógicos que puedas haber desarrollado.

¿Qué tan de acuerdo estás con los siguientes enunciados? *

Marca solo un óvalo por fila.

	Totalmente de acuerdo	De acuerdo	Neutral	En desacuerdo	Totalmente en desacuerdo
Siento que puedo aplicar en mis clases lo aprendido durante el proceso					
Lo que aprendí no es posible de ser aplicado en mis clases					
Mi participación me llevó a reflexionar en relación a mis prácticas de enseñanza del inglés					
Mi participación me llevó a reflexionar en relación a mis prácticas de evaluación del inglés					
La metodología que utilizo constantemente en mis clases tuvo algún cambio					

Mis creencias metodológicas cambiaron al ser partícipe de la creación del instrumento	⬭	⬭	⬭	⬭	⬭
Pasé de una metodología principalmente basada en gramática a una enfocada en el propósito comunicativo	⬭	⬭	⬭	⬭	⬭
Ninguna de las experiencias compartidas me resultó novedosa	⬭	⬭	⬭	⬭	⬭

¿De qué otra forma la experiencia de diseñar un instrumento diagnóstico impactó tus creencias y prácticas docentes? *

Área de desarrollo personal/profesional en base a la experiencia de diseño colaborativo.	En esta sección debes reflexionar sobre la experiencia de creación de un instrumento de manera colaborativa, tomando en cuenta las tareas que se hicieron y los aprendizajes tanto disciplinares como pedagógicos que puedas haber desarrollado.

¿Qué tan de acuerdo estás con los siguientes enunciados? *

Marca solo un óvalo por fila.

	Totalmente de acuerdo	De acuerdo	Neutral	En desacuerdo	Totalmente en desacuerdo
Personalmente el trabajo en equipo no es de mi agrado	⬭	⬭	⬭	⬭	⬭
Me sentí cómodo o cómoda trabajando en equipo	⬭	⬭	⬭	⬭	⬭
Mi foco metodológico fue influenciado por la opinión de las y los miembros del equipo	⬭	⬭	⬭	⬭	⬭
Mi foco evaluativo fue influenciado por la opinión de las y los miembros del equipo	⬭	⬭	⬭	⬭	⬭
El trabajo en equipo me ayudó a analizar distintos puntos de vista sobre la enseñanza de lenguas, que antes no había considerado.	⬭	⬭	⬭	⬭	⬭

¿De qué otra forma la experiencia de diseñar un instrumento colaborativamente impactó tus creencias y prácticas docentes? *

Área de desarrollo general

¿Qué tan de acuerdo estás con los siguientes enunciados? *

Marca solo un óvalo por fila.

	Totalmente de acuerdo	De acuerdo	Neutral	En desacuerdo	Totalmente en desacuerdo
La metodología con la que se trabajó en la creación del instrumento fue la adecuada	⬭	⬭	⬭	⬭	⬭
El tiempo destinado a la creación del instrumento fue el adecuado	⬭	⬭	⬭	⬭	⬭
Los temas que se abordaron y analizaron (fatiga, estrés, etc) fueron relevantes para el desarrollo de las tareas asignadas	⬭	⬭	⬭	⬭	⬭
Las instancias de feedback entre equipos me resultaron útiles	⬭	⬭	⬭	⬭	⬭

Si quisieras compartir algún comentario que no fue abordado en los enunciados anteriores, por favor, exprésalo aquí.

The Role of Test Anxiety on Performance on Standardized Language Tests

Josefina C. Santana

Universidad Panamericana, Guadalajara, Mexico

English has become the most widely used language in the world of international business and science. For this reason, universities in different countries have established English language proficiency as a graduation requirement. This is the case of most Mexican universities, especially those that belong to the private sector. English is generally accredited by an exam, and the required score depends on various factors such as university, city, field of studies, or others.

The present study was carried out in a private university in the city of Guadalajara, Mexico. For over 20 years, this university has required proof of English proficiency for graduation. Students must take any of the international standardized exams (TOEFL, IELTS, TOEIC, BULATS, CAE, etc.) that are available on the market. If they do not obtain the required score, they cannot receive their college degree. Most of the students manage to obtain the required score on their first attempt; thus, they accredit the language and can graduate without problems. However, some students appear to have difficulties obtaining the required score. This may be due to a variety of reasons; they do not know the exam and do not prepare properly, or they have stopped studying English for a while and have lost practice in the language, among other reasons. However, there is also the case of students who demonstrate that they are capable of using the language for their classes or their work, but they do not manage to obtain the required points in the standardised test, even after several attempts.

The present study posited test anxiety as a possible reason for failure. Although students can take any standardised test to accredit language proficiency, this study focused on the *Test of English for International Communication* (TOEIC), as it is the most widely used among this population. According to their field of study, students must obtain between

750 and 795 points in the exam to be able to accredit the language, and thus graduate. The objectives of the study were firstly, to determine how many of the students taking a TOEIC test at the university as a graduation requirement showed signs of test anxiety. The second objective was to evaluate the impact of test anxiety on the TOEIC scores of these students.

The following sections will describe the theoretical framework, studies into test anxiety and language exams, interventions to mitigate anxiety, the study itself, including the instruments and the procedures, the data analysis, the discussion of the findings, and finally, some pedagogical implications.

Theoretical Framework

Test Anxiety

Furlan et al. (2014) define test anxiety as "an emotional reaction to situations in which aptitudes are evaluated, its central characteristic being concern about possible failure or poor performance on tasks" (p. 32). Test Anxiety (TA) can be divided into two components: worry and emotionality (Liebert & Morris, 1967). Worry is a cognitive dimension of anxiety, and it manifests itself through negative thoughts. Those who suffer from cognitive anxiety on a test doubt their knowledge and abilities, compare themselves with others, and believe that they will fail the test. Avoidance attitudes, such as procrastination, are also cognitive signs of TA. Emotionality, on the other hand, has physiological manifestations, such as nervousness, headache or stomach ache, nausea, sweaty hands or increased heart rate (Cassady & Finch, 2014; Furlan, 2013).

Some test takers may suffer a moderate form of anxiety. This can be positive, as it leads the student to take the test seriously, to prepare for it, and to put into practice strategies to solve the exam. However, too much anxiety is detrimental, as it generates frustration in the learner, and can affect cognitive performance (Eccius-Wellmann et al., 2017; Furlan et al., 2014).

Test anxiety has been extensively studied since the 1950s, and it has been calculated that between 15 and 22% of students suffer some form of TA (von der Embse et al., 2018). Any number of factors may cause test anxiety

(TA). These include poor self-image, high expectations, parental pressure, individual ability, and poor study habits. TA is found at all educational levels, although it tends to increase the further the student advances, so that it is more common in university students than in elementary school children (Hurley & Padró, 2006). Among university students, TA affects older students more than younger ones (Álvarez et al., 2012), and those who study humanities more than students in other fields (Grandis, 2009). Putwain (2007) found that the sex, ethnicity, and socioeconomic status of the student are predictors of variance in test anxiety. Thus, it is less common among men and among people of a higher socio-economic level. Other studies (Cassady & Johnson, 2002; Núñez-Peña et al., 2016) have also found greater TA among women than among men. They attribute it to the self-expectations women tend to have regarding academic performance.

TA also appears to be affected by the cultural context. Though test anxiety can be observed in all cultures, some tend to experience it more frequently than others. It is more common in the Middle East and in Latin America, for example, than it is in Western Europe (Putwain, 2007). Cross-cultural studies of test anxiety have shown that test anxiety differs in structure and in its manifestations (Furlan et al., 2009).

Language Proficiency Tests and Anxiety

Standardised language proficiency tests have three characteristics that can increase the level of anxiety experienced by the test-taker. These are that the exam is in another language (Horwitz, 2010), that it is a high-stakes exam (Sommer & Arendasy, 2014; von der Embse et al., 2018), and, finally, that the exam has a strict time limit (Onwuegbuzie & Seaman, 1995).

Rumsey et al. (2016) carried out a study among people migrating to Australia who must take an IELTS exam. Test anxiety came about from different factors related to the test -changes in immigration rules, the cost of the exam, especially if they needed to pay it more than once, the score requirements, and the feeling that it was not appropriate to their needs- as well as the test content itself.

A study into anxiety and language proficiency tests as a graduation requirement for university students in Taiwan (Wu & Lee, 2017) found cognitive anxiety had a greater impact than emotional anxiety on test scores. The authors also indicate that the perceived importance of the test was a strong generator of TA. This was true even when students held a positive view of the test and the requirement.

For their part, Santana & Eccius-Wellmann (2018) studied anxiety before the TOEIC test among university students in Mexico. They found that women obtained significantly higher averages on an anxiety questionnaire and significantly lower scores on the TOEIC test. Additionally, female students showed higher cognitive anxiety, whereas men scored higher on items measuring physiological anxiety. The authors conclude that test anxiety significantly affects women in this population. Another Mexican study of test anxiety among university students (Santana & Ventura Michel, 2020) found that the participants did not prepare thoroughly for the exam. The majority said they had studied English during secondary or high school, but their courses had focused on communication, and not on grammar. Thus, they found the grammar section of the test especially stressful.

Test Anxiety Reduction Strategies

Because testing is a normal part of education, and because TA is so widespread, several studies have looked into interventions to reduce anxiety, with differing results. A Chilean study (Contreras-Soto et al., 2019) explored the use of portfolio writing activities to reduce test anxiety among sixth-graders. The authors found that this intervention was successful. In contrast, the intervention carried out by Myers et al. (2021), who tried expressive writing and stereotype threat mitigation training to reduce TA, was not as successful. They found that their interventions had no effect on anxiety nor on test performance.

On the other hand, training students to control their emotions appears to be effective (Liu et al., 2021; McDonald & Boud, 2003; Serrano Pintado & Escolar Llamazares, 2014), as does training in the use of metacognitive strategies (Xu & Huang, 2018), which allows students to plan, monitor, and evaluate their learning.

Materials and Methods

Context of the Study

The study took place at a private university in western Mexico. The institution is small, with a population of between four and five thousand undergraduate students. In 1998, the university took the decision to establish English language proficiency as a graduation requirement because, in those days, knowing English offered students a competitive advantage and meant access to better jobs, better pay, and greater opportunities to study abroad. Proficiency was to be demonstrated by means of the TOEFL test, one of the few available options in Mexico at that time, and the requirement was set at 490 points. Other exams were later added, and the score was gradually increased until now it is between 550 and 580 on the TOEFL, depending on the field of studies, with its equivalent on other tests.

Many of the students have access to bilingual schools or study abroad programs. They take an English diagnostic test upon entry to the university, and more than 80% of entering students place at a B2 level or above, according to the Common European Framework of Reference parameters (Santana et al., 2017). Thus, the score requirements are considered attainable.

Participants

The participants were 138 students who took a TOEIC exam in order to accredit their knowledge of English, and thus graduate. The respondents were 53% women and 47% men; the majority studied business (60.2%), while 18.8% studied engineering and 21% were in the area of humanities, which includes the Schools of Law, Communication, and Education. Students tend to take the exam in the last years of their studies, or even after they have finished, so the average age of the participants was 23 years old.

Instruments

The study measured the correlation between the TOEIC test score and an anxiety questionnaire, detailed below.

CAFEU

Before taking the TOEIC test, participants responded to the *Questionnaire of Test Anxiety among University Students* (CAFEU, for its initials in Spanish) developed by Grandis (2009). The questionnaire consists of 34 items that include questions about cognitive, avoidance, and physiological anxiety. It is answered on a five-point Likert scale, where 1 corresponds to *never* and 5 corresponds to *always*. The score will range from 34 (no anxiety) to 140 (high anxiety). For this study, the CAFEU showed a Cronbach's alpha of .938, which indicates good reliability. Although there are other anxiety scales, this one was selected because it was developed specifically for university students and in Latin America. The questions were in Spanish, as in the original scale (See Appendix for the questions in English).

TOEIC

The *Test of English for International Communication,* or TOEIC, is a standardized test to measure proficiency in the English language. It was developed by Educational Testing Services (ETS) in 1979 and is currently applied in some 160 countries around the world (Powers & Schmidgall, 2018). The version administered for this study consists of two parts; one measures listening comprehension and the second measures reading comprehension, and includes grammar and vocabulary items. Both sections use multiple-choice items and consist of 100 questions each. In its entirety, two hours and 30 minutes are required for the administration of the test. The exam score will be in the range of 10 to 990 points.

Data Collection

Students who had signed up to take a TOEIC test at the university were given the CAFEU questionnaire. Information was added to the questionnaire explaining the purpose of the study and how the data would be treated. They were told that if they did not wish to participate, they should return the questionnaire without responding to it. They were told that by responding to the questionnaire, they were giving consent to use their data for the purposes of the study. It was also explained that responding to the questionnaire or not would not affect their scores. Nobody refused to participate, but 67 questionnaires were returned with incomplete answers, so they were not included in the analysis. The questionnaire asked for participants' school registration number and their field of studies. The registration number was collected in order to match the questionnaire to the exam results. The TOEIC was administered under the conditions of timing, spacing, and security established by ETS to ensure the reliability and validity of the exam.

Data Analysis

A statistical comparison was carried out by field of studies and by sex to determine which subgroups present greater anxiety. Likewise, a multiple linear regression was performed to determine the correlation between anxiety as measured by the CAFEU, and the test results.

Results

Responses to the CAFEU questionnaire showed little evidence of test anxiety among the study population. The mean was 62.31 points out of a possible 140, which indicates moderate anxiety. Engineering majors report the highest anxiety (64.26 points), followed by humanities, with 63.81 points, and Business majors, with 61.15 points. A t-test of equal means shows that these differences are not statistically significant ($p=.948$), so it can be considered that the level of anxiety among this population does not depend on their area of study.

Regarding the results of the TOEIC test, the mean was 749.09 (lower than what is required by this particular university). The engineering students obtained the highest scores; their average was 804.80. Among Business students, the average was 750.06, and in humanities, the average was 696.37.

Business majors need the highest scores (780 or 795, depending on the area of study), so engineers are the only ones that, on average, pass the test. We see that engineering students report higher anxiety but perform better on the exam.

When comparing by sex, it is seen that women report greater anxiety than men (66.89 points vs. 56.35). In contrast, the TOEIC mean is lower for women than for men (722.24 points vs. 784), as seen in Table 1.

Table 1

Mean scores CAFEU and TOEIC for male and female students

	Gender	N	Mean	Standard deviation
CAFEU	M	60	56.35	14.737
	F	78	66.89	22.718
TOEIC	M	60	784	148.566
	F	78	722.24	144.666

T-tests were carried out to determine if the differences between scores for male and female students were statistically significant. Table 2 shows the t-test for CAFEU.

Table 2

Comparison of CAFEU means by sex of the participants

Test of independent simples

		Levene Test for equality of variances		T test for equal means		
		F	Sig.	T	Gl	Sig. (Bilaterial)
Score	equal variances are assumed	10.534	.001	3.426	136	.001
	equal variances are not assumed			3.514	122.87	.001

The difference between CAFEU scores for male and female students is statistically significant (.001), indicating that women report significantly more TA than men do. In contrast, the difference in TOEIC means between men and women is not significant (p=.807), as seen in Table 3.

Table 3

Comparison of averages on the TOEIC by sex

Test of independent samples

		Levene Test for equality of variances		T test for equal means		
		F	Sig.	T	Gl	Sig. (Bilaterial)
Score	equal variances are assumed	0.60	.807	-1.727	136	.086
	equal variances are not assumed			-1.725	133.229	.087

These results indicate that sex does not have a significant impact on TOEIC scores. Thus, a linear regression analysis test was carried out to determine if the TOEIC score is dependent on test anxiety as measured by the CAFEU. Table 4 shows the relationship between these two variables.

Table 4

T-test to measure the correlation between TOEIC scores with respect to CAFEU

		Non-Standardized coefficients			
Model		B	Standard error	t	Sig.
1	(Constant)	851.324	40.243	21.154	.000
	CAFEU	-1.641	.614	-2.670	.008

Test anxiety has a negative impact on the TOEIC score, as demonstrated by the negative coefficient. The p-value (.000) indicates that the correlation between TOEIC and CAFEU is significant.

The linear regression line is TOEIC = 851.324 – 1.641CAFEU. This indicates that an increase of 10 points on the CAFEU reduces the TOEIC score by 16 points.

Discussion

In general, it can be said that the population under study does not show great anxiety before the TOEIC test, even though for them it represents the difference between obtaining their degree or not. This contradicts the findings of Hurley and Padró (2006), who found that high-stakes exams tend to cause greater anxiety. Participants' socio-economic level may play a role in mitigating TA in this population (Putwain, 2007).

A statistically significant difference was found in the level of anxiety between men and women, with women having a higher score. This finding agrees with other studies (Cassady & Johnson, 2002; Núñez-Peña et al., 2016; Santana & Eccius-Wellmann, 2018), who found that women show more anxiety before exams, compared to men in equal situations. In contrast, the different fields of study show no significant differences regarding TA. These findings differ from those of Grandis (2009), who concluded that the students who present more cognitive anxiety before exams are those in the field of humanities, while those in engineering show more avoidance anxiety. In our study, engineering students showed greater anxiety than other students did; at the same time, engineering students obtained higher TOEIC scores. This may be explained as an effect of positive anxiety (Eccius-Wellmann et al., 2017; Furlan et al., 2014); because they feel anxious, students are motivated to study and prepare for the exam.

Finally, it was found that, excepting the engineering students, as mentioned above, the greater the anxiety, the lower the TOEIC test score. These results coincide with those found by Santana and Eccius-Wellmann (2018). For most of the population studied, the level of anxiety did not significantly affect the score. However, it must be highlighted that even a moderate score on the CAFEU may have a negative impact on the TOEIC results, as shown by the linear regression. Thus, it is important to consider that anxiety may be a factor in the failure to obtain the required score. At the same time, because women suffer more from TA than men do, they may be victims of unintended discrimination. These results indicate that it is necessary to address anxiety, even if only a few test-takers suffer from it.

Pedagogical Implications

Institutions that require their students to present a standardised test to prove their language proficiency should consider including TA reduction strategies in their curriculum. These strategies can be divided into three categories: test preparation, emotional control, and alternative assessment.

Test preparation strategies help students to become familiar with the exam they will take. This familiarity will, in turn, reduce uncertainty and mitigate anxiety. Examples of test preparation strategies include learning about the exam, its format, the time allotted for each section, and the types of questions that are included. Because many students nowadays study English through communicative methods with little focus on grammar, this should also be included in the course. Advanced courses at this university tend to focus on reading and listening activities, on specialised language according to the field of studies, or on academic writing. They should not assume that students no longer need grammar practice.

Interventions to teach emotional control include stopping negative thoughts and relaxation (Serrano Pintado & Escolar Llamazares, 2014). Mindfulness training to reduce anxiety has also been found useful in some cases (Carsley et al., 2015; Cunha & Paiva, 2012). Some examples of mindfulness activities include breathing exercises, relaxation, or even colouring, in the case of younger learners. Metacognition training is also a useful strategy for test-takers (Xu & Huang, 2018). This involves having students evaluate their strengths and weaknesses regarding the test they will take. They create a plan to overcome their weaknesses, including a timeline and monitoring their progress towards their goal.

The abovementioned interventions should be useful to prepare most students for an upcoming proficiency test, regardless of whether they feel moderate anxiety or no anxiety at all. However, it must be considered that there are students who suffer severe and debilitating anxiety when faced with an exam. These students may need psychological intervention, rather than mere study skills training (Serrano Pintado & Escolar Llamazares, 2014).

Finally, institutions should consider alternatives to exams as a means of proving proficiency. At the university where this research took place, English was established as a graduation requirement considering that students would need it in their professional lives. For the students, however, the test has become the end, and not the means to an end. The university should consider offering other options to prove language abilities, such as professional portfolios, or community service involving English (see Santana & Ventura Michel, 2020). These alternatives would be important for test-takers with severe forms of anxiety.

Conclusions

The present study sought to explore the possible role of test anxiety in the accreditation of an English language proficiency test, the TOEIC. The level of anxiety reported by students who took the TOEIC test to accredit English as a requirement for their university studies was measured and compared to their test results.

The first objective of the research study was to determine if TA is a widespread phenomenon among the study population. The results indicate that it is not; most respondents stated feeling little or no anxiety when faced with the test. The second objective was to explore how TA affects test scores. The findings indicate that there is a moderate negative effect and that increased anxiety leads to a lower score on the TOEIC.

Thus, the study found that anxiety explained poor performance in only a small percentage of the participants. However, for this small group of students, TA may mean the difference between graduating from the university or not. Therefore, it is worth focusing on reducing the levels of anxiety so that this factor has the lowest possible incidence. Increased test preparation and strategy training may be useful in this regard.

References

Álvarez, J., Aguilar, J. M., & Lorenzo, J. J. (2012). La ansiedad ante los exámenes en estudiantes universitarios: Relaciones con variables personales y académicas. *Electronic Journal of Research in Educational Psychology, 10*, 333–354.

Carsley, D., Heath, N. L., & Fajnerova, S. (2015). Effectiveness of a classroom mindfulness coloring activity for test anxiety in children. *Journal of Applied School Psychology*, *31*(3), 239–255. https://doi.org/10.1080/15377903.2015.1056925

Cassady, J. C., & Finch, W. H. (2014). Confirming the Factor Structure of the Cognitive Test Anxiety Scale: Comparing the Utility of Three Solutions. *Educational Assessment*, *19*(3), 229–242.

https://doi.org/10.1080/10627197.2014.934604

Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, *27*(2), 270–295. https://doi.org/10.1006/ceps.2001.1094

Contreras-Soto, A., Véliz-Campos, M., & Véliz, L. (2019). Portfolios as a strategy to lower English language test anxiety: The case of Chile. *International Journal of Instruction*, *12*(1), 181–198. https://doi.org/10.29333/iji.2019.12112a

Cunha, M., & Paiva, M. J. (2012). Text anxiety in adolescents: The role of self-criticism and acceptance and mindfulness skills. *The Spanish Journal of Psychology*, *15*(2), 533–543. https://doi.org/10.5209/rev_sjop.2012.v15.n2.38864

Eccius-Wellmann, C., Lara-Barragán, A. G., Martschink, B., & Freitag, S. (2017). A comparison of mathematics anxiety profiles between Mexican and German students. *Revista Iberoamericana de Educación Superior*, *8*(23), 69–83. https://doi.org/10.22201/iisue.20072872e.2017.23.3011

Furlan, L. A. (2013). Efficacy of an intervention aimed at reducing test anxiety in Argentinean university students. *Revista Colombiana de Psicología*, *22*(1), 75–89.

Furlan, L. A., Cassady, J. C., & Pérez, E. R. (2009). Adapting the cognitive Test Anxiety Scale for use with Argentinean university students. *International Journal of Testing*, *9*(1), 3–19. https://doi.org/10.1080/15305050902733448

Furlan, L. A., Ferrero, M. J., & Gallart, G. (2014). Ansiedad ante los exámenes, procrastinación y síntomas mentales en estudiantes de la Universidad Nacional de Córdoba. *Revista Argentina de Ciencias del Comportamiento*, *6*(3), 31–39. https://doi.org/10.30882/1852.4206.v6.n3.8726

Grandis, A. (2009). Evaluación de la ansiedad frente a los exámenes universitarios. [Unpublished doctoral dissertation].Universidad Nacional de Córdoba.

Horwitz, E. K. (2010). Foreign and second language anxiety. *Language Teaching*, *43*(2), 154–167. https://doi.org/DOI: 10.1017/S026144480999036X

Hurley, M., & Padró, F. F. (2006). *Test anxiety and high stakes testing*. Pervasive, pernicious, punitive and policy-driven. *The International Journal of Learning: Annual Review* 13 (1): 163-170 https://doi.org/10.18848/1447-9494/CGP/v13i01/48170

Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: a distinction and some initial data. *Psychological Reports*, *20*(3), 975–978. https://doi.org/10.2466/pr0.1967.20.3.975

Liu, Y., Pan, H., Yang, R., Wang, X., Rao, J., Zhang, X., & Pan, C. (2021). The relationship between test anxiety and emotion regulation: the mediating effect of psychological resilience. *Annals of General Psychiatry*, *20*(40). https://doi.org/10.1186/s12991-021-00360-4

McDonald, B., & Boud, D. (2003). The Impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education: Principles, Policy & Practice*, *10*(2), 209–220. https://doi.org/10.1080/0969594032000121289

Myers, S. J., Davis, S. D., & Chan, J. C. K. (2021). Does expressive writing or an instructional intervention reduce the impacts of test anxiety in a college classroom? *Cognitive Research: Principles and Implications*, *6*(1). https://doi.org/10.1186/s41235-021-00309-x

Núñez-Peña, M. I., Suárez-Pellicioni, M., & Bono, R. (2016). Gender differences in test anxiety and their impact on higher education students' academic achievement. *Procedia - Social and Behavioral Sciences*, *228*(June), 154–160. https://doi.org/10.1016/j.sbspro.2016.07.023

Onwuegbuzie, A. J., & Seaman, M. A. (1995). The effect of time constraints and statistics test anxiety on test performance in a statistics course. *The Journal of Experimental Education*, *63*(2), 115–124. https://doi.org/10.1080/00220973.1995.9943816

Powers, D. E., & Schmidgall, J. (2018). *The TOEIC® Tests: A Brief History: Vol. III*. Educational Testing Service. https://www.ets.org/s/toeic/pdf/research-compendium.pdf

Putwain, D. W. (2007). Test anxiety in UK schoolchildren: Prevalence and demographic patterns. *British Journal of Educational Psychology*, *77*(3), 579–593. https://doi.org/https://doi.org/10.1348/000709906X161704

Rumsey, M., Thiessen, J., Buchan, J., & Daly, J. (2016). The consequences of English language testing for international health professionals and students: An Australian case study. *International Journal of Nursing Studies*, *54*. https://doi.org/10.1016/j.ijnurstu.2015.06.001

Santana, J. C., & Eccius-Wellmann, C. (2018). Gender differences in test anxiety in high-stakes English proficiency tests. *Electronic Journal of Foreign Language Teaching*, *15*(1), 100–111. https://e-flt.nus.edu.sg/wp-content/uploads/2020/09/santana.pdf

Santana, J. C., García-Santillán, A., & Escalera-Chávez, M. E. (2017). Variables affecting proficiency in English as a second language. *European Journal of Contemporary Education*, *6*(1),138-148. https://doi.org/10.13187/ejced.2017.1.138

Santana, J. C., & Ventura Michel, K. M. (2020). A qualitative look at test anxiety and English language proficiency tests. *International Journal of Assessment and Evaluation*, *27*(2), 15–29. https://doi.org/10.18848/2327-7920/CGP/v27i02/15-29

Serrano Pintado, I., & Escolar Llamazares, M. del C. (2014). Description of the general procedure of a stress inoculation program to cope with the test anxiety. *Psychology*, *05*(08), 956–965. https://doi.org/10.4236/psych.2014.58106

Sommer, M., & Arendasy, M. E. (2014). Comparing different explanations of the effect of test anxiety on respondents' test scores. *Intelligence*, *42*(1), 115–127. https://doi.org/10.1016/j.intell.2013.11.003

von der Embse, N., Jester, D., Roy, D., & Post, J. (2018). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *Journal of Affective Disorders*, *227*(September 2017), 483–493. https://doi.org/10.1016/j.jad.2017.11.048

Wu, J., & Lee, M. C.-L. (2017). The relationships between test performance and students' perceptions of learning motivation, test value, and test anxiety in the context of the English benchmark requirement for graduation in Taiwan's universities. *Language Testing in Asia*, *7*(1). https://doi.org/10.1186/s40468-017-0041-4

Xu, J., & Huang, Y. T. (2018). The mediating effect of listening metacognitive awareness between listening test anxiety and listening test performance. *Asia-Pacific Education Researcher*, *27*(4), 313–324. https://doi.org/10.1007/s40299-018-0388-z

Appendix

CAFEU translated into English.

1. During exams, my hands perspire.
2. When I have been in an exam for a while, I feel discomfort in the stomach and need to defecate.
3. Even though I can see well, when I start reading the exam, my eyes blur and I don't understand what I'm reading.
4. I am almost always late for an exam and when that happens I do not show up to take it.
5. When I finish an exam, I have a headache.
6. In exams, I feel that I lack air. I feel faint, very hot or a general malaise.
7. When a teacher stands next to me, it makes me very nervous and I cannot keep working
8. I get nervous when I see the teacher arrive in the classroom with the exams
9. My hands feel stiff during exams
10. Before entering the test, I feel "something" in my stomach
11. Just by reading or listening to the questions of the test, I go blank.
12. When I finish an exam, I feel sad, thinking that I did it incorrectly despite not having the results yet.
13. I think long before deciding whether to take an exam
14. While I'm answering an exam, I think I'm doing it wrong
15. I feel very bad if my classmates finish and start to turn in the exam before me
16. While I am taking the exam, I think that the teacher is constantly looking at me
17. I want to smoke during the exam
18. I really want to go to the bathroom during the exam
19. I get sick and look for any excuse not to do an exam
20. For me it is a comfort or relief when, for whatever reason, an exam is postponed
21. Despite having studied, I think that I will not be able to pass

22. Before beginning the examination, I think that I know nothing and that I should suspend it.
23. I cannot sleep the night before the test.
24. It makes me nervous if there are many classmates and / or teachers in the exam.
25. During the exam I felt dizzy and nauseous.
26. Before I turn in the exam, my mouth is dry; I cannot swallow.
27. If I sit at the front of the classroom, I get more nervous.
28. I get more nervous if the exam has a time limit to finish it and I do it worse or I do not finish.
29. When I'm doing an exam I feel like my heart beats faster.
30. When I enter the testing room, my legs shake.
31. I feel nervous in exams with few classmates.
32. When I finish the exam, I feel like I will not pass it.
33. I think that in the exam my nerves will make me forget everything I studied.
34. I take a long time to answer most of the questions and I'm almost always the last one to turn in the test.

Participatory Assessment Designing a Learner-driven Test

Mario Moya
University of East London, UK

Assessment, evaluation, and tests are contested terms in education, where different definitions have been attempted without any achieving consensus. This is possibly one of the reasons why the topic remains interesting and current for teachers whose different views and practices have given rise to innovative approaches where the judgement of learners' performance is much fairer, equitable, and reliable. In this article, assessment is seen as a cycle where evaluations and tests are set up at a particular moment to elicit a behaviour or a performance to measure learners' knowledge, competencies, and skills against a benchmark (Green, 2014). The assessment/test distinction is of no minor importance: the meaning of assessment is more encompassing as it highlights its dynamic nature, characterised by continuity and iteration, allowing for reflection and growth (Fulcher & Davidson, 2007). Assessment is a continuous process where the boundaries between a start and a finish stage are not always clear-cut. Therefore, it becomes necessary to acknowledge that it is impossible to separate assessment from learning since they are complementary and cumulative. Additionally, assessment involves reflection, which is a crucial element of learning, mainly when the punitive part, traditionally associated with the judgement of an assessor based on grades and scores, is removed. The purpose of this article is to argue for the benefits of a participatory approach when designing assessment tasks involving the voice of the learners. This stance follows the premises of dialogic teaching (Alexander, 2008), claiming that learning emerges from the interactions between teachers and learners where participation involves self-determination, exercised by promoting learner agency. As it happens in dialogue, learners and teachers actively engage in conversations and agree on criteria to determine performance as levels of success, with feedback used developmentally for learners to produce an action plan to improve their language performance.

Literature Review

Assessment and its Purposes

Assessment is a crucial part of education (Falchikov, 2013) and is a source of information for learners concerning their understanding of language facts and the performance of linguistic competencies. The primary purpose of assessment is the identification of learners' strengths and areas for development for future lesson planning or more efficient interventions, ensuring continuous improvement of both teaching and learning (Earl, 2004). Assessment data can be gathered in various ways, such as homework, observations, tasks, or classroom interactions (Arends, 2012). Embedding assessment as a form of classroom-based practice is necessary for teachers to show their learners' progression (Muijs & Reynolds, 2018; Suah & Ong, 2012) and to reflect on their teaching approaches (Tomlinsom, 2010). Teachers partake in a wide range of assessment roles, which can be challenging, as this is not a simple task since assessment depends on its purposes, the knowledge, as well as the language competencies and skills to be assessed (Suah & Ong, 2012).

Over the years, different approaches have been developed to make assessment more manageable. For instance, assessing learners' performance regularly during lessons is known as *formative assessment* (Brookheart, 2018). A distinctive feature of this type of assessment is the use of feedback as a scaffolding mechanism for progression (Black & William, 1998). Some studies (Shepard et al., 2018) have demonstrated that such an approach, which usually focuses on discrete items, results in more extensive learning than testing. Conversely, *summative assessment*, akin to testing, is typically conducted at the end of a learning unit to examine learners' ability, knowledge of a particular topic, or the performance of skills. Summative assessment allows an assessor to gather data and produce a report on progress (Weurlandera et al., 2012) as the learners' performance is awarded a grade. Although traditionally, these two assessment categories have been presented separately, Harlen (2005) argues that formative and summative assessments are not two distinct types, but they are complementary whilst acknowledging that the literature shows them separately simply because of a difference in purpose. At this stage, it is pertinent to echo Newton's (2007) claim, who argues that assessment goals can be interpreted in many ways, not necessarily restricted to those traditionally associated with scores to determine performance outcomes. Nonetheless, the author indicates that these purposes boil down to three primary levels: judgement, decision, and impact. The judgement level examines the technical aim of an assessment event (Newton, 2007), focusing on the assessment outcome as it produces a judgement through a grade. The decision level focuses on the reviews and actions to support decisions for the promotion or demotion of learners, whilst the impact level ensures that students continue to remain motivated when learning and achieve common ground for a topic or skills. This level provides valuable information for lesson planning and the evaluation of teaching methods.

Earl's (2004) seminal work on types of assessment provides further clarity to the apparent dichotomy between formative and summative assessment by introducing the concepts of *assessment of learning* (AoL), *assessment for learning* (AfL), and *assessment as learning* (AaL). AoL is usually completed at the end of a term, course, or unit, utilising a test or exam that includes various questions based on the material covered over a period of study. AoL is summative since the results are usually in marks or grades based on the correct answers reported to stakeholders, there is little to no feedback, and students cannot gain direction for improvement. The outcome of this type of assessment only shows areas where learners performed well and those that need improvement. This type of assessment dominates the standardised examination systems in English as a foreign language and is externally judged as the most reliable type. However, research carried out to measure the impact of summative assessment on reliability, accuracy, and its advantages seem to be limited (Earl, 2004; Marzano, 2006).

In turn, AfL emphasises formative assessment as it frequently happens from the initial stages of learning (Earl, 2004). It focuses on strengths and weaknesses, helping learners unlock their potential to progress onto the next learning phase. A distinctive feature of this type of assessment is the use of feedback, a dialogue between teacher and learners signposting the subsequent learning and teaching steps to ensure gaps are addressed. AfL requires teachers to use their knowledge and understanding of the students and their contexts to identify particular learning needs. The practice of AfL then requires a considerable amount of pedagogical expertise, as well as subject-specific and curricular knowledge (Ball et al., 2008; Manizade & Mason, 2010). Whilst Black and William (1998) promoted the concept of AfL, Dann (2014), claims that the use of this type of assessment and feedback has led to no improvement or, in some areas, to no improvement at all, worsening the learners' performance.

Lastly, AaL extends the range of formative assessment by emphasising the learners' role in the assessment practice, in other words, their *agency*, which is one of the pillars of participatory pedagogy. It is pertinent to point out that agency is a theoretical construct as it is motivation and intelligence. A classical definition of learner agency indicates that this is "the socioculturally mediated capacity to act" (Ahearn 2001, p.112), stressing the situated human nature in sociocultural terms and the role of mediation. However, the definition of agency is not free from nuances since the "capacity to act" is open to different interpretations. For instance, Mercer (2012, p. 42) argues that "whilst an individual's capacity to act is widely accepted as being socio-culturally, contextually and interpersonally mediated, it also needs to be understood in terms of a person's physical, cognitive, affective, and motivational capacities to act." In other words, the main goal of AaL is for learners to become assessors (Earl, 2004) by active use of metacognitive strategies involving monitoring their learning and using feedback to underpin their learning and master skills. This type of assessment prepares learners for interacting with everyday situations outside the confines of a classroom by encouraging them to make their own judgements and decisions (Brookheart, 2018). In the same vein, Shepard et al.,(2018) state that learning that takes a socio-cultural stance and is linked to students' everyday life leads to higher engagement. However, AaL is a complex interplay of assessment, teaching and learning with the notion that learners must understand their own learning progress and goals through various processes involving cognition and a willingness to be active and engaged. One of the critical features of AfL and AaL is the use of feedback, a mediating tool rooted in the sociocultural theory to ensure that learners become active and take ownership of their learning experience. Jiménez Segura (2015) adds that learners must abandon the passive role of following instructions in a process led by an academic expert (i.e., teachers) and be actively involved in self-regulating their learning. Although this type of assessment practice may prepare students to take responsibility for their learning, this suggests that this type of assessment may only be implemented with "more able students" who have already achieved some self-determination.

"Assessment as learning" as a metacognitive practice

A sociocultural perspective explains that human activity is mediated by tools or artefacts emerging and shaped by the socio-cultural-historical context (Lantolf & Thorne, 2006). The educational tradition in the Western world has largely been mediated by the same tools, for example, language, textbooks, and displays, amongst others. By *practice,* Arnseth (2008) means the work of cultural extension and transformation in time. In education, the acquisition of new knowledge and skills provides individuals with opportunities for their self-actualisation and the group to which they belong. In this sense, Lave & Wenger (1991, p. 35) acknowledge that "learning is not merely situated in practice… but [it] is integral part of generative social practice in the lived-in world." This, therefore, means that practice involves learning how to problem-solve in the lived-in world, this is, the world as it is experienced in social practice, which is relational and agency-driven, where experiences are developed and challenged in and through social relationships.

Within this perspective, a school and all the social-cultural practices that take place in that setting is seen as a system structure where knowledge cannot be separated from the context of its use or the situation where it occurs (Brown et al., 1989). These relationships allow the development of metacognition, encompassing two distinct dimensions: the self (individual) and co-regulation (shared cognition). Garrison & Akyol (2013) relate metacognition to the process of inquiry at three levels of *presence*, claiming that:

> Metacognition enhances and refines the inquiry process in a collaborative constructivist learning environment (…). Metacognition is manifested at the convergence of the social, cognitive, and teaching presences in proportions reflective of the nature of the task and the capabilities of the participants. Social presence creates a purposeful environment in which students can have a connection to what others are thinking. Through cognitive presence, students have an increased understanding and awareness of the inquiry process (i.e., metacognition) which, in turn, helps them improve their regulation of cognition (…). Teaching presence (metacognitive awareness) encourages participants to become cognitively aware and develop regulatory skills for self and other's learning. (2013, p. 85)

One of the keywords in the above definition is *presence*, which means *interaction* in the sociocultural theory. An individual is present in a group or a community of practice when interacting with others. It is essential to remind the reader that socio-constructivists claim that learning occurs in two distinct stages: intrapersonal or intramental, and interpersonal or intermental (Vygotsky, 1978). This position is then very relevant to understanding why assessment is a continuous process constructed and negotiated with others through language.

Dialogic Teaching and Participatory Pedagogy

A wide range of research on the interactions between teacher-learners and learners-learners in different language learning settings indicates that teachers tend to dominate exchanges and that most of those interactions are ineffective (Donoso & López, 2020). Most of the time, learners passively listen to the teacher in whole-class sessions or work individually

(Kostadinovska-Stojchevska & Popovikj, 2019). A participatory pedagogy based on dialogue as a tool for teaching and learning is not new (Mercer, 2000; Alexander, 2008). Alexander (2008) claims that in dialogic interactions, learners are exposed to alternative perspectives and required to engage with another person's views in ways that challenge and deepen their own conceptual understandings.

The findings of Blatchford et al.'s study (2006) in the context of social (participatory) pedagogic research into groupings (SPRinG) showed that all the learners were more likely to be involved in tasks when they had a chance to exchange views. Additionally, the number of negative behaviours amongst students was considerably low whilst learner-learner interactions were more productive, and there was less need for teacher support. When the learners worked in groups, lessons tended to be more tasks than person focused. The researchers concluded that a programme like SPRinG encouraged more connectedness amongst learners whilst increasing opportunities for students to learn from one another as a direct result of the type of talk they used, which Wegerif (2008) calls *exploratory*.

Talk as a form of collaboration has been identified as a tool facilitating learning and the views of Bruner (1996), Halliday (1993), Mercer (2000), and Alexander (2008) are of relevance for a participatory pedagogy. Mercer (2000, p. 4) claims that "language is not just a means by which individuals can formulate ideas and communicate them, it is also a means for people to think and learn together." He identifies exploratory talk as one encouraging learning and defines it as:

> … [the type of talk] in which partners engage critically but constructively which each other's ideas. Relevant information is offered for joint consideration. Proposals may be challenged and counter-challenged, but if so reasons are given and alternatives are offered. Agreement is sought on the basis for joint responses. Knowledge is made publicly accountable and reasoning is visible in the talk. (Mercer, 2000, p. 98)

In turn, Wegerif (2008) distinguishes *exploratory talk* from other types of peer talk, such as *cumulative talk*, where the speakers build positively but uncritically on what the others have said, and *disputational talk*, characterised by disagreement and individualised decision-making. Exploratory talk can also be described as an orientation towards shared cognition for reaching consensus, for example, by exploring viewpoints in a group. This is what leads Mercer (2000, p. 8) to use the term *interthinking* to refer to the "joint, coordinated intellectual activity which people regularly accomplish when using language."

The model of participatory pedagogy based on interthinking in Latin America and the Caribbean is not a transient fad. Emerging from the work of Paulo Freire (2006), different models of school management and classroom practices have emerged in many countries in the area. One, in particular, is that of the *Escuela Nueva* (Colbert & Arboleda, 2016) in Colombia, whose set-up fosters the social construction of knowledge through dialogue and interaction. The authors highlight the approach of *Escuela Nueva* by indicating that:

> …teachers evolve from someone who imparts information or knowledge to someone who becomes a facilitator, interacts with the community, links knowledge with the local context and recognises accomplishments in social behaviour. In turn, a central role is granted to the child's thinking by supporting the construction of knowledge while working in small groups, establishing a direct relationship between learning and the physical and community environment, respecting the different paces and by fostering personal and social development in democratic and autonomous contexts. (Colbert & Arboleda, 2016: 393)

The above is an example of how a primary school within the context of a Spanish-speaking environment uses a participatory approach where teachers and learners position themselves as agents whilst promoting each other's participation through dialogic teaching and learning (Kumpulainen & Lipponen, 2012). The following experience considers a participatory approach to the design and conduct of an assessment built upon the principles expounded in this literature review.

Applying a Participatory Approach to Assessment Design and Implementation

The breakout of the COVID-19 pandemic in March 2020 severely disrupted long-established educational practices, changing paradigms and ways of imparting guided instruction in times of rapid changes and uncertainty. Most educational organisations had to close and move to an online provision with very little or no experience of the medium. Within this context, 18 commercial aviation pilots had to sit a standardised English examination to gain full accreditation in the industry. The exam, which usually takes place at a designated centre, has five components: technical vocabulary, English for communication, reading comprehension, listening comprehension, and three simulation tasks to assess the candidates' ability to communicate in different scenarios. The exam is equivalent to Level B2/C1 of the European Common Reference for Languages, and, under normal circumstances, each part lasts 60 minutes. The deadline for the exam could not be moved, although the Assessment Board gave each group of candidates some flexibility to meet the assessment criteria. While the methodology was made flexible, the standards, including procedures, such as blind double first and second marking and moderation, remained unchanged.

By the time consultancy was requested, the candidates had gathered *ad hoc* to revise contents and were working in small online groups. Without a clear indication of their current level of performance, it was necessary to understand the nature of the examination, the teaching and learning approach used during training, and details concerning the exam to provide advice. This first action involved a needs analysis followed by the observation of online interactions to produce a catalogue of learning strategies used during the revision phase. This stage also consisted of collecting samples including summaries, vocabulary lists, mind-maps, and multiple scribbles on textbook pages. The next step involved the use of the students' documents to create a corpus, which alongside the course syllabus, contributed to the identification of the language items and the contexts in which these items were used. The fieldwork was followed by conversations between the consultant and the learners on their learning preferences, using their work during their revision to prompt ideas. The conversations gave learners some space to talk about the language contents of the exam and their feelings towards it, and secondly, it was also an opportunity for the consultant to use the information the participants provided to select different types of test tasks. The analysis of the data only focused on the revision of the course contents and provided additional insight for the selection and gradation of tasks at the level of design. Three different versions of the exam were produced and piloted in five different settings, followed by several iterations until the final version was agreed. An example of how the procedure was put into practice is provided in the vignette section below.

The focus on revision was not arbitrary but followed the premises of peer mediation aligned with the tenets of the socio-constructivist paradigm. According to this view, learners working together can co-construct contexts in which expertise emerges as a feature of the group rather than residing in any given individual in the group (Lantolf, 2000). The sociocultural theory holds that human forms of mental activity arise in our interactions with other members of our culture and the specific experiences we have with the artefacts produced by our ancestors and our contemporaries. When learners revisit knowledge or rehearse competencies in groups, they reorganise different forms of consciousness. Lantolf (2000) argues that:

> This reorganisation occurs as a result of the culture-specific interactions we have with other individuals and with the artefacts constructed and deployed by the culture. These artefacts may be relatively simple physical objects such as paper and pencil used to help us remember what to purchase at the store or to carry out multiple digit arithmetic operations. (p. 79)

The experience of many teachers and learners alike is to underestimate the potential of group revision as a space where dialogue constructs artefacts to promote and scaffold learning. These artefacts produced during the revision phase were crucial in designing a new exam adapted to an online environment. Following a rigorous process of validation that included pilot testing for quality assurance, the 18 pilots sat the exam entirely designed following their input. As the exam was seen as one single event within the more extensive learning experience, this was followed up with feedback for the learners to produce an action plan where the actions became their learning objectives for the subsequent course.

Feedback was essential in the process of helping learners with the identification of the actions to meet their learning objectives. Recent literature has identified that the critical determining factor in feedback effectiveness is how individual students engage with it (Hattie et al., 2016; Lipnevich et al., 2016; Nicol, 2013). Engagement with feedback involves receiving, perceiving, interpreting, and understanding it and using it in some way to improve learning (Handley et al., 2011; Hargreaves, 2011; Nicol, 2013). However, feedback may lose its potential to influence learning in each of these stages of engagement positively. For example, if learners do not recognise feedback (reception), do not perceive it to be useful (perception), do not understand it (interpretation), or do not have time or opportunity to use it or are unwilling or unable to do so (use), feedback will not influence their learning in any positive way (Brookhart, 2018; Gamlem & Smith, 2013; Havnes et al., 2012). These stages of engagement with feedback should not be seen as a linear process. Instead, they interact as students employ cognitive and affective mechanisms in interaction with the feedback message (Lipnevich et al., 2016; Van der Schaaf et al., 2013). For this reason, the choice of words and the tone of the message were given particular attention for the learners to engage with the feedback. This was achieved by emphasising achievement instead of the knowledge and competencies they did not demonstrate in the exam. As a result, the learners saw the feedback as an opportunity to exchange ideas with their colleagues about their achievements and areas for future development.

Vignette: Including Learners' Voices in Test Design

It is essential to acknowledge the various contexts and practices around assessment in any given region or, sometimes, within one school. However, teachers should not feel constrained by traditional ways of evaluating learners imposed by educational policies or curriculum requirements that largely ignore students' needs. Although there are some excellent alternative forms of assessment such as portfolios, blogs, and documentaries, to mention a few, creativity provides endless opportunities to make tests part of our students' integral learning experience instead of a finished product. Using learners' voices can add to those lesser-known forms of classroom assessment. The experience reported in this article provides an interesting basis for teachers and students to create spaces for collaboration in the design of tests.

Embed opportunities for focused revision during lesson time. Focused revision is characterised by purpose. Therefore, learners must know what aspect of the course content they need to concentrate on, for instance, accuracy, fluency, vocabulary, pronunciation, or a combination of all. Keeping the revision focused will enable learners to develop their awareness by searching examples, engaging more actively with the learning resources.

Ask learners to participate in the test design. During instruction, learners probably have had a chance to experience different forms of input and varied tasks. As they reflect aloud on the teaching approaches they have been exposed to, teachers may gain valuable information about the effectiveness of their teaching methodology. They will be able to gauge the most appropriate tasks to capture students' performance at one time. Asking learners about their views will develop a sense of ownership of their learning, resulting in better outcomes.

Ask learners to identify critical items to be included in the exam. This stage attempts to trigger views and opinions about what constitutes an essential item and provides teachers with an opportunity to undertake formative assessment and check how much revision is needed for learners to achieve a good grade. Critical items can be recorded by, for example, asking learners to write them down on a piece of paper for the teacher to collect. After the compilation has been made, those items are to become the content of the exam.

Involve learners in the development of the assessment criteria. Learners need to know which benchmark teachers use to assess their performance for informative purposes and ensure they understand how their work is graded. Teachers can use some previous assessment criteria as a draft for learners to work in pairs creating rubrics. This task promotes interthinking and adds ownership to the marking process by making it fairer.

Allow learners to engage with feedback. Engagement does not mean just reading a set of teachers' comments but acting upon them. Sometimes learners' failure to engage with feedback is because the statements provided by the teacher are too vague and, therefore, irrelevant. Learning how to give feedback and personalising it may have a different effect, thus prompting learners to action. Modelling how to engage with feedback through action planning is an effective way for learners to know what engagement means.

Give learners time to respond to feedback. Communication is a loop where there is one speaker and a listener who alternate roles. If there is no response to feedback, this means that there is no engagement. By asking learners to identify their learning goals and think about strategies to meet them, they will be exercising their metacognitive skills. Teachers can use those goals to monitor individual learners' performance for formative purposes.

Review language when talking about exams. Comments such as passing or failing an exam are a legacy of traditional views on assessment prevalent in many school settings—the use of this type of language centres on an objectified individual who wears the label of achiever or non-achiever. Participatory pedagogy views all learners with an ability and, as such, is inclusive. Changing the language when talking about exams does not mean avoiding negative words, but it should move teachers to change their mindsets and their learners' to acknowledge that each student is good at something.

Conclusion

This article has articulated why assessment should be seen as an integral part of the students' learning experience instead of a by-product of teaching. Far from attempting to simplify issues around the complexities of assessment, the example provided here shows how the learners' voices can add to the validity and reliability constructs whilst allowing students to exercise their agency within the framework of participatory pedagogy to promote the learners' sense of ownership of their learning. For teachers, including the students' voices in the design and conduct of assessment may be challenging if they are unfamiliar with a dialogic approach; however, such a challenge is not different from those posed by more traditional forms of evaluation with the added value of enhancing the teaching and learning experience.

References

Ahearn, L. M. (2001). Language and agency. *Annual Review of Anthropology, 30*(1), 109-137.

Alexander, R.J. (2008). *Towards dialogic teaching. Rethinking classroom talk* (4th edition). Dialogos.

Arends, R. I. (2012). *Learning to teach.* McGraw Hill Open University Press.

Arnseth, H.C. (2008). Activity theory and situated learning theory: Contrasting views of educational practices. *Journal of Pedagogy, Culture & Society, 16*(3), 314–342.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59, 389–407.

Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74.

Blatchford, P., Baines, E., Rubie-Davies, Bassett, P., & Chowne, A. (2006). The effect of a new approach to group work on pupil-pupil and teacher-pupil interactions. *Journal of Educational Psychology, 98*(4), 750–765.

Brookhart, S. (2018). Summative and formative feedback. In A. A. Lipnevich & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback* (pp. 52–78). Cambridge University Press.

Brown, J.S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–42.

Bruner, J.S. (1996). *Toward a theory of learning.* Harvard University Press.

Colbert, V. & Arboleda, J. (2016). Bringing a student-centered participatory pedagogy to scale in Colombia. *Journal of Educational Change,17*, 385–410.

Dann, R. (2014). Assessment as learning: blurring the boundaries of assessment and learning for theory, policy, and practice. *Assessment in Education: Principles, Policy & Practice, 21*(2), 149–166.

Donoso, B., & López, J. (2020). Management of teacher talking time in adult students of English as a foreign language at University of Guayaquil. *Publicaciones*, *50*(2), 103–119.

Earl, L.M. (2004). *Assessment as learning: Using classroom assessment to maximise student learning*. SAGE.

Falchikov, N. (2013). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education.* Routledge.

Freire, P. (2006). *Pedagogy of the oppressed.* Continuum.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment. Advanced resource book.* Routledge.

Gamlem, S. M., & Smith, K. (2013). Student perceptions of classroom feedback. *Assessment in Education: Principles, Policy & Practice, 20*(2), 150–169.

Garrison, D.R. & Akyol, Z. (2013). Toward the development of metacognition construct for communities of inquiry. *Journal of Internet and Higher Education, 7*, 84–89.

Green, A. (2014). *Exploring language assessment and testing. Language in action*. Routledge.

Halliday, M.A.K. (1993). Toward a language-based theory of learning. *Journal of Linguistics and Education, 5*(2), 93–116.

Handley, K., Price, M., & Millar, J. (2011). Beyond "doing time": Investigating the concept of student engagement with feedback. *Oxford Review of Education, 37*(4), 543–560.

Hargreaves, E. (2011). Teachers' feedback to pupils: Like so many bottles thrown out to sea? In R. Berry & B. Adamson (Eds.), *Assessment reform in education: Policy and practice* (pp.121–133). Springer.

Harlen, W. (2005). Teachers' summative practices and assessment for learning—tensions and synergies. *The Curriculum Journal, 16*(2), 207–223.

Hattie, J., Gan, M., & Brooks, C. (2016). Instruction based on feedback. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 290–324) (2nd edition). Routledge.

Havnes, A., Smith, K., Dysthe, O., & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in Educational Evaluation, 38*(1), 21–27.

Jimenez Segura, F. (2005) Uso del feedback como estrategia de evaluación: aportes desde un enfoque socioconstructivista. *Revista Actualidades Investigativas en Educación, 15*(1), 1–24.

Kostadinovska-Stojchevska, B. & Popovikj, I. (2019). Teacher talking time vs. student talking time: Moving from teacher-centered classroom to learner-centered classroom. *The International Journal of Applied Language Studies and Culture, 2*(2), 25–31.

Kumpulainen, K., & Lipponen, L. (2012). Harnessing funds of knowledge in dialogic inquiry across formal and informal learning environments. In P. Jarvis & M. Watts (Eds.), *The Routledge international handbook of learning* (pp. 112–125). Routledge.

Lantolf, J.P. (2000). Second language learning as a mediating process. *Language Teaching, 33*, 77–79.

Lantolf, J.P. & Thorne, S.L. (2006). *Socio-cultural theory and the genesis of second language*. Oxford University Press.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Lipnevich, A. A., Berg, D. A. G., & Smith, J. K. (2016). Toward a model of student response to feedback. In G. T. L. Brown & L. R. Harris (Eds.), *The handbook of human and social conditions in assessment* (pp. 169–185). Routledge.

Manizade, A., & Mason, M. (2010). Using Delphi methodology to design assessments of teachers' pedagogical content knowledge. *Educational Studies in Mathematics, 76*(2), 183–207.

Marzano, R. (2006). *Classroom assessment and grading that work*. ASCD Publications.

Mercer, N. (2000). *Words and minds*. Routledge.

Mercer, S. (2012). The complexity of learner agency. *Apples. Journal of Applied Language Studies, 6*(2), 41–59.

Muijs, D., & Reynolds, D. (2018). *Effective teaching evidence and practice* (4th edition). SAGE.

Newton, P. (2007). Clarifying the purposes of educational assessment. Assessment in education: *Principles, Policy & Practice, 14*(2), 149–170.

Nicol, D. (2013). Resituating feedback from the reactive to the proactive. In D. Boud & E. Molloy (Eds.), *Feedback in higher and professional education* (pp. 34–49). Routledge.

Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Classroom assessment principles to support learning and avoid the harms of testing. *Educational Measurement: Issues and Practice, 37*(1), 52–57.

Suah, L., & Ong, S. (2012). Investigating assessment practices of in-service teachers. *International Online Journal of Educational Sciences, 4*(1), 91–106.

Tomlinson, C. A. (2010). Learning to love assessment. *Educational Leadership: Journal of the Department of Supervision and Curriculum Development, N.E.A, 15,* 54—65. ·

Van der Schaaf, M., Baartman, L., Prins, F., Oosterbaan, A., & Schaap, H. (2013). Feedback dialogues that stimulate students' reflective thinking. *Scandinavian Journal of Educational Research, 57*(3), 227–245.

Vygotsky, L.S. (1978). *Mind in society*: *Development of higher psychological processes.* Harvard University Press.

Wegerif, R. (2008). Reason and dialogue in education. In B. van Oers, E. Elbers, R. van der Veer & W. Wardekker (Eds.), *The transformation of learning* (pp. 273 - 286). Cambridge University Press.

Weurlandera, M., Söderbergb, M., Schejac, M., Hultb, H. & Wernersonb, A. (2012). Exploring formative assessment as a tool for learning: students' experiences of different methods of formative assessment. *Assessment & Evaluation in Higher Education, 37*(6), 747–760.

Assessment for Learning: An Exploration of the Potential of Exit Tickets and Choice Boards to Boost EFL Learners' Achievement

Luciana Ortega
Asociación Rosarina de Cultura Inglesa / Instituto Superior San Bartolomé

Assessing students' knowledge and performance plays a key role in every EFL teacher's practice. Although instruction usually takes up most of the classroom time, there will inevitably be a moment when a teacher needs to know whether the students have actually learnt the contents taught. Summative assessment instruments are commonly used for such purpose. This type of assessment provides evidence of achievement and can be used to make decisions about the students' future courses.

The problem that many teachers experience when using formal assessment tools exclusively is that the information obtained about the students' progress achieved between the actual teaching stage and the test assigned as a knowledge-measurement instrument is usually very little. How can teachers be better informed of what happens during that blurry in-between period? Research findings indicate that one of the benefits of formative assessment is precisely that it helps to narrow the gap between the students' current knowledge and the learning goals (Greensetin, 2010). This paper will explore the potential of two formative assessment tools: exit tickets and choice boards used in 2021 in a B1+/B2 course for teenage students at Asociación Rosarina de Cultura Inglesa, a private language school in Rosario, Santa Fe, Argentina. It was an online course which met six hours per week and which prepared students for the international B2 First Cambridge English Assessment exam.

Theoretical Framework

Formative assessment practices result from a paradigmatic shift in the field of assessment which has moved from data-driven decision making to decision-driven data collection. The overarching idea is that assessments are simply procedures for drawing inferences followed by the collection of evidence needed to support those inferences or to make informed decisions. Formative assessment, then, entails both intention and action: the strategies designed and implemented to collect information about the learners' progress have to be accompanied by minute-by-minute and day-by-day decisions and actions. Based on the data collected, decisions to move forward, to clarify misinterpretations or simply to re-teach a whole point are easier to make. The very definition of formative assessment presupposes then the use of evidence to adapt teaching and meet learning needs (Black et al., 2004).

What teaching strategies can be implemented to use formative assessment as a means to enhance students' achievement? There are three main processes which take place during assessment for learning: identifying where the learners are on their learning (current knowledge), where they are going (objectives and expected outcomes) and how to get there (plans and procedures). Drawing on these processes, Wiliam and Thompson (2008) propose five key strategies that can be used to support the implementation of effective formative assessment, namely 1) clarifying, understanding and sharing learning intentions; 2) engineering effective classroom discussions, tasks and activities that elicit evidence of learning; 3) providing feedback that moves learners forward; 4) activating students as learning resources for one another, and 5) activating students as owners of their own learning.

The first strategy, sharing the lesson's objective with the learners in age-appropriate ways, can help them to understand what they are learning, why they are learning it and how they should learn it. The strategy can also help learners to reflect on whether or not they were successful. This, in turn, will foster metacognition, i.e., it will help the learners to raise awareness of their own thinking processes. The second strategy, eliciting evidence of learning, is related to effective questioning and to the concept of responsive teaching. Asking effective questions will prompt thinking and, in turn, it will help learners know what to retrieve and whether they have been successful in doing so. It will also allow the teacher to collect useful data, plan ahead and adjust instruction if necessary. Giving feedback, the third strategy, can also fulfil the function of moving forward rather than looking back on mistakes. Successful feedback does not only imply learners' improvement but also the development of problem-solving skills which learners will be able to use in future learning situations. The fourth strategy focuses on the roles performed among peers. Students do not learn in isolation. Allowing for a variety of classroom dynamics to take place will help students to know how to help one another and, most importantly, how to ask for help when needed. The fifth strategy centres around activating students as owners of their own learning to maximize the power of formative assessment. Research has shown that asking learners to prepare their own test questions with correct answers or playing the role of the teacher and explaining a concept to peers enhance the chances of retaining the new information in the long-term memory (Ebersbach et al., 2020).

Practical Application

Having outlined the theoretical aspects of formative assessment, i.e., defining what formative assessment is and summarising the positive impact that implementing this type of assessment has on students' overall performance, the second part of this work will focus on the practical teaching implications of the two strategies adopted in the course briefly described earlier: exit tickets and choice boards.

An exit ticket is a type of formative assessment linked to the objectives of a lesson. These tickets are usually issued at the end of a lesson to allow students time to reflect on what they have learnt and express their thinking about the new information. Typically, they are short and ask students to answer a few questions. Depending on the purpose, such questions may require students to document learning with prompts like "Write one thing you have learnt today."; they may emphasise the process of learning as in "Write one question you have about today's lesson."; or they may ask the students to evaluate the effectiveness of instruction by asking "Did you enjoy working in small groups today?". In other words, although the main goal of exit tickets is to check students' understanding, there are other reasons for implementing this formative assessment strategy as well: to verify whether students can solve a problem, to emphasize the essential part of a lesson and to gather questions that students may still have about a lesson. Exit tickets can vary in format and style. They can be actual slips of paper handed out to students to fill in or they can be prepared using a digital tool like Google Slides.

In short, exit tickets have a dual purpose: to provide the teacher with feedback and to challenge the students to evaluate and synthesise course content and their learning processes. They require students to think critically and write about a wide variety of classroom topics. The power of exit tickets lies in that the information collected will allow the teacher to measure not only whether the students are learning the material but also whether the learners have a superficial understanding of the information or some depth of understanding. Knowing this can help teachers prepare future lessons considering the students' actual needs.

Choice Boards, on the other hand, are usually assigned as homework. Essentially, a choice board is a menu of activities from which students can choose. In terms of format, they can be a one-page printed worksheet or they can be digitally designed with a digital tool like Google Slides or Canva. As hinted by the name, a choice board gives the learners the possibility of choosing their own adventure in terms of how they are going to complete the different activities. It allows students to demonstrate the set of skills that the teacher is assessing but gives them a choice and voice in terms of what task to do. That is why, unlike exit tickets, this strategy gives the students the power to self-regulate their learning. The level of cognitive rigour demanded in the tasks will vary in an attempt to differentiate instruction. Some learners will decide to complete the task they feel less difficult or achievable, while others may choose to challenge themselves and choose more demanding activities.

Both Exit Tickets and Choice Boards are practical teaching techniques which can support the implementation of the formative assessment strategies proposed by Wiliam and Thompson (2008). The examples discussed in this section will revisit such strategies to assess the potential of

Exit Tickets and Choice Boards in helping students understand how their success will be measured, mapping out effective instructional strategies, providing students with the information they need to better understand problems and solutions, getting students involved in the learning experience and teaching learners to self-monitor and self-regulate their learning.

The Exit Tickets and Choice Boards in this work were designed drawing on the revised version of Bloom's taxonomy of learning objectives (Anderson et al., 2001). Widely applied in the educational sphere, Bloom's taxonomy has been proven to help learners develop critical thinking skills and higher order cognitive abilities.

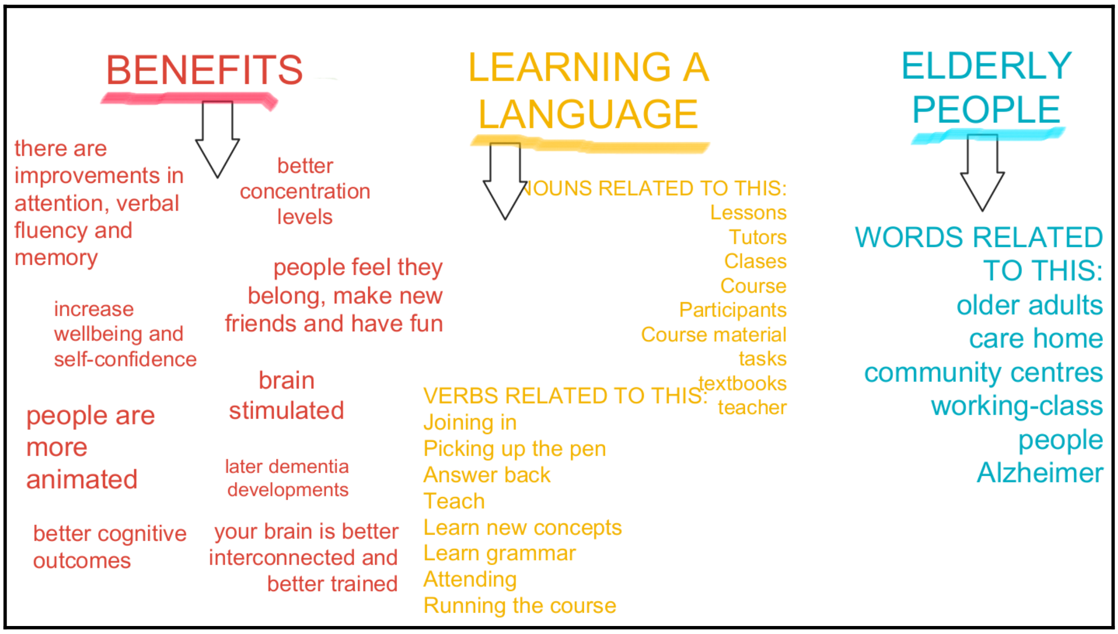
The taxonomy provides educators with a framework or organization for classifying lesson objectives. It is split into six different domains, namely remembering, understanding, applying, analysing, evaluating and creating. The taxonomy can be used as a spectrum that varies between lower-order and higher-order thinking with remembering being on the lowest order and creating on the highest. Descriptive verb lists aligned to the taxonomy objectives have been made available to help educators in the process of planning their lessons and writing lesson objectives.

Exit Ticket One

The lowest level of learning in Bloom's taxonomy is *remembering*. Exit Tickets at this level required students to recall previously learnt information. One of the first topics covered in the B2 First course curriculum is language learning. The topic was presented through an article about language learning and age. After reading the article individually and working with comprehension questions about the ideas in the text, a whole-class discussion followed. The students were asked to skim the article again and put it away afterwards. Then they were asked to work collaboratively on a Jamboard and enter words and/or phrases related to the topic dealt with in the article. Jamboard is a user-friendly app available in Google Drive which helps boost student collaboration and engagement as ideas can be sketched whiteboard-style on an interactive canvas. Figure 1 is an image of what the board looked like at the end of the activity.

Figure 1

Remember



This exit ticket shows a list of words and phrases that relate to the topic. As it can be seen in the figure, not only were the students able to remember several pieces of information from the article but they also managed to organize the items into subcategories (benefits, learning a language and elderly people), hinting here that they are also able to analyse, a higher-order skill. This retrieval activity makes learning visible and at the same time it allows both struggling and more advanced learners to have a say without highlighting their differences.

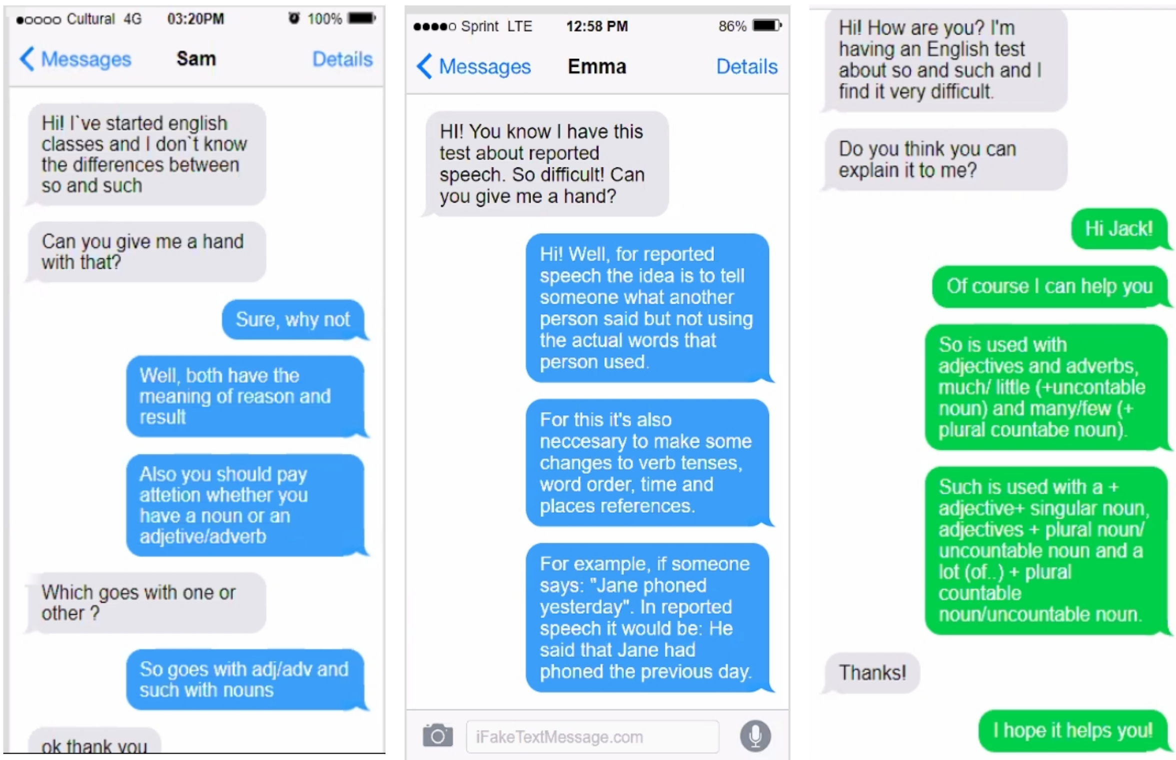
This exit ticket is an example of the formative assessment strategy proposed by Wiliam and Thompson (2008) that focuses on activating students as learning resources for one another. Collaboration was key to this activity in that each student was able to contribute to the board, but at the same time the final outcome reflected different levels of reading proficiency. The collaborative task capitalized on the power of students helping each other. While struggling learners only came up with a list of words related to the main idea of the article, high achievers grouped the disconnected words and phrases into sub-categories that made up lexical chains. By doing this, they presented a model of understanding lexical cohesion.

Exit Ticket Two

The following ticket is aligned to the second level of Bloom's taxonomy: *understanding*. Exit Tickets at this level require students to explain ideas or concepts. On this occasion, the students were asked to imagine that they had to explain the concept dealt with during that lesson to a student at a lower level or a partner that had been absent. To develop creativity and encourage engagement, the students had to use an online tool that allows them to create fake text messages and type the imaginative conversation they had with their partner or friend. Figure 2 shows examples of students' work.

Figure 2

Understand



In this exit ticket learners explain what they have learnt to a partner who has missed a lesson or to a student at a lower level. The examples shown above once again reveal very useful information for the teacher. Rather than repeating the rule of the structures of reported speech or the pattern and function of the determiner and adverb *so* and *such* respectively, the students had to select the pieces they considered most relevant for someone who had not received any instruction on the topic. Carrying out the task compelled them to revise the rules and to make them simple and clear.

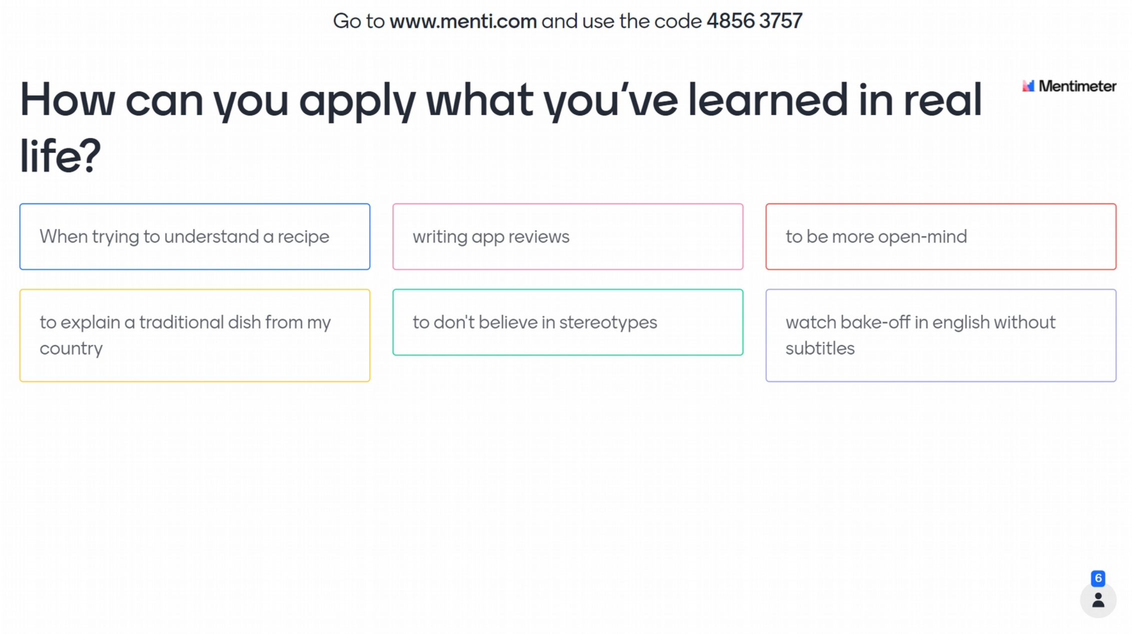
This example can be related to William and Thompson's (2008) third strategy: engineering effective classroom discussions, tasks and activities that elicit evidence of learning. The focus here is on eliciting evidence in contrast to merely asking direct questions. An analysis of the students' explanations will uncover the aspects that might need re-teaching or might act as a green light for the teacher to move on the following course content.

Exit Ticket Three

The third example is aligned to the *apply* level of learning in Bloom' taxonomy. Exit Tickets at this level require students to use information in new ways. In class, students had worked with the topic of food, recipes and mealtimes across cultures. The lesson plan integrated listening, speaking and writing skills together with the development of vocabulary. At the end of the lesson, students were asked to reflect upon ways in which the lesson content could be applied in real-life situations. The tool used was Mentimeter, an online audience engagement platform which allows to create open-ended questions, the answers to which are displayed in a bubble-like fashion. Figure 3 shows an image of what the students' answers looked like.

Figure 3

Apply



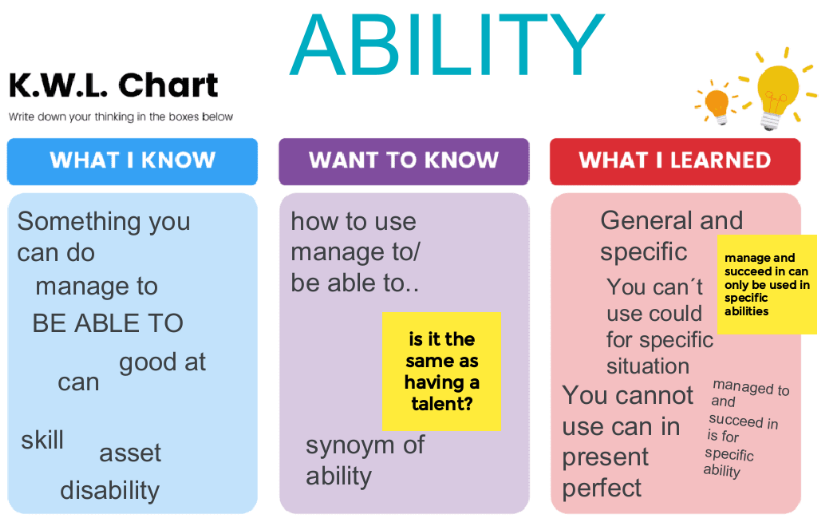
In this exit ticket learners suggest ways to apply what you have learnt in real life. While the previous exit tickets turned out to be effective ways of eliciting evidence of learning, this ticket uncovers data related to the students' self-driven motivations and that can be seen in the variety of answers entered. Some students focused on the applicability of the content to exam-related situations or contexts (e.g., writing a review or trying to understand a recipe) whereas others actually stepped out of the educational setting to consider the transferability of the concepts to authentic situations. Allowing students to read their partners' contributions (both the more direct and more elaborate ones) can help to prompt self-reflection. Hypothetical and imaginary as the answers to this ticket may be, their value lies in pushing the learners to think outside the box (or the classroom in this case) and consider the practical applications of course content. This Exit Ticket can be related to the last formative assessment strategy proposed by Wiliam and Thompson (2008): activating students as owners of their own learning. Being able to say how course content might come in useful in the future is a step forward to helping students become self-regulated learners.

Exit Ticket Four

The following exit ticket focuses on the fourth level, described in Bloom's taxonomy as *analyse*. Exit Tickets at this level require students to break information down into simpler parts to explore understandings. Figure 4 illustrates a ticket used to work on the concept of ability.

Figure 4

Analise



On this opportunity the students were asked to complete a KWL chart, in which they had to write notes about what they know (K), what they want to know (W), and what they have learnt (L) about the concept of ability (physical, mental, emotional, etc.). KWL charts are visual organizers which can be used to engage, guide and review. This task is usually effective because it involves the students in the learning goal.

In this case, the ticket issued was used at different stages of the lesson. As the figure shows, in the first column the students had to complete what they already knew about ability (the meaning of the term, grammatical aspects, etc.) This called for hypothesizing and the activation of prior knowledge. The answers entered in this column can be very useful in detecting misconceptions or inaccuracies.

Once the students are made aware of what they will be learning on a particular lesson, they are ready to start establishing connections to what they would like to learn, and a key aspect to be able to do this is curiosity. This relates to the first strategy of formative assessment that Wiliam and Thompson (2008) describe: clarifying learning intentions. The “What I want to know” column in particular focuses on making the lesson outcomes explicit. The more the students know about what the objectives of the lesson are, the higher the chances of achieving them. In the W column of the chart, the learners themselves suggest points or concepts to be dealt with. In this Exit Ticket, the students addressed the multifaceted aspects of the concept of ability. While some of them focused on the grammatical patterns to express ability (how to use manage to/be able to), others were interested in the meaning of the word (is it the same as having a talent?). Not all of the students’ interests matched the actual lesson objectives; however, the W column is a powerful tool for teachers to come to know what students feel motivated to learn and maybe include in future lessons.

In this example, working on the K column of the chart helped the students to narrow down the options in relation to the concept of ability and the W column helped them came up with more specific questions. When the time for the direct instruction stage arrived, the students were already tuned in into the topic and ready to work out the new and/or refined concepts about it. The L column provided the teacher and students with evidence of learning. Analysing the KWL chart allows students to make connections between what they knew before the lesson and what they know as a result of the learning experience.

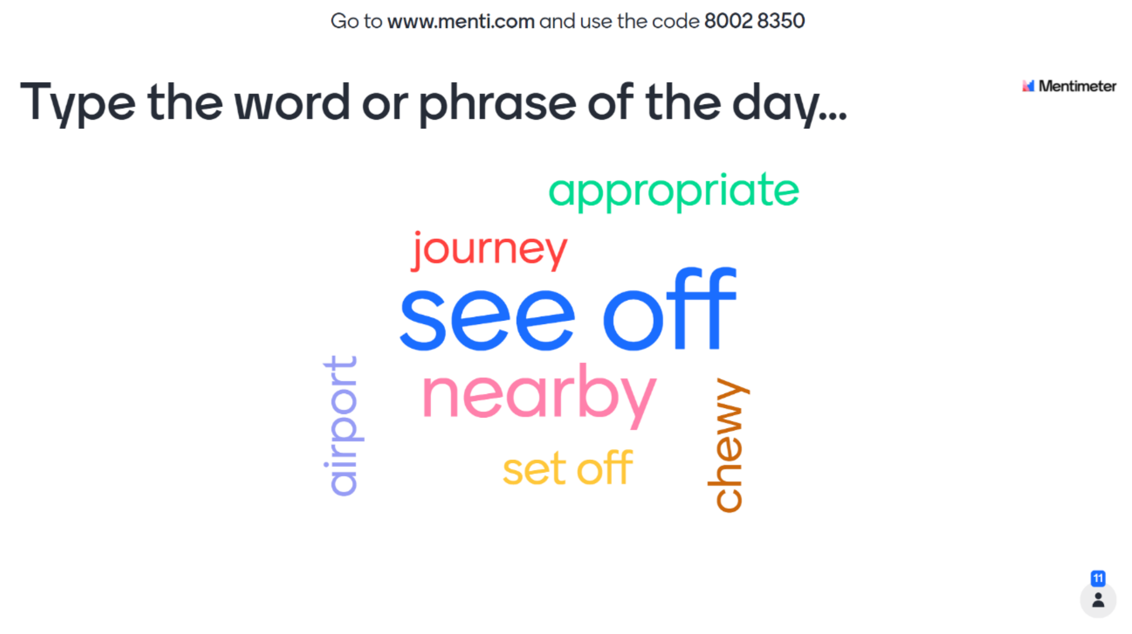
Having done this activity using Jamboard fostered collaboration among learners and helped to reinforce important points that students needed to know and remember. The KWL proved to be an effective tool which activated students' background knowledge and helped create a learning environment that allowed them to be active participants in their learning processes.

Exit Ticket Five

The next Exit Ticket is a word cloud which asks the learners to evaluate the learning experience at the end of a lesson. Exit Tickets at this level require students to justify a stand or decision, as can be seen in figure 5.

Figure 5

Evaluate



The context of the lesson in which this ticket was issued was travelling. Students had worked with a number of reading, listening, speaking and vocabulary activities. At the end of the lesson, they were asked to type the word or phrase of the day. The tool used was again Mentimeter, but this time the application automatically created a word cloud which grew and expanded as answers were submitted. In this application, the greater the number of times the same word is included, the bigger it will appear in the cloud. This allowed them to graphically show what the students considered to be the most important word of the day. In the case of the example under analysis, the biggest item was *see off*. Students had dealt with phrasal verbs connected with travelling and this one in particular was new to all students. Another word included in the cloud was *appropriate*. This relates to the focus of the lesson on the appropriateness of using words in particular contexts.

Some students considered that such discussion had been key to the development of the session. One of the most powerful strategies for formative assessment is engineering effective discussions, tasks and activities that elicit evidence of learning (Wiliam and Thompson, 2008). Having students recap on what they had dealt with resulted in a much more thoughtful set of responses. The voting system gave everybody a voice and the answers reflected the students' own learning interests or the aspects that they focused on during the lesson.

Exit Ticket Six

The last Exit Ticket example asked students to work with skills at the top of Bloom's taxonomy: *creating*. Exit Tickets at this level require students to create a new product or point of view as can be seen in figure 6.

Figure 6

Create

In this exit ticket learners create an original way to summarize or to show how they feel about the topic of the lesson. Before working on this task, the students had been working with relative clauses. This is a grammar structure generally included as content in lower-level courses, so teachers in more advanced courses might assume that students have already mastered this type of subordinate structures. Having students create an original way of explaining or expressing their feelings about a topic can be more than revealing.

On this occasion, students created memes using Canva, an online design and publishing tool. One of the features of this application is that it is user-friendly and intuitive so there is practically no need to devote time to explaining how to use it.

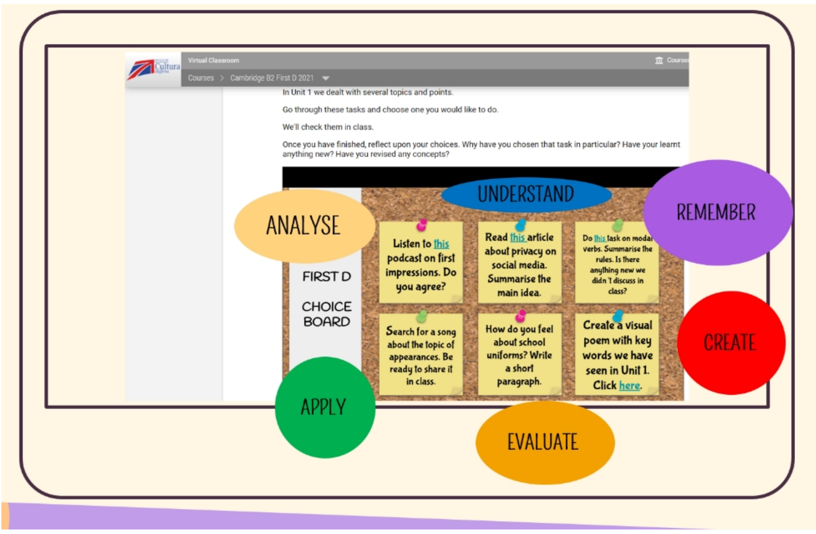
Apart from giving free reign to their imagination and unleashing students' creative potential, the analysis of the memes created reveals that students succeeded in grasping the grammar rule concerning the use of the pronoun *that* in non-defining relative clauses. At the same time, the teacher could get information about how apprehensive the students might feel when it comes to solving a multiple-choice task with a focus on pronouns. Humour played an important role here because it helped the students express their emotions freely. Sharing the memes with the rest of the group can also help clarify the concepts and develop rapport and a sense of team-building with the group. All in all, using this exit ticket in class allowed for the implementation of the second and third strategies proposed by Wiliam and Thompson (2008), engineering effective classroom discussions, tasks and activities that elicit evidence of learning, and activating students as learning resources for one another, respectively.

Choice Board

The other strategy used as formative assessment in the B1+/B2 course was Choice Boards. In essence a choice board is a graphic organizer displayed in boxes or columns but it can also take other shapes. It includes options for a variety of activities that students can choose from to either work through a concept and deepen learning or to demonstrate their understanding of a concept. Figure 7 is an example of a choice board created for the students.

Figure 7

Choice Board



This Choice Board was assigned at the end of a unit about personality and appearance. Students were asked to go through the available options and choose one to work with. Once again, the activities were aligned with the learning objectives described in Bloom's taxonomy.

The possible tasks included working on a grammar point (remember), reading an authentic article on one of the topics dealt with in class and summarizing it (understand), searching for a song related to one of the vocabulary topics and explaining the connections (apply), listening to a podcast on one of the subjects developed and reacting to it (analyse), giving their opinion and reacting to a prompt (evaluate), and creating a visual poem with key words from the unit (create).

Using choice boards turned out to be an effective strategy in the course. First of all, they are good reminders that learners learn in different ways and that task variety should be provided. Secondly, choice boards can be easily adapted to different teaching contexts (different proficiency level, age, etc.) by adjusting the complexity of tasks. The fact that students can self-select what activities to do gives teachers information about which learning preferences they have, how they feel about doing certain tasks and what they can actually do in the task.

Choice boards are goal-oriented as the tasks should be designed with a specific purpose in mind. Another benefit is that they boost engagement and develop autonomy since in the very act of choosing what task to do, learners become the owners of their own learning journey. Choice boards are also relevant as using them allows teachers to cater to different students' needs and learning styles. This is related to Wiliam and Thompson's (2008) fifth strategy, activating students as owners of their own learning. Self-regulation and metacognition are key to the development of learner's autonomy and a resource like Choice Boards can contribute to developing such skills. Metacognition involves two aspects of learning and thinking. First it involves knowledge and awareness of one's own learning and thinking and second it is the ability to plan, monitor, assess and regulate one's own learning and thinking. Giving the students the possibility of choosing the task to carry out compels them first to ponder on their own knowledge. Which tasks do they feel they will be able to do? Which seem challenging but at the same time motivating? Which will demand the revision of concepts dealt with in class? Students might ask themselves questions like these before choosing a task and in so doing they are reflecting on their own learning. Ultimately, metacognition is all about creating the best possible opportunity for students to learn what they need to learn.

Conclusions

This paper has focused on different ways in which formative assessment strategies can be implemented in an EFL class by using Exit Tickets and Choice Board tasks aligned to Bloom's taxonomy of learning objectives. An analysis of the tasks shows that they allow learning to become visible and tangible to both teachers and students. This, in turn, can help teachers make informed decisions to move ahead, re-teach or recycle course content. Formative assessment can also have a direct impact on the learners' and their feelings and attitudes towards their learning experience. Knowing what needs to be learnt and being aware of what has already been learnt fosters students' empowerment, agency and ownership, skills which will definitely facilitate language development.

References

Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of educational objectives*. Longman.

Black, P. J., Harrison, C., Lee, C., Marshall, B, & Wiliam, D (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, *86*(1), 8- 21.

Ebersbach, M., Feierabend, M., & Nazari, K.B.B. (2020). Comparing the effects of generating questions, testing, and restudying on students' long-term recall in university learning. *Applied Cognitive Psychology 34*, 724– 736. https://doi.org/10.1002/acp.3639

Greensetin, L. (2010). *What teachers really need to know about formative assessment*. ASCD.

Wiliam, D., & Thompson, M. (2008). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of* *assessment: Shaping teaching and learning* (pp. 53-82). Routledge. https://doi.org/10.4324/9781315086545-3

Assessment in Argentina's Centralized Secondary Education: Past Perfect, Future Indefinite?

Mario Luis López-Barrios

Facultad de Lenguas, Universidad Nacional de Córdoba

In this paper I intend to investigate assessment practices of past periods of Argentina's history, with a special focus on the assessment practices leading to the promotion of pupils. The question proposed in the title points to a widespread belief that secondary education of past times was superior, and that a perceived relaxation of the role of assessment is to be held accountable for the alleged decline of educational standards in present-day formal education.

Argentina's system of secondary education, i.e. provided to pupils between the ages of 12 and 18 as a follow-up of primary school, was established in 1863 under the organization of the national Ministry of Education and ended in 1992 when the Transfer of Educational Services Act (Ley 24.049) handed their control to the provinces. By the time of the creation of the national state-run system only a few secondary schools existed, notably those in Buenos Aires (Colegio de Ciencias Morales), Concepción del Uruguay (Colegio del Uruguay) and Córdoba (Colegio de Monserrat) (Mantovani, 2012). These schools were named *Colegios Nacionales*, pointing to the centralized nature of the system. The rationale behind the creation of national schools was the provision of high quality post-elementary education to create an elite that would provide the state administration with the required staff to guarantee its continuity. Consequently, pupils received an education that was firmly founded in humanism and positivism, including the study of at least two modern foreign languages as well as classical languages. Secondary education was not compulsory in the time frame discussed and was mostly accessible to middle- and high-class families until social mobility, notably between the 1940s and 1970s, prompted the expansion of the system in terms of students and schools. National schools were all boys' schools until the 1920s, when secondary schools for girls (*Liceos de señoritas*) started to be created, or national schools became coeducational.

As the main aim of national schools was to prepare for higher education, other types of secondary schools were started aiming at the training of primary school teachers (*Escuelas Normales*), vocational training (*Escuelas Industriales*) and a commercial orientation (*Escuelas Nacionales de Comercio*), thus opening up the range of future career and employment options (Mantovani, 2012; Southwell, 2011). In this paper I will concentrate exclusively on the Colegios Nacionales and Liceos de Señoritas.

There is a paucity of studies regarding the role of assessment in foreign language education in Argentina's school sector. In her study about foreign language learning and the creation of a national identity, Vila (2009) makes some references to the assessment of French in Argentinean secondary education in the 19th century, where she briefly describes typical exam questions and exam regulations. To my knowledge, no other studies have explored this aspect of the EFL curriculum in Argentina so far.

Consequently, this paper aims to contribute to our understanding of assessment practices leading to course promotion of secondary school pupils that sought to modify the status quo at three points in the past. To this end, I will focus on the information available in official documents like examination regulations, curricula and syllabuses, as well as reports of educational authorities and other sources to disentangle a possible influence of those measures on today's practices.

The Educational Context

Between 1863 and 1992 a multiplicity of curricula was implemented, so that, on average, between 1863 and 1916 a new plan replaced a previous one every three years (Southwell, 2011). Apart from these curricula and their amendments, conditions for course promotion, grade assignment policies and exam regulations were enforced by executive order until the early 1980s and by ministerial resolutions from December 1983 onwards, when democracy was restored after seven years of military dictatorship. The last curriculum, introduced in 1953, underwent slight modifications at different points until the end of centralized secondary education in 1992. The most significant changes concerned the role, aims and contents of citizenship education, mirroring the changing fate of the country, with democratically elected governments deposed by dictatorships in quick succession between the 1930s and 1983.

Details about the characteristics of formal evaluation were usually contained in the regulations for the assessment, grading and promotion of pupils (*reglamentos de calificaciones, exámenes y promociones*), or they were included as instructions in the curriculum. Some of the regulations stated specific instructions concerning ways to implement the exams and also regarding the class and home activities that could be considered as part of ongoing assessment. On occasion, these made specific references concerning the foreign languages, for example, the expected quality of a translation into the L1 (Régimen, 1926), the consideration of written activities done in class as part of the assessment (Reglamento, 1936), or the types of activities to include in term tests (Secretaría de Estado de Cultura y Educación, 1968).

Types of assessment

In this contribution I aim at describing some assessment practices in secondary education at three points in time: 1917, 1941 and 1986. As stated above, official instructions laid out the administrative procedures involved in summative assessment.

In terms of contemporary conceptions of foreign language assessment, Fulcher (2010) describes the architecture of test construction as a three-dimensional structure consisting of a model, a framework and specifications. The model of foreign language use is a theory or a set of assumptions of what it means to be proficient in a foreign language, whereas the framework stipulates the relevant contents and objectives drawn from the model that will be the focus of the assessment. Specifications describe the actual form of the assessment so that the test can designed (Fulcher, 2017).

Assessment is characterized as formal and informal. The first "involve[s] the systematic recording of results contributing to the award of scores and grades" (Green, 2014, p. 233) and its outcomes, generally in combination with other sources of evidence of learning, determine course promotion. The second consists in the collection of learning evidence, both as in-class and out-of-class activities that are not carried out under test conditions. A further distinction is that between formative and summative assessment. The former "guides the teaching and learning process" (Green, 2014, p. 233), whereas the latter focuses on measuring what has been learned. But the aim of assigning grades and thus determining course promotion, which is a main focus of this paper, is fulfilled by summative assessment.

Lastly, different types of assessment are used for specific purposes. In this study, the most relevant is the achievement test, the "assessment of the knowledge, skills or abilities developed following a period of study" (Green, 2014, p. 228).

The documents from which the information for this study was drawn mention different types of tests, *pruebas* and *exámenes*, used with specific aims. These can be distinguished as progress and exit tests (Green, 2014). Progress tests determine learning achievement during the term and are both summative and formative: summative, because they may be a mandatory requirement for course promotion, and formative, because of their potential to contribute to the learning process, especially when teachers use test results to design consolidation activities. In Argentina, exit tests are conducted after the closure of the last term, with the aim of assessing pupils who obtained a lower grade than that required for course promotion.

Methodology

For this exploratory study I have employed qualitative content analysis (Dörnyei, 2007), a type of textual analysis that relies on the tracking of relevant information to disclose relationships and create categories derived from the data. The exploratory nature of the study is related to the paucity of previous research looking into the official regulation of assessment practices in foreign language education in Argentina.

The corpus consisted in the curricula of Colegios Nacionales, the syllabuses as well as the regulations for the assessment, grading and promotion of pupils issued by the national Ministry of Education between 1863 and 1992. Given the considerable size of the corpus, I decided to narrow down the scope of the study to a few cases coinciding with the curricula issued in 1912, 1941 and 1953, and their related syllabuses and regulations. Additionally, I consulted the annual reports (*Memoria del Ministerio de Educación*) and the monthly bulletins (*Boletín del Minsterio de Educación*) of the Ministry of Education.

The search terms *lenguas extranjeras*, *exámenes* and *promoción* allowed me to single out the most relevant sources and to systematize the information found into categories. Through a process of constant comparison, I revisited the data and refined categories until the data were saturated, which implied the broadening of categories and the dismissal of irrelevant information until all data fit to satisfaction.

Three snapshots

In this section I will concentrate on three policy reforms that sought to make a significant change to the assessment practices and the requirements for the promotion of pupils of national state-run secondary education in the years 1917, 1941 and 1986.

For each of these three cases I will proceed as follows: first, I will describe the prevailing assessment practices and the model of foreign language knowledge of the historical period at hand as expressed in the documents analysed. Secondly, I will describe the reform of the assessment practices and promotion policies introduced and the reasons for their repeal expressed by the educational authorities. Whenever applicable, the assessment framework and the specifications for test construction will be considered.

Snapshot 1: Pendulum Swift

The period starts with the presidency of Hipólito Yrigoyen and ends with that of his follower, Marcelo T. de Alvear. Both belonged to the centre party *Unión Cívica Radical*, albeit with the former representing a more progressive, and the latter a more conservative orientation. Yrigoyen's predecessor, de la Plaza, had ended a long period of conservative administrations that was characterised by a firm support of free, non-religious and universal education.

When Yrigoyen assumed the presidency in October 1916, course promotion at national schools was determined by grades given for performance in reaction to impromptu questioning during class time (*notas diarias*), bi-monthly progress tests and a final oral exam (Plan de estudios, 1912). This heavy concentration on formal assessment was replaced by promotion requirements that relied on fewer instances of formal assessment.

In the time period at hand, the curriculum in force was that of 1912. It comprised six years, in which English was taught for three years in four weekly lessons. The model of foreign language use that can be inferred from the curriculum is that advocated by the Direct Method, which had first been indicated in the previous curriculum (Plan de estudios, 1905). This method was characterized by an emphasis on the oral use of the target language, and the reduced role of translation and of intensive teaching of grammar for the purpose of language analysis (Richards & Rodgers, 2001). The curriculum of 1912 stated that the main aim of foreign language education was to enable learners to read easy texts (*obras sencillas y corrientes*) and that grammar was subordinate to this end. Translation and a more prominent role of grammar were indicated for the second and third courses, together with reading, speaking and writing in simple, mostly text-based activities. Because the official instructions do not make any specific regulations for the assessment of foreign languages, neither the test framework nor the specifications can be inferred. Thus, the assessment reform that took place at that point in time will be described in general instead of from the viewpoint of foreign language education.

The regulations on grading, exams and course promotion of 1917 (Reglamento, 1917) stated that pupils could be dispensed of the oral final exam if they obtained an average grade of 7 out of 10 marks as the result of grades awarded for regular classwork and a written exam at the end of the first term. This reform relieved pupils and teachers of two stressing and work-intensive instances of formal assessment: the bi-monthly tests (*exámenes o pruebas bimestrales*) and the final oral exam. The shift was in line with educational concepts that value formative assessment with a focus on continuous assessment by collecting in-class samples of learners' work. According to Darling-Hammond et al (2020), these kinds of assessment "have been found to create higher motivation, greater agency, and higher levels of cognitive engagement, as well as stronger achievement gains" (p. 121) whereas the formal assessment practices prior to the reform were characterised by pupil anxiety caused by impromptu questioning and frequent exams. These practices may have led to "students' decreased interest in school, distancing from the learning environment, and a lowered sense of self-confidence and personal efficacy" (Darling-Hammond et al., 2020, p. 121).

The reform could not be fully implemented due to the epidemics of 1918 and 1921 (Memoria, 1918, 1921), so that it is difficult to assess the degree to which the educational community agreed to the reform. In the yearly reports of the ministry of education, comments from a handful of school principals bemoaned that the limitation of one written exam in mid-year and promotion based on continuous assessment promoted laziness, as good pupils could obtain high enough grades to attain promotion without the strain of a second term test. Additionally, some of these principals complained that 6.51 points out of 10 as average grade was too low a standard for course promotion. In fact, of a total of 44 national schools by 1922 (Departamento de Estadística Educativa, 1963), only three such views were expressed by their principals (Memoria 1918, 1920), so that they cannot be regarded as representative.

The new regulations (Reglamento, 1923) ended the practice of granting promotion without a final exam with the argument that it damaged pupils' discipline. Though it considered classroom-based assessment, a final written exam was established in every subject and the grade to obtain course promotion without an end-of-year oral exam was raised to 8 points out of 10. Thus, the decision of course promotion by examination committees re-established a higher degree of state control, by taking it away from the class teachers. Through these measures, pupils were expected to be discouraged from making less effort in non-preferred subjects, as a low passing grade in one subject could not be averaged with higher grades in learners' preferred courses.

Snapshot 2: The Exam That Never Was

The curriculum of 1941 (Decreto 101107) had divided secondary education into two stages: lower (1st, 2nd and 3rd year) and upper (4th and 5th year), including a summative exam named *examen de capacidad* to test the extent of the pupils' achievement at the end of the study of every subject or group of related subjects. This had been attempted only for the foreign languages in the curriculum of 1901 (Plan de estudios, 1901) but was given up shortly afterwards without being implemented (Decreto, 30 de octubre de 1901). In foreign language education, pupils were to demonstrate their knowledge of the language studied (English or French) after three years of learning in the lower stage. The model of foreign language use was still that advocated by the Direct Method in the first two years of learning, with the addition of translation of target-language literary texts into the mother tongue and of simple texts from the mother tongue into the foreign language (Programa, 1941, p. 114). The exam regulations established the framework for the assessment, described as an oral exam with a duration of approximately 10 minutes, in which pupils would have to evince their ability to read a text aloud, answer questions on it asked by the teacher and to write phrases – presumably dictated – on the board (Reglamento, 1941). Additionally, the exam syllabus (Inspección General de Enseñanza, 1944) laid out the specifications for the test. Some discrepancies between both are noticeable, namely that the specifications also included translation in both senses and a demonstration of grammar knowledge based on the reading text.

This reform, introduced during the presidency of Ramón S. Castillo of the Conservative Party, came into operation during 1942, so that the examen de capacidad would be first implemented in 1944. In 1943, a year after its inception, Castillo had been overthrown by a military coup. General Edelmiro J. Farrell assumed as de facto president at the beginning of the school year of 1944, when the exam was to be implemented for the first time, but the abrupt changes ensuing the coup resulted in a neglect of its planning and organization. Consequently, an executive order (Decreto 22613) was released stipulating that only the contents of the third year would be tested, even though the exam syllabus had already been published (Inspección General de Enseñanza, 1944). Furthermore, a day after the order was issued, Inspector Juan A. Madueño, in charge of foreign language education, addressed a large group of principals and teachers of English and French to instruct them on the characteristics of the exam, which he characterised as consisting of both an oral and a written section, thus adding evidence to the discrepancy mentioned above. (Madueño, 1944).

In summary, this exam was never implemented for the purpose and in the way it had been conceived and was officially cancelled by an executive order (Decreto 33313) that reinstated yearly exit exams as from the following school year. Several reasons accounted for this failed reform. First, the change in evaluation policy may have been felt as abrupt and alien to the promotion and assessment practices that had been in operation until then, creating anxiety and insecurity in all concerned: pupils, teachers, families, and principals. Second, a degree of improvisation around the exam, evinced in lack of preparation, contradictory orders and last-minute decisions conspired against its viability. In conclusion, non-consensual reforms imposed from the top instead of agreed by the educational community are doomed to failure. This is supported by a view expressed in the annual report of the Ministry of Education stating that the modifications introduced to the exam were rooted in a generalised view of its inadaptability to the school context (Memoria, 1944).

Snapshot 3: No More Numbers

The third reform was meant to introduce a deep change in the assessment practices whose washback effect was expected to influence teaching and learning as a goal-driven endeavour centred around the definition of learning objectives and their attainment. In line with this practice, the numerical grading system was replaced by a conceptual one describing the degree of achievement.

This reform was prompted during the presidency of Raúl Alfonsín, a representative of the progressive quarters of the Unión Cívica Radical, at a time when Argentina had recovered democracy after a seven-year military dictatorship in force between March 1976 and December 1983. When Alfonsín assumed the government in December 1983, secondary school pupils were assessed continuously through evidence collected from class and homework and a written test in each of the four bi-monthly terms (Reglamento, 1975). Between 1968 and 1976 pupils were promoted without a final exam provided they obtained an average grade of at least 6 points of 10 in each bi-monthly term, and of at least 4 points in the last term. This requirement was modified by the de facto regime when the average grade for promotion was raised to a minimum of 7 points with the alleged aim of "stimulating and instilling in pupils a higher concern for their academic training" (Decreto, 18 de abril 1977, p 9, my translation). This measure was based on arguments that resemble those used to repeal the reform of promotion requirements that had been enforced during Yrigoyen's presidency described earlier.

In 1986, when the reform was enacted, the curriculum was that of 1953 (Plan de estudios, 1953). In it, two foreign languages were taught: English or French in the lower cycle (1st, 2nd and 3rd year), and a different one (English, French or Italian) in the upper cycle (4th and 5th year). In the first cycle, the foreign language was taught at a weekly rate of three forty-minute lessons, whereas in the upper cycle a different foreign language from that taken in the first cycle was taught in four forty-minute lessons weekly. The English syllabus, reformed in 1969 (Secretaría de Estado de Cultura y Educación, 1969), stated the aim as the acquisition of the fundamental structures of the foreign language to enable effective communication, whereas the specific objectives mentioned the development of the four skills, suggesting a model of foreign language use following a communicative orientation. Furthermore, the syllabus specified the use of question-and-answer drills and substitution and transformation exercises to apply the learned structures in real situations. Specifications for test construction were stated in the curriculum.

The reform was rooted in a concept of learning as a dynamic, vital and continuous process of participation involving the articulation of the individual and the social aspects of the pupil's personality (Régimen, 1986). It aimed at a learner-centred approach to education to ensure pupil retention and to produce a significant change in teaching and learning by providing pupils opportunities to construct, consolidate and finally achieve the learning objectives. The new exam regulations (Régimen, 1986) did not make any specific references to foreign-language assessments, so that the reform will be discussed in general terms.

In this reform, the school year was divided into two quarters, each one conceived as a structure composed of two stages. The first and the second stages were made up of two components: the first comprised a diagnostic period followed by the development of the learning units, whereas the second aimed at the integration and consolidation of the learning experiences, with opportunities to catch up on contents and further develop skills. Learners who did not attain course objectives and could thus not yet obtain promotion, were offered a new chance to do so after the second stage. This *período de recuperación final* lasted between two and three weeks, as required by each class group. A further opportunity to complete the process and demonstrate the attainment of the course objectives by means of an exit exam was provided after the summer recess, before the beginning of the new school year (Régimen, 1986).

A new grading system was introduced to replace the traditional numerical scores by descriptive statements indicating the degree of attainment of the learning objectives: *superó*, meaning that the pupil had achieved over the expected level, *alcanzó satisfactoriamente / muy satisfactoriamente*, when the objective was attained satisfactorily or very satisfactorily and *no alcanzó*, indicated the pupil had not attained the objective (Resolución Ministerial 332, 1989). Moreover, a further new feature of these regulations was the instruction that teachers should encourage pupils to self-evaluate their own performance to promote their active and responsible participation in the assessment process.

This reform ended with the new regulations for the promotion of pupils established by the government of Alfonsín's follower, Carlos Menem, in 1992. One of the reasons expressed in the new ministerial order (Resolución Ministerial 94, 1992) was the need to ensure a timely and suitable communication of pupil achievement to all the parties involved. This can be interpreted as a measure to restore tighter control and pressure on pupils to demonstrate learning gains. Accordingly, the numerical grading system was reintroduced, together with the requirement of obtaining an average of 6 points out of 10 to be promoted. Furthermore, the two-stage structure of the school year was replaced by a division into three terms. Even if one of the preliminary considerations of the reform reaffirmed the nature of learning as a process, the reintroduction of a school year structure consisting of smaller units did not reflect the same conception of learning embodied in the reform of 1986. This difference is further evinced in the place devoted to the description and functioning of the teaching and learning model in the regulations of 1986 and 1992. In the first, the organisation of the school year is presented as the first chapter, followed by the regulations of the assessment system, whereas in the second, the order is the reverse. Clearly, the reform discussed in this section placed the learning process at the centre, with its assessment as a part of it, whilst the reform that repealed it placed assessment first.

In brief, another opportunity to bring about a positive qualitative change was dismissed too early to know its impact in the long run. Yet again, old practices were reintroduced in replacement of more progressive measures. Nevertheless, pupils who had not obtained promotion were assessed by their teachers after the third term only in those areas of the syllabus they had unsatisfactorily achieved, thus keeping to the idea of "catching up" that featured in the reform of 1986, albeit without the instance of teacher guidance and support included in that proposal. This new assessment practice put an end to the December final exam before an examination board that had been a feature of secondary school summative assessment up until the end of the dictatorship.

Conclusions and Implications

This study is intended as a contribution to the historiography of ELT in Argentina, and it seeks to raise awareness of the influence of evaluative practices of the past on present-day assessment and requirements for promotion of pupils in secondary-school education. Knowledge of past practices could help understand their pervasiveness in the present.

It seems that assessment and promotion regulations before and after the reforms discussed in the three snapshots worked as instruments to curb slackness, which was considered the main reason of low learning standards (Cometto, Arce, Porta, 2019). Thus, assessment was viewed as a disciplinary measure rather than as an instrument to promote positive washback and formative assessment to cause a sustained change in terms of teaching practices and learning outcomes.

A second conclusion is that the three reforms were short lived so that their likely effects in the long run cannot be estimated. These prematurely aborted reforms were followed by a reestablishment of the old and tried: frequent formal assessments instead of a heavier reliance on informal assessment of in and out-of-class work, and numerical grading versus descriptive grading. Such measures indicate a prevalence of behaviourist, competitive approaches to education that were brought back again and again. A school system that aims to contain and provide learning affordances is not compatible with one that places high demands on achievement, leaving many behind in the race.

As a limitation of this study, I recognise that the nature of the reform discussed in snapshot 2 is different from that of snapshots 1 and 3. Snapshot

2 describes a failed attempt to introduce a type of assessment that had not been practiced before, consisting of an end-of-cycle rather than an end-of-year exam, aiming at testing achievement over a three-year period. Without considering the complex nature of the assessment per se and the socio-political context of its implementation and abolishment, a nationwide picture of pupil achievement with a diagnostic instead of a summative purpose could have provided valuable information for educational planning. Snapshots 1 and 3 deal with measures affecting the organisation of the teaching and learning processes and the assessment requirements leading to the promotion of pupils. In both cases, the assumption of pupil underachievement induced by the reform led to their repeal.

A further limitation is the unavailability of testimonials. As said in snapshot 1, the counter-reform was backed by a small number of opinions expressed by school principals, whereas in snapshot 3 the resolution by the Ministry of Education (Resolución 94, 1992) stated that views gathered during debates prompted by the Ministry were considered to introduce the counter-reform, but no details were provided as to whose views were represented. In snapshot 2, only one testimonial was mentioned as support to abolish the exam system.

Despite these limitations, the three cases dealt with allow an insight into the tensions existing in a context that is popularly deemed to be exempt of ideological debates: education.

References

Cometta A. L., Arce, A. & Porta, M. J. (2019). *Notas para pensar la Educación Secundaria en Argentina: una mirada socio-histórica*. Nueva Editorial Universitaria.

Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B. & Osher, D. (2020). Implications for educational practice of the science of learning and development. *Applied Developmental Science*, *24*(2), 97-140, DOI: 10.1080/10888691.2018.1537791.

Decreto 101107, 23 de septiembre de 1941. *Boletín del Ministerio de Justicia e Instrucción Pública, IV*(19), 730-747.

Decreto 22613, 22 de agosto 1944. *Boletín del Ministerio de Justicia e Instrucción Pública*, VII*(58)*, 1112-1114.

Decreto 33313, 11 de diciembre de 1944. *Boletín del Ministerio de Justicia e Instrucción Pública, VII*(58), 1717-1718.

Decreto, 18 de abril 1977. *ADDENDA A Informaciones y Documentos, VIII*(67), 9-10.

Decreto, 30 de octubre de 1901. Ministerio de Justicia e Instrucción Pública (1902). *Memoria del Ministerio de Justicia e Instrucción Pública. Año 1902.* Talleres Tipográficos de la Penitenciaría Nacional.

Departamento de Estadística Educativa (1963). *Enseñanza media años 1914-1963. Tomo 2.* Ministerio de Educación de la República Argentina.

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.

Fulcher, G. (2010). *Practical Language Testing*. Hodder Education.

Fulcher, G. (2017). Standards and frameworks. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 29-44). de Gruyter.

Green, A. (2014). *Exploring Language Assessment and Testing. Language in Action*. Routledge.

Inspección General de Enseñanza (1944). *Inglés. Programa para el examen de 3er año y normas para la recepción del mismo.* Ministerio de Justicia e Instrucción Pública de la República Argentina.

Ley 24.049 Transferencia de Servicios Educativos, 1992. Congreso de la Nación Argentina http://servicios.infoleg.gob.ar/infolegInternet/anexos/0-4999/448/norma.htm

Madueño, J. A. (1944). Qué deberían ser y cómo debiera tomarse los exámenes de idioma extranjero. *Boletín Informativo de la Asociación Argentina de Profesores de Idiomas Extranjeros,* 6, Noviembre de 1944.

Mantovani, J. (2012). *Bachillerato y formación juvenil*. UNIPE: Editorial Universitaria.

Memoria, 1918. Argentina. Ministerio de Justicia e Instrucción Pública (1919). *Memoria del Ministerio de Justicia e Instrucción Pública. Año 1918.* Talleres Gráficos de la Penitenciaría Nacional.

Memoria, 1920. Argentina. Ministerio de Justicia e Instrucción Pública (1921). *Memoria. Año 1920*. Talleres Gráf. Arg. de L. J. Rosso y Cía.

Memoria, 1921. Argentina. Ministerio de Justicia e Instrucción Pública (1922). *Memoria. Años 1921 -1922.* Talleres Gráf. Arg. de L. J. Rosso y Cía.

Memoria, 1944. Argentina. Ministerio de Justicia e Instrucción Pública. (1946). *Memoria: año 1944*. Penitenciaría Nacional.

Plan de estudios, 1901. Argentina. Ministerio de Justicia e Instrucción Pública. (1901). *Memoria: año 1901*. Penitenciaría Nacional.

Plan de estudios, 1905. Argentina. Ministerio de Justicia e Instrucción Pública (1905). *Planes y Programas de los Estudios Secundarios y Normales*. Didot.

Plan de estudios, 1912. Argentina. Ministerio de Justicia e Instrucción Pública (1913). *Plan de estudios. Programas analíticos.* Talleres Gráficos de la Penitenciaria Nacional.

Plan de estudios, 1953. Argentina. Dirección General de Enseñanza Secundaria, Normal, Especial y Superior (1953). *Planes y Programas de Estudio.* Ministerio de Educación de la Nación.

Programa, 1941. Argentina. Dirección General de Enseñanza Media y Normal (1941). *Programas para los cursos del Ciclo Básico Común a los*

Estudios de Bachillerato y Magisterio. Secretaría de Educación de la Nación.

Régimen, 1926. Argentina. Dirección de Instrucción Pública (1926). *Régimen de promociones, clasificaciones y exámenes.* Talleres Gráficos de la Penitenciaria Nacional.

Régimen, 1986. Argentina. Ministerio de Educación y Justicia (1986). *Régimen de evaluación calificación y promoción*. CENDIE.

Reglamento, 1917. Argentina. Ministerio de justicia e instrucción pública (1919). *Reglamento de clasificaciones, exámenes y promociones*. Talleres Gráficos de la Penitenciaria Nacional.

Reglamento, 1923. Argentina. Ministerio de Justicia e Instrucción Pública (1923). *Reglamento de clasificaciones, exámenes y promociones*. Talleres Gráficos de la Penitenciaria Nacional.

Reglamento, 1936. Argentina. Departamento de Instrucción Pública (1936). *Reglamento de clasificaciones, exámenes y promociones*. Antonio Mentruyt.

Reglamento, 1941. Argentina. Departamento de Instrucción Pública (1941). *Reglamento de clasificaciones, exámenes y promociones.* Talleres Gráficos de la Penitenciaria Nacional.

Reglamento, 1975. Argentina. Ministerio de Cultura y Educación (1975). *Régimen de calificaciones, exámenes y promociones para los establecimientos de enseñanza media y su reglamentación*. CENDIE.

Resolución Ministerial 332. 14 de marzo de 1989. http://www.bnm.me.gov.ar/giga1/normas/0474.pdf

Resolución Ministerial 94. 29 de enero de 1992. http://www.bnm.me.gov.ar/giga1/normas/5874.pdf

Richards, J.C. & Rodgers, T. S. (2001). *Approaches and methods in language teaching* (2nd ed.). Cambridge University Press.

Secretaría de Estado de Cultura y Educación (1968). *Disposiciones especiales de la reglamentación general del régimen de calificaciones, exámenes y promociones para los establecimientos de enseñanza media de la Administración Nacional de Educación media y superior*.

Secretaría de Estado de Cultura y Educación. (1969). *Planeamiento de la tarea escolar (a nivel del profesor) y evaluación del rendimiento escolar. 6: Departamento de Lenguas Extranjeras*. Administración Nacional de Educación Media y Superior.

Southwell, M. (2011). La educación secundaria en Argentina. Notas sobre la historia de un formato. In G. Tiramonti (ed.), *Variaciones sobre la forma escolar* (pp. 35-70). Homo Sapiens Ediciones.

Vila, S. (2009). *Enseignements des langues et construction d'une identité nationale: le cas de l'Argentine au XIXème siècle* [Unpublished doctoral dissertation]. Université Paris III- Sorbonne nouvelle.

FAAPI
Federación Argentina de Asociaciones de Profesores de Inglés

LAALTA
Latin American Association for Language Testing and Assessment

ISBN 978-987-98045-2-0