

Project Description

University: Trinity College Dublin

Project (Challenge/Tool/Action): #7: LMS platforms and innovation in English language assessment

1. Introduction:

This challenge responds to the need to ensure that English language assessment uses all the tools available in the language classroom which can assist in measuring language proficiency. This includes virtual learning platforms and applications which ensure the authenticity of language assessment through providing context, choice and multiple forms of language input. Whilst many innovations have been made in pedagogy, creating learner-centred classrooms which deploy task-based curricula, language assessment has often lagged behind, and traditional language tests are often still used even though they may not capture much of a language learner's repertoire or measure the breadth and scope of their communicative proficiency. However, such tests are seen as reliable and easy to use. This action attempts to use some simple interactive formats and tasks (oral presentations, mindmapping and blogging) as contextualised ways of delivering assessment opportunities for teacher, peer and self-assessment, with rating scales that have been constructed using the descriptive apparatus of the Common European Framework of Reference for Languages.

2. State of the art:

Innovation in language assessment shifts what is often the final stage of the the curriculum design cycle (summative assessment) to become an integrated part of teaching and learning activities in the classroom (formative assessment), not just assessing language learning, but also assessing *for* learning. In other words, assessment should be viewed from a learner-centred perspective as much as any other part of the curriculum. However, it can be very difficult for many language learners who have been accustomed to a, teacher-led classroom to adapt to such a shift. As Little (2007, p. 23) notes, "few learners will arrive at their first class ready to take complete charge of their own learning; for most, self-management in learning will be something they have to learn, to begin with by taking very small steps". Providing scaffolding (Wood, Bruner & Ross, 1976; see also Thomsen, 2003 for discussion of scaffolding the language classroom) allows learners to take

control of their own learning through learner involvement in the process of learning, learner reflection and target language use (Little & Ushioda, 1998; Little, 2007). The assessment tools outlined below attempt to respond to this teaching and learning context, and to provide scaffolding throughout the assessment phase of language curriculum design.

For the researcher and test designer, language testing has the aim of providing reliable and valid measures of specific language constructs such as speaking, writing and so forth, and those involved in test design have a dual concern – defining the construct to be measured, and deciding upon the best way of defining or delineating that particular construct. However, this concern is not typically shared by test-takers or those who administer a gatekeeping function based on test results: these end-users generally take scores for granted, and mostly place trust in the validity and reliability of the testing exercise. This is particularly the case for large-scale language proficiency tests which tend to be deployed across the globe for a variety of purposes. However, the life of the test does not begin and end in these scores, but instead it ripples back to language classrooms, where much time is spent in test preparation, and foreword to the school, university or workplace, where decisions are based around test scores often without discussion or understanding of what can be extrapolated from the results, as the newspaper articles cited above help illustrate. Test designers are aware of such effects, and much research has been conducted on the (i) washback and (ii) impact of language tests (Wall 2005; Spolsky 1997; Alderson and Wall 1993).

Washback can be defined as how a test influences the activities which go on in a language classroom before testing, or what Alderson and Wall describe as something that ‘compels teachers and learners to do things they would not necessarily otherwise do because of the test’ (Alderson and Wall 1993, p. 115, their emphasis). Washback can be positive and negative in terms of the types of changes in a curriculum that a test can induce. For example, incorporating spoken interaction into a test is likely to lead to washback in the classroom, and encourage teachers and learners to place more emphasis on the skills involved in dialogue, listening comprehension and creating appropriate spontaneous utterances. But washback can also be detrimental to learners’ competences, by focussing for example on inauthentic aspects of a test or indeed on how to pass the test itself. By creating tests that are authentic, closely related to the target language use domain, and delivered in ways that are clear, fair and transparent, language assessment can create positive washback in the classroom. Our assessment tools aim to create such washback, especially through diversifying forms of feedback beyond just teacher feedback.

Online language learning and assessment in virtual learning environments allows an opportunity for sustained self-assessment which can be scaffolded by the curriculum designers. It has been argued, supported by research findings (e.g., Rea-Dickens, 2006; Goto Butler & Lee, 2006), that self-

assessment embedded in learning, or immediately following, and related to, a particular task (on-task rather than off-task) leads learners to self-assess more accurately. Moreover, detailed and informative descriptions of language competency rather than more abstract concepts of overall progress and effort generally seem to help students to break down the monolith of language learning into manageable skills and subskills, and to identify much more clearly where they have made progress, and to recognise areas of difficulty. In the same way that we learn a language by using the language, we learn most about assessment when we assess, or in the case of the learner, when they are co-assessors and self-assessors.

The approach to language learning and language use outlined in the Common European Framework of Reference will allowed us to implement innovation in online language assessment, to generate a set of shared assessment and goal-setting instruments, scaled to different proficiency levels, and adapted to various assessment needs. A needs-based, learner-centred curriculum is concerned principally with providing structures and mechanisms to help students learn in a way that best suits their interests, priorities and learning styles, and, above all, to operate successfully in the target language both in the classroom, and in a target language context.

In this E-LENGUA challenge, we focus on specific, action-oriented descriptions of language use, joined-up learning and assessment materials, and embedded self-assessment procedures, as we endeavour to help student language learners to become lifelong language users. The “Can do” descriptor approach of the CEFR facilitates the implementation of self-assessment procedures. Even through the small step of asking students to consider their current language ability before enrolment in an online module, self-reflection can be encouraged from the outset.

Through the design of three innovative real-life assessment activities which will test learners’ spoken and written production (an interactive presentation format, mindmapping and Blackboard blog), we attempt to provide viable alternatives which can be embedded in virtual learning environments, used and adapted by teachers, and benchmarked to the Common European Framework of Reference’s proficiency levels.

3. Target groups:

The target group of learners is based on English language learners studying English mostly for purposes of advancing their academic studies, whether at a very basic level, or at the highest level for further education purposes. These learners range from those aiming for A2 level proficiency through to those aiming for C2 level, near-native speaker-level proficiency. Our core target group is modelled on the university students registered at Trinity College in our English language programmes. The university offers a variety of not-for-credit modules tailored to the needs of students whose first language is not English. At present, these modules do not avail of Blackboard in any systematic way, and assessment methods remain rather static and

traditional (reports and in-class oral presentations). In addition to this core group, we are aiming to provide online assessment tools that will be available freely and publicly to the designers of English language programmes anywhere in the world, and adaptable to other virtual learning environments such as Moodle.

4. Methodology:

Three tools were adopted in our efforts to ensure a learner-centred and needs-based assessment framework: (1) an interactive presentation format for spoken production and interaction, (2) a mindmap format for written production and (3) a blog for written interaction. Each of these formats allow authenticity in assessment topics, authentic and meaningful interaction, scaffolding, and a variety of assessment methods (teacher, peer, self). Development of appropriate rating scales of language proficiency are described below under (5).

Assessing Spoken Production Skills in an innovative way which can be deployed for testing learners of all proficiency levels. Our approach to assessing spoken production skills uses the *PechaKucha* approach to testing English Spoken Production Skills (Japanese: ペチャクチャ, translation: 'chit chat'). The PechaKucha approach to presentations employs 20, mostly visual, slides which are shown for 20 seconds each (6 minutes and 40 seconds in total). The format keeps presentations concise and fast-paced, and was first used in Japan in 2003 for multiple presentations by young designers. This approach is increasingly been used in Europe for crowdsourcing, pitches etc. in 'PechaKucha' nights. For the English speaking tests, students are requested to present on any topic of their choice, using this visual slide format as their prompts. The rating scale is provided below under (5). These presentations can be delivered synchronously and asynchronously.

Assessing Written Production Skills through [Popplet](#). The Popplet application provides a platform for the creation of mindmaps and presentations, incorporating visual supports as well as text. It can be used by a variety of proficiency levels, from A1 through to near-native speaker, and provides a single platform through which writing can be assessed from a variety of perspectives, enabling peer assessment through the comments feature as well as teacher assessment. Assessment purposes in an interactive manner entails selecting the topic in advance. The rating scale is provided below under (5).

Assessing Written Production Skills through Blog functions in Blackboard. A Blog is a personal online journal that is updated regularly and is intended for public viewership online. Blackboard Learn provides a function for enrolled users to view and write blogs. This kind of extensive and interactive writing encourages students to express their ideas clearly in an authentic context and in a social environment. They also provide a way of showcasing students' interests and achievements. Blackboard Learn Blogs comprise the Blog entry function (text, images, links, embedding YouTube videos etc.); the Comments function (remarks or responses to Blog entries made by the assessor and peers). Blogging for assessment purposes in an interactive manner entails selecting the topic in advance, allowing only the owner of the Blog to add Blog

entries but allowing all other class members to view and add comments. The rating scale is provided below under (5). Assessors can grade the quality of the discussion, as well as the number of entries and comments that are made by an individual or the group of learners. Grades for blogs are managed in the Blackboard [Grade Center](#). The Blackboard platform offers accessibility to learner groups with visual impairments, and given the range of students' needs, we will ensure that our tools have textual equivalents or alternative text for content, that screen labels are used appropriately, and that any PDFs are appropriately edited for screen readers and accessible from a range of devices.

5. Development of the project:

This aspect of our report provides details of the testing rating scales which have been piloted by small groups of volunteer English language teachers enrolled in the TCD M.Phil. module in Language Testing as part of the validation process. The scales were validated through collaborative discussion, trial ratings of anonymised examples of student work, and through group sorting of the scales by subconstruct and proficiency level (i.e. deconstructed, and reconstructed in groups to see if the reconstruction matched the original.) Further validation is planned in the next phase and outlined in the chronogram, using the English language students enrolled in TCD's summer English programmes.

These rating scales, both holistic and analytic, for written production and spoken production, are designed to be accessible from a Blackboard or Moodle platform and used across a variety of language proficiency levels. The rating scales have been adapted from the descriptors provided by the *Common European Framework of Reference for Languages*, based on an analysis of the target language use domain and the most relevant language constructs.

A1-B1 CEFR benchmarked rating scales, Writing Production (Mindmaps and Blogs)

WRITING	A1	A2	B1
LANGUAGE	uses simple sentence forms uses limited range of basic vocabulary	Uses basic vocabulary reasonably appropriately. Uses simple grammatical forms with some degree of control. Errors may impede meaning at times	Uses everyday vocabulary generally appropriately, while occasionally overusing certain lexis. Uses simple grammatical forms with a good degree of control. While errors are noticeable, meaning can still be determined.
COMMUNICATIVE ACHIEVEMENT	Produces a simple and short text	Produces text that communicates simple ideas in simple ways.	Uses the conventions of the communicative task in generally appropriate ways to communicate straightforward ideas.
ORGANISATION	Text is consists of short phrases	Text is connected using basic, high frequency linking words.	Text is connected and coherent, using basic linking words and a limited number of cohesive devices.

C1-C2 benchmarked Analytic Rating Scale, Written Production (Blogs)

	1	2	3	4
Content	The information provided has little or no relation to the topic.	The information provided has limited knowledge about the topic. Overall, it is lacking essential information.	Includes essential information about the content. Subject knowledge appears to be decent.	All information provided is extremely detailed, relevant and appropriate to the content. The topic is presented thoroughly.
Structure & Organization	The main points of the presentation are presented in a haphazard manner. There is no clear or organized structure.	Content is organized for the most part, but is illogical at times. The structure of the presentation is adequate.	Includes a coherent presentation structure, mostly easy to navigate.	Ideas and information are organized and transitioned well. The structure is very easy to navigate.
Grammar & Vocabulary	Grammar structures are not suitable for academic writing. Many spelling errors and uses basic vocabulary.	Several grammar and spelling mistakes. The complexity of vocabulary is inadequate, but does not hinder the overall comprehensibility.	Minor grammar inaccuracies presented. There is a wide variety use of vocabulary, but it is restricted in places.	Grammar structures and use of vocabulary is presented logically. Contains almost no inaccuracies.
Independence of Thought	There is an absence of independent thought. Student's voice is not apparent within his/her presentation .	Independence of thought is evident throughout the presentation, but it is not developed properly.	Independent thought is expressed appropriately. Clear evidence to support the argument should	The independence of thought is executed clearly and is supported accurately and academically. The student fully grasps the subject matter.

A1-B1 CEFR benchmarked Analytic Rating scales, Spoken Production & Interaction (PechaKucha style)

	A1	A2	B1	B1+
SPEAKING LANGUAGE	only produces isolated words or memorised utterances • cannot produce basic sentence forms	• uses simple vocabulary to convey personal information • has insufficient vocabulary for less familiar topics • attempts basic sentence forms but with limited success, or relies on apparently memorised utterances • makes numerous errors except in memorised expressions	is able to talk about familiar topics but can only convey basic meaning on unfamiliar topics and makes frequent errors in word choice • rarely attempts paraphrase produces basic sentence forms and some correct simple sentences but subordinate structures are rare • errors are frequent and may lead to misunderstanding	manages to talk about familiar and unfamiliar topics but uses vocabulary with limited flexibility • attempts to use paraphrase but with mixed success produces basic sentence forms with reasonable accuracy • uses a limited range of more complex structures, but these usually contain errors and may cause some comprehension problems
DISCOURSE MANAGEMENT	Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.	Can use basic turn taking devices in simple and routine tasks requiring a direct exchange of information, feelings and opinions on familiar and routine matters.	Can use simple turn taking devices . in discussion in a familiar situation making relevant points and responding to reach a shared understanding	Can interact with small degree of fluency and spontaneity that makes regular interaction with native speakers quite possible with little strain for either party.
PRONUNCIATION	Speech is often unintelligible	mispronunciations are frequent and cause some difficulty for the listener	uses a limited range of pronunciation features • attempts to control features but lapses are frequent • mispronunciations are frequent and cause some difficulty for the listener	uses a range of pronunciation features • attempts to control features but lapses are frequent • mispronunciations are frequent and cause some difficulty for the listener
GLOBAL ACHIEVEMENT	• pauses lengthily before most words • little communication possible	speaks with long pauses • has limited ability to link simple sentences • gives only simple responses and is frequently unable to convey basic message	• cannot respond without noticeable pauses and may speak slowly, with frequent repetition and self-correction • links basic sentences but with repetitious use of simple connectives and some breakdowns in coherence	usually maintains flow of speech but uses repetition, self correction and/or slow speech to keep going • may over-use certain connectives and discourse markers • produces simple speech fluently, but more complex communication causes fluency problems

C1-C2 CEFR benchmarked Holistic Rating Scale, Spoken Production & Interaction (PechaKucha style)

C1	<p>Can handle difficult and even hostile questioning fluently and clearly.</p> <p>Can elaborate on complex topics confidently and articulately to an audience unfamiliar with it, structuring and adapting the talk flexibly to meet the audience's needs.</p>
B2	<p>Can depart spontaneously from a prepared text and respond clearly to questions raised by members of the audience, often showing remarkable fluency and ease of expression.</p> <p>When asked can give a clear, detailed explanation of presented topics with highlighting of significant points, and relevant supporting detail.</p>
B1	<p>Can take follow up questions, but may have to ask for repetition if the speech was rapid</p> <p>Can give a straightforward explanation of presented topic which is clear enough to be followed without difficulty most of the time, and in which the main points are explained with reasonable precision.</p>
A2	<p>Can cope with a limited number of straightforward follow up questions if he/she can ask for repetition and if some help with the formulation of his/her reply is possible.</p> <p>Can give a short explanation of presented topic without giving much detail and without a high level of clarity or precision.</p>

6. Chronogram:

The next stage of the project is based on three approaches to assessment described above: PechaKucha, Popplet and Blackboard Blogs, in assessing written and spoken skills. During the first stages of the project, the testing framework has been outlined and the rating scales have been developed and piloted. The next stages are:

January-March 2017

Completing the design phase of all assessment tools (prompts, instruction rubrics, rating scales etc.) and uploaded to VLE. Transforming validated scales into self-assessment and peer-assessment documents.

April-June 2017

Research and development activities including focus groups of learners and instructors, assessment trials under test conditions, collaboration with

Blackboard support team at TCD. Training workshop open to all TCD language instructors in English. Finalising checks regarding use of analytic or holistic rating scales and teacher/learner preferences.

July-August 2017

Final preparations in advance of new cohort of students who will use the tools in their modules from Sept 2017 onwards. Tools reviewed by summer English students taking intensive programme in August.

Sept-December 2017

Investigation of assessment tools in practice from teacher and learner perspectives, administration of evaluation tools, investigating scale validity and reliability (e.g. face validity, construct validity, inter-rater reliability etc).

January-June 2018

Completion of research and development phase of the assessment tools: publication of full assessment framework and specification for use at large. Research analysis completed, gathering together all data (attitudinal and test scores) into final report.

7. Dissemination:

Conference paper: Lorna Carson, Daniela Modrescu and Mona Syrbe, November 2017 : National Conference of the Irish Association for Applied Linguistics (IRAAL) – paper presentation on ‘Innovation in language pedagogy : the E-LENGUA project and English language assessment’

Training workshop on the E-LENGUA tools in June 2017, open to all TCD language instructors in English for Academic Purposes.

Workshop facilitated by Lorna Carson at the Jean Monnet Summer School, 10-14 July 2017, Trinity College Dublin ‘Assessing languages in a multilingual context: Innovative Tools and the E-LENGUA project’.

8. Bibliography:

Alderson, J. C., & Wall, D. (1993). Does washback exist?. *Applied linguistics*, 14(2), 115-129.

Carson, L. (2010). Innovation and Autonomy in an Institution-wide Language Programme in, editor(s) O'Rourke, B. & Carson, L. , *Language learner autonomy: Policy, curriculum, classroom*, Oxford, Peter Lang, 2010, pp151 – 170

Carson, Lorna, *Fostering engagement with the Common European Framework of Reference for Languages and the European Language Portfolio: Learning from good*

practice in university language centres, *Language Learning in Higher Education*, 6, (2), 2016, p275 – 282

Council of Europe (2001). *Common European Framework of Reference for Languages: Teaching, Learning and Assessment*. Strasbourg, Council of Europe.

Little, D. (2007). Language learner autonomy: Some fundamental considerations revisited. *International Journal of Innovation in Language Learning and Teaching*, 1(1), 14-29.

Spolsky, B. (1997). The ethics of gatekeeping tests: what have we learned in a hundred years?. *Language testing*, 14(3), 242-247.

Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory (Vol. 22)*. Cambridge University Press.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of child psychology and psychiatry*, 17(2), 89-100.