



# **European Association for Computer Assisted Language Learning**

---

## ***THE EUROCALL REVIEW***

**Volume 22, Number 2, September 2014**

Editor: Ana Gimeno

ISSN: 1695-2618

<http://www.eurocall-languages.org/publications/review>

---

## Table of Contents

<i>Article: Sustainability in CALL Learning Environments: A Systemic Functional Grammar Approach</i> Peter McDonald.....	3
<i>Article: Lessons Learned in Designing and Implementing a Computer-Adaptive Test for English</i> Jack Burston and Maro Neophytou .....	19
<i>Article: How EFL students can use Google to correct their "untreatable" written errors</i> Luc Geiller.....	26
<i>Article: Constructing an evidence-base for future CALL design with 'engineering power': The need for more basic research and instrumental replication</i> Zöe Handley.....	46
<i>Article: Podcasts for Learning English Pronunciation in Igboland: Students' Experiences and Expectations</i> E.E. Mbah, B.M. Mbah, M.I. Iloene and G. Iloene .....	57
Editorial Board .....	69

## Article:

### ***Sustainability in CALL Learning Environments: A Systemic Functional Grammar Approach***

**Peter McDonald**  
**J.F. Oberlin University, Tokyo (Japan)**

---

[pmcdonal@obirin.ac.jp](mailto:pmcdonal@obirin.ac.jp)

#### **Abstract**

This research aims to define a sustainable resource in Computer-Assisted Language Learning (CALL). In order for a CALL resource to be sustainable it must work within existing educational curricula. This feature is a necessary prerequisite of sustainability because, despite the potential for educational change that digitalization has offered since the nineteen nineties, curricula in traditional educational institutions have not fundamentally changed, even as we move from a pre-digital society towards a digital society. Curricula have failed to incorporate CALL resources because no agreed-upon pedagogical language enables teachers to discuss CALL classroom practices. Systemic Functional Grammar (SFG) can help to provide this language and bridge the gap between the needs of the curriculum and the potentiality of CALL-based resources. This paper will outline how SFG principles can be used to create a pedagogical language for CALL and it will give practical examples of how this language can be used to create sustainable resources in classroom contexts.

**Keywords:** CALL, Multimodality, Systemic Functional Grammar, Sustainability, Curriculum innovation.

#### **1. Introduction**

The rapid development of ubiquitous technologies provides new opportunities for learners to express themselves. In the pre-digital age, learners could only express themselves through written text (the written mode) or spoken text (the oral mode). By contrast, learners in a computer-assisted environment can quickly combine these two modes and even add visuals or sound. These new multi-modal texts are already widely used in courses that use Moodle/Blackboard, computer mediated communications (CMC) such as posts or blogs, and digitally created compositions such as PowerPoint or video presentations.

Nevertheless, new pedagogical opportunities create new pedagogical challenges. Despite the successful additions introduced by computer-assisted language learning (CALL) to established curricula, overall, the curricula that educational establishments deliver to students have fundamentally remained the same. Curricula, to a large extent, are still based on old technologies (paper, pens, and textbooks) and traditional classroom methodologies. At present, classrooms still follow the 19th century industrial model, in which a large group of students sit at separate desks while a teacher delivers a pre-prescribed, traditional curriculum (Collins & Halverson, 2009). Owing to the

scarcity of computer labs in such teaching contexts, CALL is often limited to a set number of classes each term.

Curricula are difficult to change because of “situational constraints” (Cuban, 2001) that educational institutions encounter. Educational institutions have developed complex systems (budget, size, number of stakeholders, inbuilt working practices, and so on), which makes it difficult for them to adapt to change effectively (Kennedy, 2013). Likewise, teachers are constrained. Considering that teachers must successfully meet the needs of all institutional stakeholders, introducing innovation in the classroom may be difficult for them. Teachers must necessarily depend on shared and established educational knowledge.

Such a shared pedagogical knowledge, the foundation upon which curricula are developed, does not exist for CALL. Therefore, without this shared knowledge, CALL resources cannot be integrated into the existing curricula. The current study supports the view (Kress & Van Leeuwen, 2006; Royce, 2002) that systemic functional grammar (SFG) can help provide this missing pedagogical framework as it provides teachers with a multi-modal meta-language that can work alongside existing pedagogical meta-languages. Equipped with a multi-modal pedagogical language, teachers can create “sustainable CALL resources.” A sustainable CALL resource, as defined in this paper, can: 1) work alongside existing classroom resources that have been established by teachers to meet the needs of the curriculum, and 2) help to prepare students for using the new and exciting affordances that digital technology offers.

## **2. Expanding existing pedagogical languages to create sustainable CALL resources**

### *2.1. The role of pedagogical languages in creating sustainable resources in the curricula*

In the pre-digital society dominated by the printed-word pedagogical languages, from the traditional grammar of Latin and Greek to the modern approaches such as Consciousness Raising (C/R), have always played a role in creating classroom resources. For example, in teaching reading or listening, standard classroom activities (textual comparisons, assessments, examination of writers’ textual choices, and so on) are possible because the underlying textual relationships in written/spoken texts can be explicitly expressed. Thus, the various methods of linguistic description that already exist for written/spoken texts, whether traditional grammar or modern communicative approaches such as discourse analysis or pragmatics, assist the teaching of these texts in the classroom.

The pervasiveness of digital media in our society, however, is changing the nature of text, and thus, the established methods of linguistic description that teachers use in the classroom are no longer sufficient for teaching modern texts. In the past, texts predominantly utilized the alphabet to send their messages; by contrast, digital devices of today combine words, visuals, and audio to create multi-modal texts where the traditional linear structure of reading left to write on a page is challenged by a visual hyperactive reading path that follows the rules of visual design, as well as the rules of the written language (Kress, 2003, pp. 35–60).

Furthermore, digitalization is significantly changing the traditional written text itself. In a modern digital society in which the means of production have shifted and writers are now assuming many of the publication tasks that were once considered specialized, written texts that vary from traditional academic essays to modern tweets and blogs incorporate a wide array of visual features, such as bullet points, tables, and emoticons. Another change that the digital revolution has helped to introduce is the pedagogical acceptance of popular media texts that incorporate multi-modal relations, such as comic books, videos, or computer games. These texts, which may not have been considered

useful in the classroom 30 years ago, are now being recommended as essential additions to the modern curricula (Hagood, 2008).

Considering these fundamental changes in the nature of texts, our existing methods of linguistic description must be updated so that teachers can talk about multi-modal texts explicitly in the classroom, in the same way as they can currently talk about traditional printed word or spoken texts in the classroom. Indeed, creating a multi-modal pedagogical language is important because research suggests that multi-modal texts are more complex than has been accounted for in existing classroom approaches.

## *2.2. The complexity of multi-modal texts*

Although multi-modal texts may appear to have simple means of presenting information, the textual relationships underlying such texts may be complex. In order to effectively comprehend a multi-modal text, the reader, viewer, or listener must engage in "parallel processing" (Luke, 2003, p. 399). In this type of processing, the receiver must initially (and perhaps unconsciously) decode different semiotic systems, including the spatial system of design to decode the images and the linear system of the alphabetic text to decode the words, and then interpret how the systems combine to deliver a singular meaning. Moreover, Unsworth (2008, p. 378) reports the effect of "naturalization," in which these complex underlying semiotic relationships can be hidden by multi-modal writers to create cohesive texts. In the context of Teaching English as a Second Language (TESOL), parallel processing and naturalization are extremely complex because learners not only have to process different modes, but also have to translate these modes from their second language (L2) into their first language (L1).

TESOL research has shown that two multi-modal relations, defined in SFG research as concordance and complementation, can produce complex effects on comprehension. Lui's (2004) research on L2 multi-modal comprehension suggests that images only support comprehension when the graphic text clearly reiterates the same information as the written text. In SFG grammar research, this relationship is defined as a graphic/alphabetic text relationship of "concurrency" (Unsworth, 2008). Positive support occurs in concurrent relationships when the students' proficiency level is just below the level of the alphabetic text. In this case, students can use the images to infer the meaning of the words. However, concurrent text relationships may also result in redundancy when the students' proficiency level is above the level of the alphabetic text. In this case, the students do not need the graphic text to infer the meaning of the words (Lui, 2004).

Other negative effects on comprehension can be observed in multi-modal relationships of "complementation" (Unsworth, 2008). In a relationship of complementation, the graphic text and the alphabetic text contain closely related information that augment, rather than reiterate, each other in some way. Relationships of complementation can result in incomprehension or miscomprehension when the students' proficiency is lower than the words in the text. Incomprehension occurs when the lack of textual integration prevents students from using the graphic text to infer the meaning of the words, which renders them unable to understand the text. Miscomprehension occurs when students make the wrong assumption about the graphic/alphabetic text relationship. They assume that the graphic text reiterates the information in the written text; that is, the graphic text can support the words. However, the lack of harmony between the alphabetic and graphic text clues creates difficulties in processing. The students then make wrong inferences about the text. Thus, the graphic text hinders the comprehension of the written text (Lui, 2014).

Therefore, a multi-modal pedagogical language can allow for teachers and students to decode complex semiotic relationships in a meaningful way that can be applied to their

existing teaching contexts, as will be demonstrated in Parts 3 and 4. A multi-modal pedagogical language can therefore help teachers create sustainable CALL resources. This process can be assisted by SFG, given that it is a communicative approach to language learning (Halliday & Matthiessen, 2004) and it can therefore work alongside established classroom approaches to language and learning such as C/R. In C/R language users work with the language in use, making a series of assumptions about the language, rules of thumb, which can be adjusted to suit the needs of the communicative situation (Rutherford, 1987).

### 3. The SFG theoretical model for creating sustainable CALL resources

#### 3.1. SFG and reading image-based multi-modal texts

The Kress and Van Leeuwen (2006) model for analyzing visual texts serves as the theoretical basis of the ideas presented in this paper. This paper aims to demonstrate that the semiotic model of language can be used in the classroom in a practical and simple manner. Therefore, this study will not provide a full account of the model, rather it will only focus on the use of the Kress and Van Leeuwen SFG model to create sustainable resources in classroom contexts, as outlined in Section 4.

##### 3.2.1. The compositional function

In alphabetic texts, Fries (1994, p. 230) points out that the placement of clauses in a written text determines the importance of the information placed within the clause. This concept is also true of images: the placement of elements in an image such as a picture or a web page determines the visual importance of the elements. This study will focus on two compositional elements, namely, framing and salience. Framing refers to the way elements (image elements include words, pictures, hyperlinks, and others) are connected or disconnected through frame lines. Salience refers to the prominence ascribed to one image element over another by varying an image's size, color, contrast, and by choosing to place elements at the top, bottom, center, or margins of a picture.

##### 3.2.2. The representational function: narrative images versus concept images

An image can represent two things to the viewer, a narrative event or a concept. Artists create narrative events in images by joining the participants (people, animals, objects, and so on) together with an imaginary line called a 'vector' (Kress & Van Leeuwen, 2006, p. 59). Figure 1 shows an excerpt from *Macbeth: The Graphic Novel* (McDonald, Haward, Dobbin & Erskine, 2008, p.8), where panels 1 and 5 are narrative images. The fire is the vector. The witches' attention is focused on the fire, and the fire is connected to the witches by framing, salience, and color. This communicates to the reader that the main action of the image is centered on the witches and the fire. In concept images, the participants are not represented in action; that is, no vector joins them. By contrast, the participants are represented in a fixed state of being, such as a portrait painting. In Figure 1, panels 2, 3, and 4 are all concept pictures. In these panels, the witches' faces are given salience through a close-up view staring in the direction of the viewer, as in a portrait.

##### 3.2.3. The interpersonal function: offer images versus demand

Images can interact with viewers in two ways: by offering information or by demanding attention. A simple way to understand the difference between offer and demand is to make a comparison between teaching a lesson in front of a group of students with having a face-to-face conversation with one student. In the classroom setting, the speaker (the teacher) is offering information to the class; the speaker is at a distance from the class; and the students can choose between listening to the speaker and thinking about other things that are unrelated to the lesson. By contrast, a face-to-face

situation requires that the participants demand attention from one another; that is, they need to directly focus on what is being said.

Referring again to the series of Macbeth panels, we see that Panel 1 and 5 are offering information to the viewer. The viewer is asked to observe the scene from a distance, as well as choose their own reading path: the viewer can begin with the text boxes, the background image, or the main image of the witches cooking a spell in a cauldron. Alternatively, Panel 2, 3, and 4 demand attention from the viewer. They are confronted with a talking head image and are asked to focus directly on the words.



Figure 1: Macbeth panels (Used with permission © Classical Comics Ltd.).

### 3.3. Classifying images into types

Once teachers and students have a working knowledge of these underlying principles, they can begin to classify images into types by asking a series of C/R questions, as provided in Table 1 below. Thus, applying the questions to the Macbeth texts, we see that Panel 1 and 4 are narrative/offer pictures, in which the illustrators/writers ask viewers to observe events. Meanwhile, Panel 2, 3, and 4 are demand/concept pictures. The illustrators present an idea, not an action, and demand an emotional response from the viewer. However, as mentioned above, multi-modal text relations can create very complex texts. Therefore, as suggested by the underlying principles of C/R, teachers would be dealing with rules of thumb, rather than proscribed rules, when applying these theories to the classroom.

<p><b>The Compositional Function</b></p> <ol style="list-style-type: none"> <li>1. Which elements are most salient in the image? (How is this salience created? Is it created through placement, color, or contrast?)</li> <li>2. How are the elements of the image framed? (Are the participants/elements joined together?)</li> </ol>
<p><b>The Narrative Function</b></p> <ol style="list-style-type: none"> <li>1. Is the image representing a narrative? (Does the image portray an event? Does it have a vector?)</li> <li>2. Is the image representing a concept? (Are the participants not joined in action together? Are they staring at the viewer or into the distance? Does the image portray an idea rather than an event?)</li> </ol>
<p><b>The Interactive Function</b></p> <ol style="list-style-type: none"> <li>1. Is the image interacting with the viewer by <i>offering</i> information to the viewer?</li> <li>2. Is the picture interacting with the viewer by <i>demanding</i> attention from the viewer?</li> </ol>

Table 1. C/R questions for classifying image-based texts.

#### 4. Creating sustainable CALL resources: applying the SFG grammar model to the classroom

##### 4.1. Sustainable CALL resource 1: converting written compositions to multi-modal compositions

This multi-modal task is relevant to beginner writing classes in which students are taught how to write five-paragraph essays using the meta-language of topic sentences, supporting sentences, concluding sentences, and the function of each type of sentence. For example, topic sentences represent general ideas whereas supporting sentences provide details about the topic using examples or explanations. In the established curricula activities in my teaching institution, students use such classifications to create their own original paragraphs and deconstruct teacher-created paragraphs. Figure 1 (Appendix) shows a comparative paragraph, which is a teacher-created example of a written text that students are expected to follow. Figure 2 (Appendix 1) shows the deconstructed paragraph using the traditional classroom meta-language, which is a common classroom activity.

In the multi-modal activity, students convert the written paragraph to a multi-modal presentation using presentation software. Students re-read their paragraphs, select the key words that communicate the overall ideas, perform a Google image search to find a supporting image that helps to visually communicate the main ideas to the listener, and rewrite the sentences to adapt them to the spoken mode, if necessary. A teacher-created example of a presentation is provided in Figure 4 of the Appendix.

As shown in the examples, the SFG model's classification of images into types, using the compositional, representational, and interactive functions, provides teachers with the pedagogical tools to outline clear organizational patterns for students when they construct their classroom presentations. In this example, students are encouraged to follow a general-specific textual pattern. Thus, the information represented by the topic sentence and the concluding sentence (Appendix, Figures 1 and 2, sentences 1 and 5) that comprise the general ideas in the written paragraph are represented by concept/offer images (Appendix, Figures 3 and 4, slides 1 and 5), whereas supporting sentences that contain examples (Appendix 1, Figures 1 and 2, sentences 2, 3, and 4) are all represented by concept/demand images (Appendix 1, Figures 3 and 4, slides 2, 3, and 4). Moreover, knowledge of visual/verbal text relations of concurrence and complementation enables teachers to provide clear advice to students. In this L2 setting students are encouraged to use simple slides with strong relations of concurrence and



simple relationships of complementation to facilitate comprehension (Figure 3 column 4).

Therefore, the SFG model creates sustainability because new multi-modal skills can be taught alongside the established curricula. For example, naturalization, identified in Section 2 as a multi-modal skill, is already being taught in writing curricula through textual patterns. Research shows that deconstructing textual patterns, such as cause-effect, compare-contrast, problem-solution, has a positive effect on language learning (Hoey, 2001). Likewise, teaching the way texts are designed across different genres is now part of our established writing curricula (De Voss, 2010) and, interestingly, Computer-Mediated Communication (CMC) is now being defined as a separate genre that can be taught in the classroom (Marchand, 2013). The SFG model provides teachers and students with a multi-modal framework that can be used to unpack and teach naturalized components that comprise multi-modal texts in the same manner that they can unpack alphabetic texts.

From a visual perspective, unpacking the text in this manner makes it possible to introduce students to 'image juxtaposition' (McCloud, 1994). Image juxtaposition involves combining different types of images together in a sequence to create meaning. The process is another example of naturalization that is effectively utilized in multi-modal texts, but is often overlooked by the untrained eye. In this task, students can see how images can be juxtaposed in way that creates a clear textual pattern, as outlined above. Similar multi-modal conversion tasks can easily be created for the other textual patterns that are taught in the curricula. For example, non-computerized texts such as the Macbeth text (Figure 1) can be used to introduce students to image juxtaposition in traditional classroom settings. In this narrative text, the writers/illustrators use narrative/offer images to set the scene and portray the event (the spell being cast) juxtaposed with three concept/demand pictures, not only to attach the reader to the witches who are casting the spell emotionally, but also to provide salience to the spoken text that cataphorically points the reader to a key event (their future meeting with Macbeth).

Finally, converting the written text to a multi-modal text reinforces established curricula skills, such as peer review, revision, and rewriting, in a creative and engaging manner. Hence, students must reflect on their individual writing, discuss it with peers, and evaluate its meaning, clarity, and communicative competence. Students must re-read, edit, rewrite, and summarize as they convert the written mode to the visual/verbal mode, which is then converted back to the written mode. Moreover, learner autonomy is encouraged through the navigation of presentation software and the use of search engines in English.

#### *4.2. Sustainable CALL resource 2: making multi-modal comparisons*

This multi-modal task is designed for reading and writing curricula in which students learn how to work with different genres from different discourse communities. In this task, students perform the standard curriculum task of making textual comparisons between two texts of different registers, which is a task that is usually done with traditional written texts. The SFG model, however, as outlined above, allows for the expansion of the standard curricula comparison activity to include multi-modal texts.

The task is composed of two parts. In part one, students compare and contrast the homepages of BBC (traditionally considered as a politically neutral creator of serious news) with the Daily Mail (traditionally considered as a creator of popular, right-leaning news). In part two, students create a multi-modal report. Selfe (2004) has many examples of different types of multi-modal reports that could be used. To perform the

task, students use the multi-modal C/R questions in Table 1 combined with the multi-modal C/R questions in Table 2.

#### Homepage Composition

1. Which elements on the home pages are most salient?
2. How do the sites use color, framing, and placement of images —top to bottom, left to right?
3. What kind of participants occupies the most salient positions on the page?

#### Headlines and Images

4. What types of images are used on the homepage?
5. What kind of language is used in the headline text?
6. Can you identify any relationships of concurrence or complementation between the headline texts and the image texts? If so, why do you think the writers created this relationship?

#### Body Texts, Headlines and Images

7. Can you identify any relationships of concurrence and complementation between the main texts/images/ and the headlines?
8. If so, why do you think the writers created these relationships?

#### Homepages and Political Stance

9. Do you think the political leanings of the news sites are noticeable through the image texts? In the headlines? In the body texts? If so, can you identify the elements that communicate the political stance? If not, why not?
10. Choose two news sites in your L1 that take different political stances. Compare the images/headline and texts. How are they similar? How are they different?

Table 2. Multi-modal C/R questions for analyzing news sites.

Comparing the sites from a multi-modal perspective, compositional elements allow the sites to create different registers. As Figure 2 shows, through variations in the compositional elements in column one, the BBC and the Daily Mail create differently looking sites to serve their different news functions. The BBC creates a relatively neutral-looking site by balancing the salience of its stories (i.e., a wide range of medium-sized images and headlines), clearly summarizing the type of stories through framing and placement, and using color (the formal blue) and images of politicians and professionals to reinforce the serious tone of the website. Moreover, “foregrounding,” a technique often used in advertising and newspapers, in which negative images accompany negative words and positive images accompany positive words was not particularly evident. By contrast, the Daily Mail created a popular and entertaining-looking site by giving salience to a small number of compelling stories for their perceived readership. The stories are clearly organized by size, with the most compelling stories given the most salience. In contrast to the BBC, the most common images were of celebrities and members of the public. Moreover, in the Daily Mail, headline to image foregrounding using graphic-written text relations of concurrence was clearly evident.

Thus, by comparing the sites, students review how different texts create different registers, which is a task they study in the written mode. However, with their knowledge of the compositional function, they can study the registers from a wider multi-modal perspective. The final point regarding the compositional function is that compositional principles are easily transferable to the construction of many other texts. Whether students are creating an academic essay or designing a web site from a template, compositional decisions must be made: Which elements should be most salient? Where are the different elements placed in the texts to best represent or support the writers’ ideas or opinion? Are the elements appropriate for the audience or argument?

Compositional Elements	BBC	Daily Mail
<b>Salience</b>	large number of small pictures with headlines	small number of large pictures with large headlines, as well as medium and small pictures with medium and small headlines
<b>Framing</b>	clearly framed and separated its stories by spaces and frame lines	used very few frame lines and separated stories by very small spaces
<b>Placement</b>	placed in clearly organized separated lines form left to right, top to bottom, under clearly discernible categories, for example, News, Asian Pacific, Business	stories are organized through interest, from top to bottom, with large stories at the top and smaller stories below them, right hand margins are used for the smallest stories and hyperlinks
<b>Color</b>	dark formal blue	light blue
<b>Foregrounding</b>	both images and words took a relatively neutral stance	positive images were often accompanied by positive words and negative image were often accompanied by negative words
<b>Image Participants</b>	international politicians, professionals, some celebrities and members of the public	a number of celebrities and members of the public, some politicians and professionals

Figure 2. Compositional analyses of BBC and Daily Mail homepages.

Using the representational and interactive functions, the SFG model allows for a comparison of the communicative roles that images play when they are combined with words to create meaning in authentic text contexts (the news sites), and when they are combined with words in non-authentic contexts (the classroom-created report, which the students create in part two of the task). This communicative role is fundamentally different, as explained below.

The communicative role of the images in news sites (at least those that have been analyzed in the writer's classroom) was to set the scene for the readers before they read the written text, to connect the readers to a key orienting point or concept in the written text, or to enhance or embellish one part of the written text. To accomplish these goals, a complex combination of different types of images was used with different types of graphic written/text relations.

The most common type of images used to perform these functions were concept/offer images, similar to the ones shown in Appendix 1, Figure 4, slides 1 and 5. Concept/demand images, similar to the images found in the Macbeth text Panel 2, 3, and 4, were found in stories that were intended to elicit an emotional response, to shock, or to stimulate the reader. For instance, a close-up image of a participant laughing or crying would be used to convey joy or sadness. Narrative/event images, which might be considered as common in news stories, were relatively rare. This observation is understandable because the majority of stories were not breaking news, and the texts provided details about events that had already occurred, and thus, narrative event images might not be relevant. Furthermore, the practical difficulty of finding narrative/event images of actual events from the time that they happened to press time may also account for their limited use.

In analyzing how the newspaper sites use graphic/written text relationships of concurrence and complementation, the latter was more common than the former.

Concurrence had two main functions: to draw the viewer's attention to a main participant, for example, the government, the army, an actor, and to create foregrounding, as pointed out previously. The limited use of concurrence, and the frequency of relationships of complementation, raises student awareness that although news sites are image rich texts compared to traditional printed texts, the alphabetic text still carries the main illocutionary force of the text because words are more efficient at conveying detailed meaning than image-based texts in this context and many other authentic text contexts.

Unlike classroom-based texts, overusing concurrent relationships in authentic texts would create redundancy. Although strong relationships of concurrence do occur in some authentic texts (e.g., in children's stories where reiteration is comforting for young learners or in the visual instructions that accompany the assembling of household objects) (Stenglin & Iedema, 2001), most L1 texts are proscriptive. That is, the texts drive the reader forward, and as explained above, the image-text is therefore used to augment or add information to the written text. Hence, concurrence occurred in news sites to create foregrounding.

Language students, reared in the use of concurrent images that provide linguistic support, may need teacher support when reading such online news sites and other image rich L1 homepages pages, because the texts may create miscomprehension or incomprehension. Indeed, compared with the traditional printed press reading paths, online reading paths are very challenging: students must navigate a wide array of images-texts, alphabetic-texts, hyperlinks, and advertisements. Depending on the student's proficiency level and experience, this process may be a very difficult task.

In contrast to the authentic text, the communicative role of images in the student report is to support linguistic understanding. Therefore, the types of images that students choose and the role that images play in relation to the spoken text will be fundamentally different from the authentic text. In the report, students can be encouraged to create multi-modal texts that have strong concurrent relationships to enable ease of comprehension in an L2 audience setting and follow clear academic patterns (Sustainable Resource 1 discussed above provides an example of the type of model students could follow), in contrast to the newspaper genre in which images text relations are far more complex. Finally, in creating the classroom report, not only are students working with new multi-modal skills, but they are also applying skills taught in the reading and writing curriculum, such as autonomous research, summarizing, note taking, paraphrasing, critical thinking, and citing sources.

#### *4.3. Sustainable CALL resource 3: evaluating multi-modal materials*

The multi-modal meta-language, outlined above and summarized in Tables 1 and 2, gives teachers the tools, not only to create sustainable resources from their existing classroom practices, but also to evaluate multi-modal classroom materials, such as publisher-created videos, software presentations, and online materials adopted for classroom use, for their sustainability in relation to the existing curricula. Table 3 provides examples of questions that might be included in such an evaluation process. The goal of the questions is to evaluate whether the digital materials support the goals of the existing curricula and whether they diverge from the curricula in ways that are inappropriate.

- What curriculum goals is the multi-modal features intended to support?
- Are the multi-modal relations of concurrence and complementation appropriate for the levels being taught, or will the visuals create redundancy, miscomprehension, and/or incomprehension?
- How effectively does the video involve the students in the text through its use of demand/offer concept/narrative images?
- Is the visual component composed in a way that is appropriate for the student’s level, age, and learner type?
- Is the visual component composed in a way that is appropriate for the institution?

Table 3. Multi-modal questions for material evaluation.

For example, in the established curricula in my teaching context, the *World Link Textbook* series uses a video course book to expand on the textbook materials and recycle the linguistic components in natural settings and situations (Stempleski, 2013). In a short excerpt (see script in Figure 4 of the Appendix) from Video Course Workbook 2 (Unit 1, City Living, pp. 8-9), the lesson reviews the past tense of verbs using a discussion on keepsakes. In this example, the keepsake that triggers the recollection of Tara, a character in the video, is a pendant. As Table 4 shows, demand/concept images, such as close-ups of the pendant and Tara’s face, accompany the key communicative phrases of the script: “it’s a pendant from my grandmother,” and “she gave it to me when I was 18 years old”.

Speaker	Script: Verbal Text	Visual Text
1) Sun-hee	How about this?	Concept/demand image showing a close up of the pendant
2) Tara	Now that is my favorite keepsake. It’s a pendant from my grandmother. She gave it to me when I was 18 years old.	Concept/demand of showing close up of the pendant Concept/offer showing close of Tara’s face
3) Sun-hee	For your birthday?	Concept/demand showing close up of Sun-hee’s face
4) Tara	No. It was in my first year of college and things were rough. I had no friends. I hated my classes. I did not think I could make it. And one day my grandmother told me a story.	Concept/demand showing a close up of Tara’s face

Table 4. Text excerpts from *World Link Video Course Book*.  
Textbook Extract Used with permission © Cengage Learning.

In this scene, the demand/concept images, similar to the face-to-face close-ups of Panel 2, 3, and 4 of the Macbeth text (Figure 1), create interactions between Tara, who is re-telling the story, and the viewers, by clearly focusing the viewers on the speakers and the keepsake at a key point in the text. The images allow viewers to identify with the speaker emotionally, thus reinforcing the communicative component of the lesson, which is the expression of the emotional value of keepsakes. Moreover, the demand/concept images allow students to pick out variations in facial expressions and intonation that the actress uses when expressing the key phrases, providing clear models for students to replicate when re-telling their own past stories. In a traditional textbook in which students listen to an audio recording with minimal visual support, such emotional content is very difficult to establish.

Furthermore, the video does not overuse concurrent verbal/visual text relations that could make the linguistic goals of past tense re-telling redundant. The past tense

recollection story (Table 4) relies on spoken text re-telling; the concept/demand images do not reiterate the key events in the story. This verbal/visual text relationship creates a positive teaching opportunity because, given that past tense re-telling is the linguistic aim of the lesson, the use of concurrent relationships would create redundancy at this level.

Through the development an appropriate pedagogical language, this ability to evaluate the extent to which multi-modal resources are appropriate for teaching linguistic goals, is a key feature in creating sustainable materials in the long term. For example, digital games are recommended for educational use because games have interactive features that create intrinsic motivational factors lacking in traditional classroom textbook materials. Such factors include encouraging participation through player investment in characters and game development, creating opportunities for player decision making, systems of reward and merit, competition, and interacting storytelling with play (Miller, 2004, pp.198-199). Nonetheless, games that employ these motivational features are not currently available for language learning contexts. In addition, the extent to which current video games on the market are directly beneficial to the curricula is debatable (Gee, 2011).

The multi-modal pedagogical language outlined above can be used by designers and material developers to aid in the creation of a new generation of classrooms and/or self-study materials that incorporates the motivational features of gaming, while ensuring that the materials are relevant for L2 contexts. For example, when digital computer games are created, a traditional alphabetic-based script accompanies the digital script. Knowledge of textual relations and their effects on comprehension at different levels of proficiency can ensure that the scripts for language learning digital games are appropriate for students' levels. Alternatively, understanding how images are composed to create different emotional reactions from viewers can help developers design visual/alphabetic interfaces that support linguistic goals, as well as creating stimulating multi-modal features.

## **5. Conclusion**

Given the demands of working with institutionally created curricula, one of the most challenging questions confronting language teachers is an opportunity cost question: will sending your students to the computer room be an appropriate use of classroom time? The concept of sustainability that is outlined in this paper is designed to address this question. If teachers have a multi-modal pedagogical language available to them, class activities such as the creation of a multi-modal composition need not be regarded as separate or distinct from teaching the established curricula.

Nevertheless, the primary focus of this paper is short-term adaptability to overcome situational constraints; thus, teachers use their existing pedagogical knowledge, coupled with SFG multi-modal pedagogical knowledge, to create sustainable classroom resources for CALL. In the long term, however, teachers cannot overcome situational constraints individually. Moreover, the SFG model alone is not sufficient to address the challenges of preparing students for effective communication in digital environments. To achieve long-term sustainability, researchers, practitioners, curricula developers, classroom material designers and textbook publishers must develop a pedagogical language that embraces a new multi-disciplinary approach to language and learning for the digital age.

**Appendix. Example of sustainable resource 1: multi-modal conversion activity.**

1. The country is better than the city because there is a lot of pollution in the city. 2. In the city there are many types of pollution: noise, tobacco smoke, gas exhaust, and acid rain. 3. Pollution is bad for our health and puts humans at risk of diseases such as cancer. 4. The countryside is free from pollution and there is less risk of disease. 5. For this reason, I prefer to live in the countryside than in the city.

Figure 1. Teacher-created Written Paragraph.

Sentence Number	Type of Sentence	Function
1.	Topic Sentence	Introduce the general idea
2.	Supporting Sentence	Support general idea with an example
3.	Supporting Sentence	Support general idea with an explanation/example
4.	Supporting Sentence	Support general idea with an explanation
5.	Concluding Sentence	Repeat the main idea

Figure 2. Teacher-created Written Paragraph Deconstructed.

Slide Number	Type of Image	Function	Verbal/Visual Textual Relation
1.	Concept/Offer	Introduce the main idea	Complementation
2.	Concept/Demand	Support general idea with an example	Concurrence
3.	Concept/Demand	Support general idea with an example	Concurrence
4.	Concept/Demand	Support general idea with an example	Concurrence
5	Concept/Offer	Repeat the main idea	Complementation

Figure 3. Teacher-created Presentation Deconstructed.

### Slide 1

**Spoken Text**  
The country is better than the city because there is a lot of pollution in the city.



### Slides 2, 3 & 4

**Spoken Text**  
In the city, there are many types of pollution: noise, gas exhaust, and acid rain.

### Slide 3



### Slide 2



### Slide 4





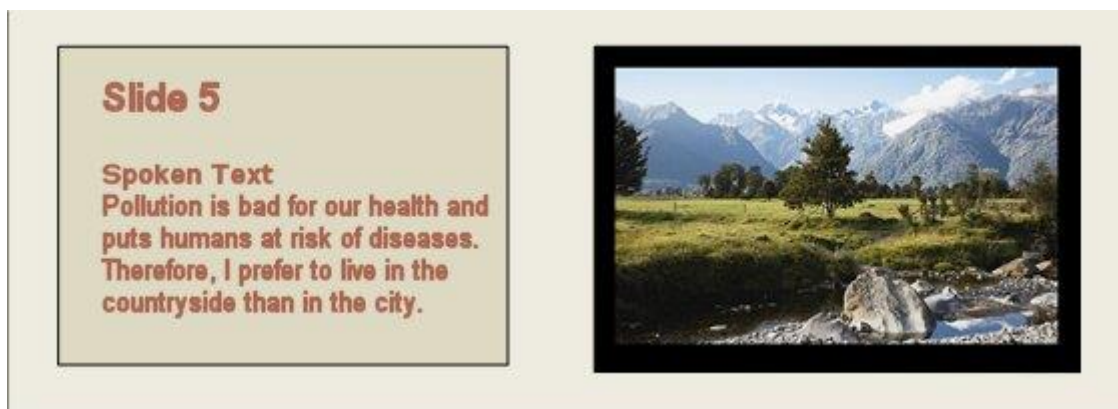


Figure 4. Teacher-created presentation.

## References

- Collins, A. & Halverson, R. (2009). *Rethinking education in the age of technology*. Columbia: Teachers College.
- Cuban, L. (2001). *Oversold and underused: computers in the classroom*. Cambridge, MA: Harvard University Press.
- DeVoss, D.N., Eidman-Aadahl, E., & Hicks, T. (2010). *Because Digital Writing Matters*. CA: John Wiley & Sons.
- Fries, P.H. (1994). On Theme Rheme and Discourse Goals. In M. Coulthard (ed.). *Advances in Written Text Analysis*. (pp. 229- 249). New York: Routledge.
- Gee, Paul, (2011). Reflections On Empirical Evidence On Games and Learning. In T.Sigmund & J.D. Flechter (Eds.), *Computer Games and Instruction*. (pp. 223-232). Charlotte NC: Information Age Publishing
- Hagood, M. (2008). Intersections of Popular Culture, Identities, and New Literacies. In J. Coiro et al (eds). *Handbook of Research into New Literacies*. (pp. 377-407). New York: Erlbaum.
- Halliday, M.A.K. & Matthiessen, C.M.I.M. (2004). *An Introduction to Functional Grammar*. London: Hodder and Arnold.
- Hoey, M. (2001). *Textual Interaction*. Oxon: Routledge.
- Kennedy, C. (2013). Models Of Change and Innovation. In K. Hyland & C. Wong (Eds.), *Innovation and Change In English Language Education*. (pp. 13-27). Oxen:Routledge.
- Kress, G. (2003). *Literacy in the New Media Age*. Oxon: Routledge.
- Kress, G. & Van Leeuwen, T. (2006). *Reading Images*. London: Routledge.
- Liu, J. (2004). Effects of comic strips on L2 learners reading comprehension. *TESOL Quarterly*, 38(2), 225-243.
- Luke, C. (2001). Connectivity, Multimodality and Interdisciplinarity. In *Reading Research Quarterly*, Vol.38, No.3, 397-403.
- Marchand, T. (2013). Speech in written form? A corpus analysis of computer-mediated communication. *Linguistic Research*, 30(2), 217-242.
- McCloud. S. (1994). *Understanding Comics: The Invisible Art*. New York: Harper Collins.
- Mc Donald, J. Haward, J. Dobbin, N. Erskine, G. (2008). *Macbeth: The Graphic Novel*. Bristol: Classical Comics Ltd.

- Miller, C.H. (2004). *Digital Storytelling*. Oxford: Elsevier.
- Unsworth, L. (2008). Multiliteracies and Metalanguage: Describing Image Text Relations as a Resource for Negotiating Multimodal Texts. In J. Coiro et al. (eds). *Handbook of Research into New Literacies*. (pp. 377-407). New York: Erlbaum.
- Royce, T. (2002). Multi-modality in the TESOL classroom. In *TESOL Quarterly*. Vol. 36, No.2, Summer 2002. 191-205.
- Rutherford, W .E. (1987). *Second Language Grammar and Teaching*. New York: Pearson Education.
- Stempleski, S. (2013). *World Link: Developing English Fluency*. Singapore: Cengage
- Stenglin, M. and Iedema, R. (2001). How to Analyse Visual Images: A Guide for TESOL Teachers. In A. Burns. & C. Coffin. *Analyzing English in a Global Context*. (pp.194-208) London: Routledge.
- Selfe, C.L. (2007). *Multi-Modal Composition*. NJ: Hampton Press.

## Article:

### ***Lessons Learned in Designing and Implementing a Computer-Adaptive Test for English***

**Jack Burston\* and Maro Neophytou\*\***  
**Language Centre**  
**Cyprus University of Technology**

---

\*jack.burston @ cut.ac.cy | \*\*maro.neophytou @ cut.ac.cy

#### **Abstract**

This paper describes the lessons learned in designing and implementing a computer-adaptive test (CAT) for English. The early identification of students with weak L2 English proficiency is of critical importance in university settings that have compulsory English language course graduation requirements. The most efficient means of diagnosing the L2 English ability of incoming students is by means of a computer-based test since such evaluation can be administered quickly, automatically corrected, and the outcome known as soon as the test is completed. While the option of using a commercial CAT is available to institutions with the ability to pay substantial annual fees, or the means of passing these expenses on to their students, language instructors without these resources can only avail themselves of the advantages of CAT evaluation by creating their own tests. As is demonstrated by the E-CAT project described in this paper, this is a viable alternative even for those lacking any computer programming expertise. However, language teaching experience and testing expertise are critical to such an undertaking, which requires considerable effort and, above all, collaborative teamwork to succeed. A number of practical skills are also required. Firstly, the operation of a CAT authoring programme must be learned. Once this is done, test makers must master the art of creating a question database and assigning difficulty levels to test items. Lastly, if multimedia resources are to be exploited in a CAT, test creators need to be able to locate suitable copyright-free resources and re-edit them as needed.

**Keywords:** Computer-Assisted Testing, CAT, English, placement, test authoring.

#### **1. Background**

In our Language Centre, as in many European universities with an EFL course requirement, the linguistic level of incoming students can vary across the entire range of the Common European Framework of Reference for Languages (CEFR) scale. Since all first-year students at our university have to complete a two-semester B1 level Academic English course as a graduation requirement, those who enter the university with English language proficiency below this level risk not only failing the course but also failing to obtain their degree. As there is neither time in the schedule nor funding for remedial classes, at the start of every academic year an urgent need arises to identify weak students in order to provide them with counseling and self-study guidance. To meet this need, our Centre previously carried out diagnostic evaluation

using a commercial paper and pencil test (MacMillan), in-class oral interviews and a writing assignment. Although this procedure gave satisfactory results, it was time consuming to administer and evaluate, with results not being known for at least two weeks after the start of classes. In order to improve diagnostic efficiency, we turned to computer-based testing since such evaluation can be administered more quickly, automatically corrected, and the outcome known as soon as the test is completed.

## **2. Computer-based test options**

### *2.1. Non-adaptive tests*

In seeking an alternative to our previous diagnostic testing procedures, one non-adaptive online option was considered: DIALANG. DIALANG attracted our attention because it evaluates a wide range of skills (reading, writing, listening, grammar and vocabulary) in English as well as more than a dozen other European languages. So, too, it is freely accessible and aligned with the CEFRL. However, since it is non-adaptive, students have to answer all questions at whatever level they self-select for testing. In a class environment this can be problematic since the test can take longer to administer than the time available in a single session. So, too, DIALANG is based on a relatively small question inventory and, being the product of a long completed EU project, lacks funding for ongoing maintenance and development. Moreover, since DIALANG does not run over the Internet (or even a local area network server), it must be individually installed on all computers. Aside from the initial complications this can entail when several labs have to be used, it also restricts flexibility should access to suitably configured labs change at the last moment. Added to these constraints, DIALANG provides no record keeping at all. At the end of a test, students are given their result, but can only write it down or, provided a printer link is available, hand in a screen print of it. For these reasons we were obliged to look elsewhere for a computer-adaptive alternative for our diagnostic testing.

### *2.2. Computer-adaptive test design*

Computer-adaptive tests are based on Item Response Theory (Hambleton, Swaminathan & Rogers 1991). The simplest, and most frequently implemented, are constructed according to a single parameter Rasch model (Rasch, 1980), which is governed only by the difficulty of the item and the ability of the person located on the same continuum. In such a test, responses are sought to questions of pre-established difficulty level. Students who can consistently answer questions at difficulty level X are deemed to demonstrate X level proficiency. A computer-adaptive test (CAT) automatically adjusts to the proficiency level of students by presenting easier questions following incorrect responses and more difficult ones after correct answers.

By targeting questions within a range that a student can consistently answer correctly, a CAT can be administered using a relatively small number of question items. Compared to a traditional non-adaptive test, which typically might contain 75-100 questions, a CAT can usually determine a student's language proficiency level in 25 questions or less. Although any particular student may at most see only a couple of dozen test items, in order to have a sufficient number of items in reserve at various levels of difficulty, the operation of a CAT requires a question database several times this size. It also requires a computer-based algorithm to select the questions to be presented, determine the correctness of responses, and adjust the difficulty level of subsequent questions accordingly.

### 2.3. Computer-Adaptive Tests

#### 2.3.1. Commercial tests

The most comprehensive, and undoubtedly best known, computer-adaptive programme for evaluating foreign language proficiency is the Brigham Young University *CAPE* (Computerized Adaptive Placement Exams). It tests grammar, vocabulary and reading comprehension and is aligned with the American Council for the Teaching of Foreign Languages (ACTFL) proficiency guidelines: novice, intermediate, advanced and superior. In its most recent iteration, known as *webCAPE*, it includes tests for six languages including L2 English. As its name implies, it is Internet-based and so can be accessed without installation on local computers. The *CAPE* series is based on a very large question database (nearly 1000 items per language) and provides statistically reliable results with detailed record keeping. However, its use comes at a cost (e.g., \$1,700/year for 500 students, if paid by the University) which our Centre simply could not afford. Alternatively, the cost (\$10) of taking the *CAPE* can be passed on directly to students, which in our public institution was not an option.

#### 2.3.2. Free tests

Fortunately, two cost-free CAT creation options are available as an alternative to a commercial test: *Concerto* and *SLUPE*. Of the two, *Concerto* is by far the most flexible and powerful. Distributed by the University of Cambridge, *Concerto* is an online R-based adaptive testing platform. Being open-source, it can be fine-tuned to the evaluation of competence in virtually any domain. That being said, its implementation requires the services of a computer programmer fluent in R and someone with a solid background in statistical analysis. On the one hand, this makes it an ideal choice where such expertise is available. On the other, as in our case, it puts *Concerto* out of reach when the required technical expertise is not accessible.

Though much more limited in its capabilities than *Concerto*, *SLUPE* (Saint Louis University Placement Exam) has the great advantage of requiring no programming ability or statistical expertise of test creators. *SLUPE* is a user-friendly CAT authoring system which requires only that test makers create their own question database. It allows two types of testing format:

- a) Text-based: multiple-choice questions with four options and only one correct answer.
- b) Audio/video-based: a set of five True/False options, 0-5 of which may be correct answers.

Questions and answers are simply entered into an online text box. Audio and video prompts can either be uploaded to the *SLUPE* website or linked to an external source (e.g., YouTube). Test makers assign a difficulty level of 1-4 (easy-hard) to each question. By default, the four difficulty levels within *SLUPE* correspond to semester divisions. However, these can be associated with whatever proficiency scale test authors choose. Once questions have been added to the database, *SLUPE* takes care of everything else. Like *Concerto*, *SLUPE* is web-based and so requires no local computer installation. Each test is associated with a specific URL which instructors give to students along with a log-in id and password. The CAT algorithm underlying *SLUPE* automatically handles question presentation based on difficulty levels and keeps detailed records of student responses: the questions they attempted, whether they were answered correctly or not, and their final placement level. It also tracks results organized by test item responses, thus allowing subsequent statistical analysis of actual question difficulty levels. For language teachers like ourselves, with minimal technical and/or financial support, *SLUPE* was an obvious choice when starting out to create a CAT.

### 3. The E-CAT

#### 3.1. Test creation

While *SLUPE* enormously simplifies the technological and computational aspects of CAT creation, the quality of placement obtained with it very much depends upon the teaching experience and testing expertise of would-be test makers.

#### 3.2. Theoretical considerations

As with any test, construct validity (Cronbach & Meehl, 1955) arguably must be the primary consideration, i.e., does the test actually assess what it claims to evaluate? In the case of our test, dubbed the E-CAT, its intended purpose was to assess the general L2 English proficiency of first-year university students. In particular, it sought to identify the weakest students, those below A2 (CEFR), in order to provide them with appropriate counseling and self-study guidance.

Attaining construct validity is challenging for any CAT used for language proficiency assessment, all the more so when aligned with the CEFR. By definition, CEFR criteria are all performance-based, i.e., they describe what students are able to do with the language in given situations. On the other hand, by design, all computer-adaptive tests are based on fixed answer responses (e.g., multiple-choice questions), which most easily targets grammar and vocabulary knowledge. Typically, listening and reading comprehension are the only performance-related language skills tested in a CAT. As a consequence, the construct validity of any CAT-based assessment of language proficiency depends critically upon the content validity of the grammar and vocabulary that is tested, i.e., the degree to which their mastery is representative of a given proficiency level. In the case of the CEFR, content validity equates to the mastery of those elements of grammar and vocabulary that allow defined language functions to be successfully performed. While listening and reading comprehension tasks allow receptive language skills to be tested, it is also possible to assess more active skills by using prompts (text as well as audio) to solicit communicatively appropriate responses. For example:

Audio Prompt - *They live on a shoe string nowadays.*

(Possible text-based responses, 0-5 of which may be correct)

- Yes, they have it pretty easy.
- Yes, they have little money.
- They should buy sandals.
- They are just stringing you along.
- They are frugal, they'll get by.

#### 3.3. Practical considerations

Owing to their fixed nature, *SLUPE* questions are subject to two notable constraints. Firstly, while audio-video-based listening comprehension testing is easily accommodated through the use of multiple true-false questions, reading comprehension tasks cannot be effectively exploited. Text-based prompts can only be associated with a single multiple-choice question, i.e., one text passage cannot serve as the basis for multiple comprehension questions. It could easily take a student a couple of minutes to read a passage of any substance, which is far too long to devote to a single question. Secondly, while question prompts may be in written, oral or video form, only text-based responses are supported. As a consequence, *SLUPE* cannot be used to present audio-based communicatively appropriate responses (see 3.2 above).

Although the creation of text-based questions is very straightforward, the exploitation of audio and video resources as question prompts is considerably more demanding. Finding appropriate materials can be very time consuming and, once located, copyright

permission must be obtained for their use. Because of the complications involved in obtaining copyright permission, would-be test creators are well advised to limit their search for audio-video materials to copyright-free or creative commons sources.

Aside from general copyright permission, the exploitation of audio-video resources makes two other demands on test makers. Firstly, copyright usage must allow the material to be modified in order to extract just that portion of the audio-video file needed as a test prompt. Typically, this would be no more than 60-90 seconds from a passage that might run for five minutes or more. Secondly, the test creator must either possess the editing skills needed to modify audio-video resources or have access to technical assistance to get the job done.

In principle, *SLUPE* can operate with as few as 52 test questions:

- 10 text-based at four levels (= 40 items)
- 3 audio/video-based at 4 levels (= 12)

However, statistical reliability requires at least twice this number of test items in the database. The E-CAT was first created with 112 testing items. Subsequent to initial testing, this was increased to 144. The E-CAT test was pilot tested in April 2013 with approximately 200 students during the second semester in their compulsory first-year course. In September-October 2013 approximately 450 first year-students sat the test. Another 350 students sat the test in March-April of 2014.

#### 3.4. Difficulty level calibration

For our purposes, in assigning question item difficulty, the *SLUPE* semester levels 1-4 were equated with CEFR A2, B1, B2 and C1. Since *SLUPE* places students who score above the top level in semester 5, we equated this with C2.

By definition, the proficiency level of a student taking an IRT-based CAT is equated with the difficulty level of test-items that are correctly answered. Consequently, the reliability of such placement is critically dependent upon the accuracy of the difficulty level assigned to each question. Although *SLUPE* itself allows question difficulty levels to be determined freely by whatever means test makers choose, until a question database has been administered to a reasonably large number of students, i.e., several hundred at least, there is no way of knowing with any certainty the actual difficulty of any question. This can only be determined by an *ex post facto* analysis of the relative frequency with which questions were answered correctly or incorrectly.

In principle, it is possible to create a CAT on the basis of a question database previously analyzed for difficulty level, for example one derived from an earlier paper and pencil version of a test. However, doing so assumes that differences in testing conditions (e.g., with or without the use of a computer) and student populations will not significantly affect question difficulty levels. In the absence of an existing question database of known difficulty level, as was our case, the initial assignment of item difficulty of necessity can only be done intuitively. In any event, however difficulty levels are initially determined, a CAT question database needs to be recalibrated several times based on actual responses from a representative student population before reliable placement can be assumed. Very often, especially at the early stages of CAT development, the recalibration of item difficulty level results in gaps being created in the database which have to be filled by the creation of new test items at the levels that have been vacated. The difficulty level accuracy of these additions then needs to be validated through the analysis of subsequent administrations of the CAT.

While the easiest and most difficult items in a question database are relatively easy to identify, i.e., those which the most students answer correctly or incorrectly, any detailed determination of question difficulty level can only be done by proper statistical

analysis. Even the most experienced language teachers cannot intuitively assign question difficulty levels with any high degree of accuracy. Compared to the statistical analysis of student responses, our initial estimations of question difficulty level in the E-CAT were correct less than half of the time, with considerable standard error and many discrepancies of 2-3 levels. Following the first recalibration, the statistical analysis of the second administration of the test again revealed an accuracy rate of less than 50% in question difficulty assignment, but this time with a considerably lower standard error of measurement. Moreover, 91% of the level assignments resulting from the recalibration were within +/-1 level of the statistical estimates of question difficulty. Analysis of the third iteration of the test demonstrated further improvements in test accuracy, with 72% of the difficulty settings agreeing with the statistical estimates.

### 3.5. Placement results

As a reference point for placement accuracy, the E-CAT results from its third pilot testing were compared against our instructors' evaluation of their students' proficiency level based on a whole semester (and in some cases an entire academic year) of class performance. Across all levels, the E-CAT agreed exactly about 40% of the time, with no more than +/-1 level divergence in another 48% of the placements. Below the A2 level, which was our primary concern, exact agreement was higher at 50% with no more than +1 level divergence in another 33% of the placements. Overall, then, in well over 80% of the cases the E-CAT successfully placed students with reasonable accuracy in less than one class period compared to instructors who had the advantage of at least an entire semester to make their judgment. As the accuracy of question difficulty levels improves through continued statistical analysis of test results, it is expected that so, too, will placement accuracy.

## 4. Conclusion

Based on our experience with the E-CAT, we can say with confidence that it is definitely feasible for language teachers without computer programming skills to create reliable computer-adaptive tests using the freely accessible *SLUPE* authoring programme. That being said, the process is neither quick nor effortless. Above all, it requires collaborative teamwork to succeed, which in our case involved five experienced language teachers. Initial test construction, learning how to use the *SLUPE* system and even more so building an operational question database, can be expected to take a whole semester. If multimedia resources are to be effectively exploited, test creators need to be able to locate suitable copyright-free resources and re-edit them as needed. Undoubtedly, the most challenging and critical aspect of question creation is the proper assignment of difficulty level. As our experience demonstrates, on their own, even the most experienced language teachers are unlikely to get this right more than half the time. Since by definition the reliability of any CAT-based student placement is directly determined by the accuracy of question difficulty assignments, access to *ex post facto* statistical analysis of item difficulty levels is essential. At least two pilot testing sessions, typically spread over two semesters and involving several hundred students, are required to evaluate placement results and adjust the question database accordingly.

For those fortunate enough to have the financial resources to pay the recurrent fees for the use of a commercial language test such as *webCAPE*, constructing a CAT may very well appear to be too demanding a task. On the other hand, making a virtue of necessity, once a locally developed CAT is operational it has one great advantage over any commercial test. Having been calibrated against the local student population for which it is intended, the difficulty level of its test items is much more closely matched to the proficiency of its test takers, with correspondingly greater placement accuracy. In cases where the native language of students being assessed is quite different from that



typically used to calibrate a commercial CAT, e.g., L1 Greek, Chinese, or Arabic speakers learning L2 English, this can make a significant difference.

### **References**

Cronbach, L. J.; Meehl, P.E. (1955). Construct Validity in Psychological Tests. *Psychological Bulletin* 52: 281–302.

Hambleton, R., Swaminathan, H., & Rogers, J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Publications.

Lawshe, C.H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28, 563–575.

Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danmarks Paedagogiske Institut, 1960. Reprint, Chicago: University of Chicago Press.

## Article:

### ***How EFL students can use Google to correct their "untreatable" written errors***

**Luc Geiller**  
**ATILF/CNRS, Nancy University, France**

---

[lgeiller @ gmail.com](mailto:lgeiller@gmail.com)

#### **Abstract**

This paper presents the findings of an experiment in which a group of 17 French post-secondary EFL learners used Google to self-correct several "untreatable" written errors. Whether or not error correction leads to improved writing has been much debated, some researchers dismissing it as useless and others arguing that error feedback leads to more grammatical accuracy. In her response to Truscott (1996), Ferris (1999) explains that it would be unreasonable to abolish correction given the present state of knowledge, and that further research needed to focus on which types of errors were more amenable to which types of error correction. In her attempt to respond more effectively to her students' errors, she made the distinction between "treatable" and "untreatable" ones: the former occur in "a patterned, rule-governed way" and include problems with verb tense or form, subject-verb agreement, run-ons, noun endings, articles, pronouns, while the latter include a variety of lexical errors, problems with word order and sentence structure, including missing and unnecessary words.

Substantial research on the use of search engines as a tool for L2 learners has been carried out suggesting that the web plays an important role in fostering language awareness and learner autonomy (e.g. Shei 2008a, 2008b; Conroy 2010). According to Bathia and Richie (2009: 547), "the application of Google for language learning has just begun to be tapped." Within the framework of this study it was assumed that the students, conversant with digital technologies and using Google and the web on a regular basis, could use various search options and the search results to self-correct their errors instead of relying on their teacher to provide direct feedback.

After receiving some in-class training on how to formulate Google queries, the students were asked to use a customized Google search engine limiting searches to 28 information websites to correct up to ten "untreatable" errors occurring in two essays completed in class. The findings indicate that a majority of students successfully use material from the various snippets of texts appearing on the Google results pages to improve their writing.

**Keywords:** Data-driven learning, Google-driven language learning, learner autonomy, error treatment, self-correction, language awareness.

## 1. Introduction

### 1.1. Data-driven learning (DDL)

"Data-driven learning" (DDL) was first used by Johns (1990) to refer to learners directly exploring authentic language by means of corpora, acting as researchers discovering language patterns, formulating and testing hypotheses. A number of recent studies have highlighted the usefulness of corpora and concordancers as tools to facilitate second language learning, particularly its impact on vocabulary acquisition and improved writing skills (Chambers, Conacher & Littlemore 2004; Chen 2004; Chen & Baker 2010; Jarvis 2004; Johansson 2009; Kennedy & Miceli 2010; Yoon 2008; Yoon & Hirvela 2004). As explained by Boulton (2009a: 83), DDL "can sensitise learners to issues of frequency and typicality, register and text type, discourse and style, as well as the fuzzy nature of language itself."

Reporting on their attempts to make concordance information accessible to lower-intermediate L2 writers as feedback to sentence-level written errors, Gaskell and Cobb (2004) explain that learners are willing to use concordances to work on grammar and that they are able to self-correct based on those concordances. They argue that online corpus exploration can reduce the burden on teachers, all the more so as the formal teaching of rules is not always effective in helping learners achieve more grammatical accuracy because "sentence-level writing errors seem immune to many of the feedback forms devised over the years" (p. 1). Similarly, Milton (2006) believes that encouraging learners to use online corpora for assistance "can help relieve teachers of the need to act as proofreading slaves" (p. 125). The rationale behind this is that maximizing learners' contact with English helps them detect recurring language patterns, thus increasing their language awareness in a data-driven learning process. The objective is for them "to acquire the means and confidence to self-edit in the future" (p. 131), which is in keeping with what Benson (2001) says about learner autonomy and language acquisition being dependent upon the capacity to initiate and manage one's own learning:

*Many advocates of autonomy in language learning would [...] share Rousseau's view that the capacity for autonomy is innate but suppressed by institutional learning. Similarly, Rousseau's idea that learning proceeds better through direct contact with nature re-emerges in the emphasis on direct contact with authentic samples of the target language that is often found in the literature on autonomy in language learning. (p. 25)*

But although the use of corpora in the classroom has imposed itself as an inescapable language learning tool, several barriers must be overcome before it goes mainstream. The activity is potentially time-consuming and tedious, and teachers and students can be reluctant to accept the changes to their traditional roles in the learning process. It may even be that they do not have a sufficient level of competence in ICT. More concretely, Widdowson (2000), argues that analyzing decontextualized and truncated concordance lines is an inauthentic activity and Johansson (2009) deplores the lack of empirical evidence supporting the theoretical benefits of DDL. Yoon (2008), for his part, suggests that learning style preferences can account for the slow acceptance of corpus use as an educational tool. As he puts it, "many corpus studies have regarded learners as a monolithic group rather than as idiosyncratic individuals" (p. 32). In other words, while some learners obviously benefit greatly from the approach, others do not. The challenge, then, is for teachers to adapt corpus exploration techniques to different learners so as to better cater to their individual needs.

### *1.2. Google-driven language learning*

According to Rundell (2000: n. pag.), the web "is not a corpus at all according to any standard definitions: what it is is a huge rag-bag of digital text, whose context and balance are largely unknown." Berg (2005: 2), for his part, argues that "the Web turns out to be a somewhat intractable collection of textual material, [...] a rather haphazard accumulation of digital text." The acronym GALL (Google-assisted language learning) was first coined by Chinnery (2005) who described Google as an informative, productive, collaborative, communicative, and aggregative tool with lots of pedagogical uses. Substantial research on Google as a tool for second-language learners has since then been carried out (e.g. Guo & Zhang 2007; Milton 2006; Shei 2008a, 2008b; Wu, Franken, & Witten 2009) suggesting that it plays an important role in fostering language awareness and learner autonomy. According to Bathia and Richie (2009: 547), "the application of Google for language learning has just begun to be tapped." A number of studies, however, point to problems associated with the use of Google and the web for language learning, namely the abundance of potentially unreliable data and the daunting task of scouring huge amounts of language (Berg 2005; Kilgarrieff 2001; Renouf 2003; Fletcher 2004; Robb 2003a, 2003b; Rundell 2000). Robb (2003a) calls it "a quick 'n dirty corpus tool," he warns about its use in class (2003b), explaining that queries are limited to specific words only, that there is no way of assessing the reliability of the language featured in the search results, and that these are not presented in a user-friendly format.

Several attempts at harnessing and systematizing web output have been made though. Since 1998, the University of Central England in Birmingham has been developing WebCorp1, a system for extracting linguistic data from the web, presenting examples of word usage from the Web in a form suitable for linguistic analysis. Similarly, KWICFinder2 and WebAsCorpus.org3, launched in 2007 by William Fletcher, can produce concordances from webpages. Guo and Zhang (2007) have built a customized collocations collector that can be used by language users, and Wu et al. (2009), acknowledging the heterogeneous, uncontrolled, and messy nature of web data, have explored the use of web searches as a language learning tool and used the Greenstone digital library software4 to organize raw online data that can be sifted through by language learners. But if Google enthusiasts insist on using raw online data, one way of dealing with the messiness and potential unreliability of the search results can be to use Google Custom Search5, a service launched by Google in 2006 which allows creators to select what websites will be used to search for information, thus eliminating any unwanted websites. For language learning purposes, it is thus possible to create a search engine that will only search specific news websites, for example.

### *1.3. Google use and its impact on language development*

Several studies have documented the impact of the web and search engines on language development and writing improvement (Acar, Geluso, & Shiki 2011; Clerehan, Kett, & Gedge 2003; Conroy 2010; Johnson 2004; Kennedy & Miceli 2010; Kenworthy 2004; Krajka 2000, Mansor 2007). Shei (2008a, 2008b) has shown that Google searches make it possible to compare the frequency of extended collocations (combinations of up to four words) and find the most commonly used and hence more formulaic ones. This suggests that Google output, however messy it is, *can* be used by second-language learners to explore native-speaker discourse and increase their language awareness.

Various studies have shown that some learners are keen users of information-related web services (e.g. Schroeder et al. 2010; Palfrey & Gasser 2008). Conroy (2010) reports that his students enthusiastically used Google and traditional concordancers for

language learning and error correction but that training was a key factor in getting them to use the approaches successfully. Although Google is a useful writing support tool, deciding which errors are amenable to correction needs further exploring. He also explains that students, being regular Google users, are more likely to favour the search engine than traditional corpora for which new interfaces have to be learnt, something learners sometimes find off-putting. Sun (2003) and Hafner and Candlin (2007) also found that learners preferred using Google to concordancers to learn about idiomaticity. As Shei (2008b) puts it, Google “remains a constant companion to the learner in the absence of the tutor. All the [teacher] has to do is to show the learner how to use this versatile tool” (p. 23). As explained by Boulton (2012): The objections [...] to using the web as ‘corpus’ and search engine as ‘concordancer’ have been shown to be largely theoretical, and based on criteria which are of little relevance in language teaching. The main conclusion is pragmatic and practical rather than dogmatic or ideological: if an approach or technique is of benefit to the learners and teachers concerned, it should not be ruled out automatically (Hafner & Candlin, 2007). As so often, there is likely to be a payoff between how much the teachers / learners are prepared to put in (ideally as little as possible) and how much they want to get out (ideally as much as possible). (n. pag.)

Kennedy and Miceli (2010) describe their use of the Contemporary Written Italian Corpus (CWIC) created at Griffith University to teach Italian to beginners, and especially to use corpus information to self-correct. Referring to Johns (1988), they sought to help their students develop observation strategies to extract information from concordances, developing what they call an “observe and borrow” mentality first, before progressing to an ‘observe and derive rules’ approach” (p. 1). They then explain that their aim was to “facilitate as much as possible their noticing the gap between their interlanguage and native speakers’ production,” encouraging them to explore the corpus “in search of words, expressions and even sentences that can be ‘plundered’ for use in their own compositions”—a “treasure-hunting” activity as they call it (p. 5).

#### *1.4. Error treatment in second language writing*

Whether or not error correction leads to improved writing has been much debated, some researchers dismissing it as useless (e.g. Hendrickson 1978; Kepner 1991; Sempke 1984; Truscott 1996; Zamel 1985) and others arguing that error feedback leads to more grammatical accuracy in students’ writing (e.g. Bates, Lane & Lange 1993; Bitchener et al. 2005; Bitchener 2008; Ellis 1998; Ferris & Roberts 2001; Ferris 2004; Hyland 2003; Chandler 2003). In her response to Truscott (1996), Ferris (1999) explains that it would be unreasonable to abolish correction given the present state of knowledge, and that further research needed to focus on which types of errors were more amenable to which types of error correction. In her attempt to respond more thoughtfully and effectively to her students’ errors, she made the distinction between “treatable” and “untreatable” ones: the former occur in “a patterned, rule-governed way” and include problems with verb tense or form, subject-verb agreement, run-ons, noun endings, articles, pronouns, while the latter include a variety of lexical errors, problems with word order and sentence structure, including missing and unnecessary words. Explaining that there is no handbook or set of rules to consult in order to avoid or fix those types of errors, she opted, in part, for direct correction hoping it “would, if nothing else, provide input for acquisition of these idiomatic forms” (p. 6). Noting that 50% of all errors she identified in her students’ compositions were “untreatable,” she argued that “ESL writing teachers would do well to give much more thought to how they provide error feedback regarding these different types of language forms and structures” (p. 6).

This study attempts to build on existing research into error treatment and especially the role Google can play in stimulating language awareness and enhancing self-editing

skills. “Untreatable” errors arguably occur when students are trying to emulate native speakers, working with their interlanguage, building on it using their acquired knowledge of rules and repository of words and expressions to formulate increasingly complex occurrences. The issue at stake is thus to find out if, during a self-correcting process, EFL learners can search the web and use raw online data, breaking down snippets of texts featured in Google search results, identifying and using various expressions and inherent language patterns to bring changes to their own non-native-like formulations.

## 2. Method

### 2.1. Participants

The *classes préparatoires aux grandes écoles section EC*, commonly called *prépa EC*, consist of two selective years preparing post-secondary students for competitive entry exams to France’s business schools. The program includes three hours of English teaching per week and consists in writing argumentative essays, answering reading comprehension questions, and translating newspaper articles and short excerpts from contemporary novels. The participants were 17 second-year French *prépa EC* students from a French *lycée*: 12 male and 5 female with an average age of 19 years. They all had French L1, had received at least six years of English instruction, and their levels varied from upper-intermediate to advanced (B2-C1). Since the beginning of their first year, they had been encouraged to read the press in their own time in order to complement the work done in class and gain a sense of self-direction, a key to learning languages and to learning how to learn languages (Holec 1980, 1981). It is generally agreed that autonomy cannot be taught and learned but only fostered and developed (Benson 2003:290) and the students were thus trained to scan newspaper articles in search of noteworthy linguistic material and also encouraged to compile their own lists of words and expressions spotted during in- and out-of-class “treasure-hunting activities” (Kennedy & Miceli 2010: 6).

### 2.2. Procedure

During the first step of the experiment, students were introduced in class to a customized search engine restricting searches to 28 information websites created using Google Custom Search (see Table 1), a service launched by Google in 2006 allowing creators to select what websites will be used to search for information, thus eliminating any unwanted websites and limiting the amount of potentially unreliable results. A set of explicit guidelines introduced students to working with Google by showing them how to perform simple and more advanced search options. It consisted of a description of the various search options, a series of search results screenshots, and sample corrections of untreatable errors performed with the help of the search results (details are provided in the next section). During the second step of the experiment, the students wrote two essays, I underlined a number of untreatable errors they contained, and the learners were then instructed to correct them at home using the customized search engine and send me their corrections via email. I then proceeded to analyze the types of searches they had performed, their use of the material featured in the search results and whether the correction was successful or not. At the end of the experiment, the students were given the opportunity to provide feedback on their use of Google Custom Search to self-correct their errors. They provided answers to a questionnaire featuring seven closed questions on a 5-point Likert scale and open questions for additional comments.

<b>Home page</b>	
<a href="http://www.google.fr/cse/home?cx=011764784480104570934:4qgipwv8a2q">http://www.google.fr/cse/home?cx=011764784480104570934:4qgipwv8a2q</a>	
<b>Indexed websites</b>	
<a href="http://www.bostonglobe.com">www.bostonglobe.com</a>	<a href="http://www.uk.wsj.com">www.uk.wsj.com</a>
<a href="http://www.cbsnews.com">www.cbsnews.com</a>	<a href="http://www.usatoday.com">www.usatoday.com</a>
<a href="http://www.chicagotribune.com">www.chicagotribune.com</a>	<a href="http://www.usnews.com">www.usnews.com</a>
<a href="http://www.csmonitor.com">www.csmonitor.com</a>	<a href="http://www.voanews.com">www.voanews.com</a>
<a href="http://www.edition.cnn.com">www.edition.cnn.com</a>	<a href="http://www.washingtonpost.com">www.washingtonpost.com</a>
<a href="http://www.europe-wsj.com">www.europe-wsj.com</a>	<a href="http://www.bbc.co.uk">www.bbc.co.uk</a>
<a href="http://www.ft.com">www.ft.com</a>	<a href="http://www.economist.com">www.economist.com</a>
<a href="http://www.latimes.com">www.latimes.com</a>	<a href="http://www.guardian.co.uk">www.guardian.co.uk</a>
<a href="http://www.newstatesman.com">www.newstatesman.com</a>	<a href="http://www.independent.co.uk">www.independent.co.uk</a>
<a href="http://www.nytimes.com">www.nytimes.com</a>	<a href="http://www.observer.guardian.co.uk">www.observer.guardian.co.uk</a>
<a href="http://www.online.wsj.com">www.online.wsj.com</a>	<a href="http://www.spectator.co.uk">www.spectator.co.uk</a>
<a href="http://www.reuters.com">www.reuters.com</a>	<a href="http://www.telegraph.co.uk">www.telegraph.co.uk</a>
<a href="http://www.thedailybeast.com">www.thedailybeast.com</a>	<a href="http://www.thesundaytimes.co.uk">www.thesundaytimes.co.uk</a>
<a href="http://www.time.com">www.time.com</a>	<a href="http://www.thetimes.co.uk">www.thetimes.co.uk</a>

Table 1. News websites indexed by the customized Google search engine.

### 2.2.1. First step: introducing learners to Google search

In the next two sections, simple and more advanced search options are presented respectively.

**A) Searching for exact words and phrases using quotation marks and wild cards.** Learners were first shown how to use the search engine to solve grammar problems and find collocations and idioms. By using the quotation marks around a search string, Google makes it possible to search for exact word combinations and whole phrases. It is possible, for instance, to compare prepositional constructions such as the number of hits for *"it depends on"* and *"it depends of"* (543,000,000 and 4,420,000 hits respectively) and find the most frequently used form (e.g. Shei 2008a). Another example: if learners are uncertain over the correct way of saying that a task or job requires no effort, they can enter *"it's as easy as"* in the search box and scour the results to find the answer (*it's as easy as pie*, *it's as easy as ABC*, and *it's as easy as falling off a log* being the recurring expressions). But learners can also use a wildcard (\*) in the search string to leave open a slot for one or more words. Entering *"it's a \* step forward"* in the search box enables them to retrieve a variety of adjectives used

with *step forward* in the snippets of text listed by Google. They can then select and compare the number of hits and choose the most frequently used ones (*it's a great step forward* occurs 4,170,000 times, *it's a big step forward* 676,000 times, *it's a major step forward* 496,000 times, and *it's a huge step forward* 319,000 times).

**B) Searching for expressions using word combinations.** In-class training then moved on to more advanced Google searches that rely on word combinations meant to generate snippets of texts that can be explored in search of words and expressions to plunder for use in personal sentences. The rationale behind this was that learners could scour the results and borrow the native-like linguistic material their interlanguage precluded them from formulating themselves, and then weave it into their own formulations. For example, if learners want to write about the need for politicians to implement an assault weapons ban, they were shown that by entering *ban* followed by *assault weapons* in the search box, Google generates a series of results which can then be observed and borrowed from (see Figure 1).

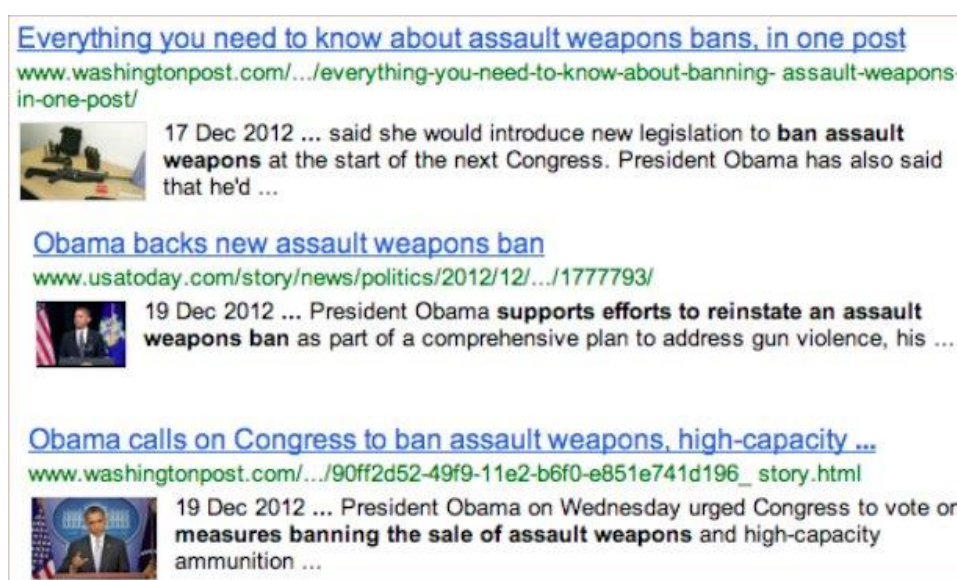


Figure 1. Selected search results for *ban assault weapons*.

Using these examples, it is possible to write a series of forceful arguments like "politicians need to *introduce new legislation to ban assault weapons*" (using the first snippet), "US politicians must make *efforts to reinstate an assault weapons ban as part of a comprehensive plan to address gun violence*" (using the second snippet), and "politicians must *vote on measures banning the sale of assault weapons and high-capacity ammunition*" (using the third snippet).

Another example: if learners are trying to express the idea that immigrants are sometimes discriminated against but don't know how to combine their words, they can enter "immigrants" followed by "scapegoats" (see Figure 2).



Figure 2. Sample search result for "*immigrants scapegoats*".

We see that "Immigrants are scapegoats for high unemployment rates" is one possibility. And using material from one snippet, the learners can then find other



noteworthy elements. Here they can enter the sentence builder "immigrants are scapegoats for" (not forgetting quotation marks) to find how else it is complemented in the press (see Figure 3).



Figure 3. Selected search results for "Immigrants are scapegoats for".

Finally, learners can use Google to check the idiomaticity of their formulations and find alternatives in case they are not native-like. To that end, they can combine the quotation mark search with the keyword search. For example, is it native-like to write "privacy issues involving Google and Facebook"? Entering the expression in the search box with the quotation marks generates no result at all. But it is not the case when the same expression is entered without the quotation marks as Google now lists a series of articles combining the words in one way or another (and not in the exact order we want them to occur as is the case when using the quotation marks). The material featured in the snippets (see Figure 4) can now be used to write alternatives like "Google and Facebook are *involved in an online privacy row*" (using the third snippet, "the latest *privacy rows involving* Facebook and Google") or "Facebook and Google have raised *privacy concerns*" (using the last snippet, "the *privacy concerns raised* by Facebook and Google").

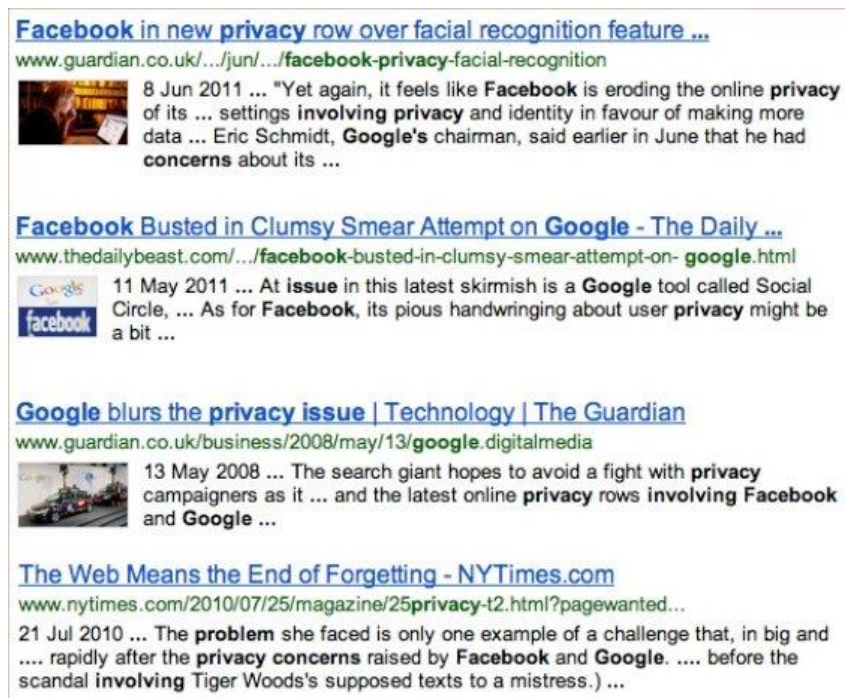


Figure 4. Selected search results for privacy issues involving Google and Facebook.

Following that initial search, the keywords spotted in the original snippets can then be used for a subsequent search. Learners will then be directed to other relevant examples. Entering "online privacy row involving Facebook and Google" (without quotation marks) generates a list of results, among which one formulation clearly stands out (see Figure 5).



Figure 5. Sample search result for *online privacy row involving Facebook and Google*.

#### 2.2.2. Second step: data collection by the instructor, self-correction by the learners

In week one, the students wrote their first in-class essay ("Should society restrict some forms of expression in order to protect its members from violence or hatred?"). The essays were then collected and one to five "untreatable" errors were identified in each of them. All students were then emailed personal charts containing the untreatable errors to be revised and were given one week to correct them on their own using the customized Google search engine. In order to exert some control over their search activities, they were instructed to submit revised passages explaining in detail how they had used Google results to improve their original passages. In week 5, the students wrote a second essay in class ("What do you think about the European Union recently winning the Nobel Peace Prize?"), received their personal charts containing up to five errors and were given one week to submit revised passages explaining the corrections.

### 3. Findings

#### 3.1. Error analysis

A total of 129 untreatable errors were identified in all 34 essays. The total number of segments improved is 67, equivalent to a success rate of 52%. The number of segments for which the correction was not successful is 36 (28%) and the number of segments for which the correction was partly successful is 16 (12.4%). Six errors (4.6%) were left uncorrected or partly so, and in four cases (3%) the students did not specify whether they had used Google in the correction process. The students' personal charts detailing the corrections made with Google Custom Search reveal six types of searches performed by the students (see Table 2 for details). One way for students to correct their errors is to perform searches on fragments of a non-native-like segment containing an untreatable error. They either initiate a direct correction that they check on Google, or use various approaches (wild card search, word combinations, etc.), and they then use elements featured in the snippets to make the necessary corrections (search type #1, used 70 times). Two other strategies consist in formulating queries after consulting a dictionary (search type #2, used 6 times) or using Google's auto-correct (alternate spelling or wording) to revise a segment (search type #3, used 3 times). In other cases the students decide to perform searches on a whole segment (or syntactically whole fragments of it). In the result snippets, they identify elements of the segment they have to correct which they use to make the necessary changes (search type #4, used 19 times). Yet another strategy consists in entering the whole segment (or syntactically whole fragments of it) in the search box. In the result snippets, although the students do *not* see elements of the segment they have to correct, Google lists articles dealing with their topic. In the snippets of text they then identify what they need to correct themselves (search type #5, used 12 times). Finally, the students sometimes perform keyword searches to which Google responds by listing articles

dealing with their topic. The students then use elements featured in the snippets to correct their segments (search type #6, used 10 times).

Search type #1		
Original segment	Revised segment	Comments
Even if war is no more a reality in Europe, there is no denying that <i>the economical war has replaced it</i> .	Even if war is no more a reality in Europe, there is no denying that <i>Europe is in an economic war now</i> .	<p>1. I first entered <i>economical war</i> in the search box and Google's auto-correct offered <i>economic war</i> as an alternative.</p> <p>2. I then entered <i>economic war</i> and saw that David Cameron once said <i>Britain is in an economic war</i>. So I used the whole expression instead of my original segment.</p>

Search type #2		
Original segment	Revised segment	Comments
<i>The liberty of expression is necessary in democratic countries but we must warn to violence.</i>	We must take steps to prevent such violence / We must pay attention to violence	I used an online dictionary to check how to say <i>faire attention à</i> in English. I then used GCS to check my correction.

Search type #3		
Original segment	Revised segment	Comments
EU is one of <i>the hugest weapons solder of the world</i> .	EU is one of <i>the biggest weapons soldier of the world</i> .	I entered the segment and Google's auto-correct offered an alternative, <i>EU is one of the biggest weapons soldier of the world</i> .

Search type #4		
Original segment	Revised segment	Comments
Freedom is the backbone of the driving force behind a "good society."	Freedom is the backbone of <b>AND</b> the driving force behind a "good society."	I entered the sentence and found a snippet making me realize that "the backbone of" and "the driving force behind" were two different expressions.

Search type #5		
Original segment	Revised segment	Comments
The newspaper Charlie Hebdo published <i>some comics which critic Islam</i> .	The newspaper Charlie Hebdo published <i>some cartoons that mocked Islam</i> .	I entered the whole passage and saw that <i>cartoons</i> was more appropriate than <i>comics</i> . I saw a better sentence than mine in the first snippet and so I used it.

Search type #6		
Original segment	Revised segment	Comments
The recent scandals in Iraq about prisoners detention.	The Iraq prison abuse scandal.	I entered <i>Iraq scandals detention</i> and found what I needed.

Table 2. Sample search types and comments.

The general coding of errors (see Table 3) reveals that the students are very creative, sometimes combining various search methods (e.g. student #13, error #8), or have an obvious predilection for one type of error correction (e.g. student #5 mainly using search type #1).

Student #	Error #									
	1	2	3	4	5	6	7	8	9	10
1	4 +	X PB3	4 +	X PB2	- PB1	4/5 ±	3 -	- PB1		
2	5 -	1 +	1 +	5 +	- PB1	? ±	1 +	1 +	1 +	
3	1 -	4 +	1 ±	2 -	2 +	4 -	1 -	1 -		
4	1 +	1 -	4 +	1 -	4 +	1 -	1 -			
5	1 +	4 -	??	1 +	1 +	1 -	1 +	1 +	1 +	? -
6	1 -	3 -	1/5 +	1 +	5/1 +					
7	1 +	1 +	1 +	4 -	1 +	1 +	2 -			
8	1 +	1 +	X	4 +	1 -	2 +				
9	4/1 ±	5 +	1 +	? -	1 +	X PB2				
10	6 +	6 +	1 +	6 +	1/6 ±	6 +	1 +			
11	1 +	??	1 -	1 -	1 ±	1 -	1 -	6 +	1 +	
12	4 +	3 +	X	? -	6 ±	5 ±	1 -			
13	? -	1 ±	1 -	1 -	1 +	1 -	1 ±	1/4/5/2 +	2 +	

<b>14</b>	4 +	1 ±	5 +	5 +	4/1 ±	6 ±	6 ±	1 +	4/1 +	6 +
<b>15</b>	1 +	X	1 +	1 +	1 ±					
<b>16</b>	? +	1 +	1 +	? -	1 +	1 +	1 +	1 +		
<b>17</b>	4/5 +	5/1 -	4 +	4 +	??	? -	??	4/1 ±		

Table 3. General error coding.

Note: The errors were identified in essays 1 and 2. To correct each error, the students performed various search types. Each search type number (1 to 6) is followed by a positive (+), a negative (-), or a plus-minus (±) sign depending on whether the correction was successful, not successful, or partly successful. The students sometimes combine various search methods, hence the succession of numbers in some cases (cf. student #13, error #8). A question mark (?) is used when the correction is not explained although a Google search was performed. Two question marks (??) are used when the correction is not explained and there is no indication that a Google search was performed, and a cross (X) is used when the segment is left uncorrected. PB1 is used when students initiate a correction after entering the whole segment in the search box and say they do not know how to use the results. PB2 is used when students say they do not know what query to formulate, and PB3 when they see elements in the search results but do not know how to use them.

### 3.2. Feedback on Google-driven language learning

Sixteen completed questionnaires were returned via email (the responses to the seven 5-point Likert-scale questions are given in Table 4). Questions 1 to 4 show that a majority of students felt comfortable with the use of basic Google search options. Question 5 indicates that the students view Google use as a good way to correct their errors and improve their English, and question 6 indicates that a majority view it as a good way to find native-like formulations in the search results. However, only nine students said that they intended to use it in the future for linguistic purposes. In the answers they provided to the open-ended questions the students explained in more detail what they liked about Google search but also raised a number of issues.

Eight students explained that the main difficulty for them was to find appropriate ways to formulate their queries. They sometimes found it difficult to identify alternatives to their non-native-like formulations because they couldn't think of any other word or expression to enter in the search box. Three of them argued that in order to use Google effectively, it is necessary for them to know what they are looking for, which implies knowing what is wrong in a segment underlined by the teacher. Other students explained that they liked how Google Custom Search could be used to discover word combinations and noteworthy formulations. One for example said she enjoyed using Google to check the idiomaticity of formulations by using quotation marks around search strings. Another student liked the idea of restricting searches to specific websites, while another one enjoyed making serendipitous discoveries when scouring the snippets of text. Two of them, however, said that they found it more effective to read newspaper articles to find noteworthy formulations. Three others said they sometimes found it tedious to have to use a search engine to correct their errors while they had other, more effective tools at their disposal (grammar handbooks, dictionaries, etc.). Two of them in fact said that they used Google Custom Search in conjunction with online dictionaries. Two others confessed they found it difficult to adapt the search results to have them fit into their original sentences. They also said it was a little frustrating to find ideas that did not exactly express the ideas they had in mind

although they constituted obvious alternatives to their original non-native-like formulations. Three students said that they sometimes felt overwhelmed with the results and simply did not know what to make of them.

<b>Closed questions (5-point Likert scale)</b>	<b>1</b> strongly disagree	<b>2</b> disagree	<b>3</b> neither agree nor disagree	<b>4</b> agree	<b>5</b> strongly agree
1. I find it easy to use Google search options.	0	12,5 %	6,25 %	31,25 %	50 %
2. I can differentiate between searches using quotation marks and searches not using quotation marks.	0	6,25 %	6,25 %	18,75 %	68,75 %
3. I know how to use wild cards in my queries.	0	6,25 %	25 %	31,25 %	37,5 %
4. I know how to use keywords in my queries.	0	6,25 %	0	43,75 %	50 %
5. I think that using Google Custom Search is a good way to correct my errors and improve my English.	0	6,25 %	12,5 %	68,75 %	12,5 %
6. I think that using Google Custom Search is a good way to find native-like formulations used in the press.	0	6,25 %	12,5 %	37,5 %	43,75 %
7. I intend to use Google (Custom Search) in the future for linguistic purposes.	6,25 %	6,25 %	31,25 %	50 %	6,25 %

Table 4. Responses to the 5-point Likert scale questions.

#### 4. Discussion

The purpose of this study was to document the way in which internet searches can act as "a tool helping second language writers make decisions about their writing" (Acar et al. 2010: 6). It can now be argued that using Google Custom Search and restricting searches to information websites is a way to increase the reliability of raw online data in so far as it maximizes the students' chances to be exposed to grammatically accurate English. For teachers who generally choose to reformulate "untreatable" passages in their students' papers, this can surely "help relieve [them] of the need to act as proofreading slaves" (Milton 2006: 125). One student for example said he found that Google was a good way to go about correcting his errors when the teacher was not around. So it seems that Google acts as a gateway to a repository of formulations that they can choose by themselves instead of relying on their teacher to provide alternatives. However, some students confessed they sometimes felt overwhelmed with the results or did not know how to formulate their queries. Several studies bearing on corpus use have reported that students feel frustrated (Lavid, 2007) or overwhelmed by considerable amounts of data (Ädel, 2010; Johns et al., 2008; Liu & Jiang, 2009; Kennedy & Miceli, 2010). Others said they found it difficult to formulate corpus queries and various studies also report on the same problem (Ma, 1994; Kennedy & Miceli, 2001; Miceli & Kennedy, 2002; Sun, 2003; Cheng et al., 2003; O'Sullivan & Chambers, 2006; Hafner & Candlin, 2007). Others still explained that analyzing Google output was no easy task, another recurring problem in studies documenting learner analysis of concordancer output (Ma, 1994; Bowker, 1998; Kennedy & Miceli, 2001; Miceli &

Kennedy, 2002; Cheng et al., 2003; Sun, 2003; Yoon & Hirvela, 2004; Lavid, 2007; Johns et al., 2008; Boulton, 2009b; Liu & Jiang, 2009; ). The challenge for teachers is thus to provide learners with appropriate training and make sure they are “adequately equipped” (Kennedy & Miceli, 2001: 81) before exploring corpora on their own.

When working on Google output, teachers are also faced with the difficult task of encouraging learners to assimilate the formulations they identify because they will inevitably risk being stigmatized for working too closely with their sources and accused of plagiarism. Donahue (2008) points to this major problem that language teachers are grappling with and makes the case that copying should nonetheless not be castigated as plagiarism:

*How do we determine at what point something is “owned”? [...] Students come to learn and we want them to appropriate knowledge and be comfortable in the discourse of the field; at what point does something —class discussion, a professor’s discourse— no longer get cited? (p.102)*

We can indeed wonder what students are supposed to make of what they read in their own time. Where to draw the line between what ought to be copied and what ought not to be? If we take a sentence like *Human cloning may be the thin end of the wedge*, it is difficult to decide whether or not, if a student reads it in a news article and subsequently uses it in an essay, the accusation of micro-plagiarism is justified. Research on the subject (e.g. Grossberg 2008; Murray 2008; Emerson 2008; Senders 2008; Bloom 2008; Bloch 2008; Adler-Kassner et al. 2008) explains that accusations of plagiarism are most often sweeping generalizations of otherwise skillful use of appropriated material. It may not be really fair to accuse students who borrow and use without referencing of intellectual theft as, when copying, they are learning to situate their discourse in relation to others’. Within the framework of this experiment, it has been shown that selective reading of Google results is a way for EFL students to write better English by skillfully copying and integrating prefabricated ideas and language into their own essays. The students never transfer extensive verbatim passages to their essays but select relevant multi-word fragments and the result is language hybridity (i.e. a combination of material identified in Google snippets and personal utterances). And while it is difficult to decide whether or not Google search is a tool helping EFL learners gain in grammatical accuracy, it is a way for them to find alternatives to their non-native-like formulations. The keyword search, used by many students, is particularly effective to that end.

For example, seeking to improve *a cartoonist who draws Mahomet*, student #10, who is writing about a scandal which recently flared up in France, enters *who draws Mahomet* and realizes that the result snippets feature the word *cartoon*. He then performs a search with a series of three keywords, *charlie hebdo cartoon* (*Charlie Hebdo* being the name of the newsweekly which originally published the controversial cartoons), and finds a *satirical weekly publishes cartoons of the Prophet Mohammed*, which he decides to use to rephrase his original idea. The same student, trying to improve *The contestation wave in Middle East against a disgusting film*, explains that he knew that *contestation wave* was incorrect yet could not come up with anything better when writing his in-class essay. So he explains that entering *protesters middle east* in the search box resulted in Google producing a link to a *New York Times* article whose title (“Protests spread in the Middle East”) he used to correct his sentence.

A successful keyword search is thus arguably the first step on the road to writing clarity. Yet it is obvious that it does not solve other problems that the students also have to tend to. When the same student uses *publishes* (instead of *published*) to refer to a scandal which erupted a few months ago, it is difficult to decide whether or not he is

aware that *spread*, which is transferred to the original essay, is used in the present tense and not the simple past in the title. In a word, while it is obvious that the students generally do recognize what they need when they see it in Google results, they are not always successful at accommodating the syntax of the segments they seek to weave into or substitute for their original written productions.

Student #1, for instance, writing about free speech and asked to improve *If the society do not established a red border, it can be a vicious circle*, explains that he doesn't know how to use Google to improve the sentence. He performs a search with the entire sentence and doesn't break it down to explore meaningful elements (e.g. *society establish a red border*) to find out if they are combined in a particular way or if Google lists articles dealing with the topic, featuring expressions that can be borrowed. In most cases, this shows that the students must already have a repository of alternatives they can use to perform their searches. These alternatives don't need to be whole syntactical segments but can be collocations or single lexical items that the student is not sure how to articulate in a complete sentence. For instance, if students realize that *establish a red border* is incorrect but know the expression *draw the line*, they can perform a search meant to find out how it is contextualized in the press. Furthermore, in order to maximize their chances of finding what they need, the students must also be able to self-correct a number of treatable errors first (i.e. write *if society does not establish* and not *if the society do not established* in the example). Indeed, Google is more likely to produce relevant examples when searches are performed with grammatically accurate, albeit awkwardly formulated, segments. In other cases, it was found that the students did make changes but on some elements only. In other words, they did not see what was wrong in their sentences. For example, student #5, asked to improve *freedom of expression is being turned into ideological injures* only corrects *injures*, opting for *injuries*, unaware that *ideological injuries* is an unlikely collocation and that it is in fact the whole idea that needs to be reformulated.

## 5. Conclusion

The web should not be dismissed as an unreliable source of data. Although it is arguably not a corpus, EFL learners can nonetheless profitably use Google for quick and easy access to authentic language in the form of selected passages from a great number of articles. In that sense, Google output is very much adapted to students who need to keep up with world events and whose ultimate goal is to emulate the language of the press. Depending on their competence, it is a vast repository of formulations that they can identify and borrow for further use in their own writing. Students can be given a significant linguistic boost if encouraged to plunder formulations featured in Google results. Such an approach implies for the students to go through an initial stage of teacher-controlled imitation (or micro-plagiarism) because initially copying native speakers will, arguably, make it possible to emulate them.

The rationale behind customizing a search engine to explore linguistic material from a selection of online newspapers is in keeping with Tribble's recommendation that the most useful corpus for EFL learners is "the one which offers a collection of expert performances in genres which have relevance to the needs and interests of the learners. Collections of relevant expert performances will exemplify the results of the desired forms of language behavior that learners are trying to achieve" (1997: n. pag.). The main objection raised by a certain number of students who took part in this study was that they sometimes felt overwhelmed with search results or could not think of ways to formulate their queries. Further research could thus profitably focus on how best to train EFL learners to use Google search results in order to self-edit.



## Websites

1. <http://www.webcorp.org.uk/live>
2. <http://www.kwicfinder.com>
3. <http://webascopus.org>
4. <http://www.greenstone.org>
5. <http://www.google.com/cse>

## References

- Acar, A., Geluso, J. & Shiki, T. (2011). How can search engines improve your writing *CALL-EJ*, 12 (1): 1-10.
- Ädel, A. (2010). Using corpora to teach academic writing: challenges for the direct approach. In: Campoy-Cubillo, M. C., Belles-Fortuño B. & Gea-Valor M. L. (eds). *Corpus-based Approaches to ELT*. London: Continuum, 39-55.
- Adler-Kassner, L., Anson, C.M. & Howard, R.M. (2008). Framing plagiarism. In: Eisner, C. and Vicinus, M. (eds.), *Originality, imitation, and plagiarism: teaching writing in the digital age*. Michigan: The University of Michigan Press, 231-247.
- Bates, L., Lane, J., & Lange, E. (1993). *Writing clearly: responding to ESL compositions*. Boston: Heinle & Heinle.
- Benson, P. (2001). *Teaching and researching autonomy in language learning*. Harlow: Pearson Education.
- Bergh, G. (2005). Min(d)ing English language data on the web: what can Google tell us? *ICAME journal*, 29: 25-46.
- Bhatia, T. K. & Ritchie, W. C. (2009). Second language acquisition: research and application in the information age. In: Ritchie, W.C. and Bhatia, T.K. (eds.), *The new handbook of second language acquisition*. Bingley: Emerald, 545-565.
- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of second language writing*, 17 (2): 102-118.
- Bitchener, J., Young, S. & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of second language writing*, 14: 191-205.
- Bloch, J. (2008). Plagiarism across cultures: is there a difference? In: Eisner, C. and Vicinus, M. (eds.), *Originality, imitation, and plagiarism: teaching writing in the digital age*. Michigan: The University of Michigan Press, 219-231.
- Bloom, L. Z. (2008). Insider writing: plagiarism-proof assignments. In: Eisner, C. & Vicinus, M. (eds.), *Originality, imitation, and plagiarism: teaching writing in the digital age*. Michigan: The University of Michigan Press, 208-219.
- Boulton, A. (2009a). Data-driven learning: reasonable fears and rational reassurance. *Indian journal of applied linguistics*, 35(1):81-106.
- Boulton, A. (2009b). Corpora for all? Learning styles and data-driven learning. In: M. Mahlberg, González-Díaz, V. & C. Smith, C. (eds.), *Proceedings of the 5th Corpus Linguistics Conference*. Liverpool: UCREL.
- Boulton, A. (2012). What data for data-driven learning? *EUROCALL 2012: Proceedings*. Nottingham: The University of Nottingham.

- Bowker, Y. (1998). Using specialized monolingual native-language corpora as a translation resource: a pilot study. *Meta*, 4: 631-651.
- Chambers, A., Conacher J. & Littlemore J. (eds.) (2004). *ICT and language learning: integrating pedagogy and practice*. Birmingham: University of Birmingham Press.
- Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of student writing. *Journal of second language writing*, 12(3): 267-296.
- Cheng, W., Warren, M., & Xun-feng, X. (2003). The language learner as language researcher: putting corpus linguistics on the timetable. *System*, 31: 173-186.
- Chen, Y. H. (2004). The use of corpora in the vocabulary classroom. *The internet TESL journal*, 10(9): n. pag.
- Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language learning and technology*, 14(2): 30-49.
- Chinnery, G. M. (2008). You've got some GALL: Google-assisted language learning. *Language learning and technology* 12(1): 3-11.
- Clerehan, R., Kett, G. and Gedge, R. (2003). Web-based tools and instruction for developing it students' written communication skills. In: *Exploring Educational Technologies Conference Proceedings*. Monash University. Retrieved from [http://www.monash.edu.au/groups/flt/eet/full\\_papers/clerehan.pdf](http://www.monash.edu.au/groups/flt/eet/full_papers/clerehan.pdf). Last accessed 25/09/2014.
- Conroy, M. (2010). Internet tools for language learning: university students taking control of their writing. *Australasian Journal of educational technology*, 26(6): 861-882.
- Donahue, C. (2008). When copying is not copying: plagiarism and French composition scholarship. In: Eisner, C. and Vicinus, M. (eds), *Originality, imitation, and plagiarism: teaching writing in the digital age*. Michigan: The University of Michigan Press, 90-103.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Emerson, L. (2008). Plagiarism, a Turnitin trial, and an experience of cultural disorientation. In: Eisner, C. and Vicinus, M. (eds.), *Originality, imitation, and plagiarism: teaching writing in the digital age*. Michigan: The University of Michigan Press, 183-195.
- Ferris, D. R. (2004). The "grammar correction" debate in L2 writing: where are, and where do we go from here? (and what do we do in the meantime...?). *Journal of second language writing*, 13 (1):49-62.
- Ferris, D. R. and Roberts, B. (2001). Error feedback in L2 writing classes: how explicit does it need to be? *Journal of second language writing*, 10(3): 161-184.
- Ferris, D. R. (1999). The case for grammar correction in L2 writing classes: a response to Truscott (1996). *Journal of second language writing*, 8(1): 1-11.
- Fletcher, W. H. (2004). Making the web more useful as a source for linguistic corpora. In: Connor, U. and Upton, T. (eds.), *Applied corpus linguistics: A multidimensional perspective*. Amsterdam: Rodopi, 191-205.
- Gaskell, D. & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System*, 32(3): 301-319.
- Grossberg, M. (2008). History and the disciplining of plagiarism. In: Eisner, C. and Vicinus, M. (eds.), *Originality, imitation, and plagiarism: teaching writing in the digital age*. Michigan: The University of Michigan Press, 159-173.

- Guo, S. & Zhang, G. (2007). Building a customised Google-based collocation collector to enhance language learning. *British journal of educational technology*, 38(4): 747-750.
- Hafner, C. A. & Candlin, C. N. (2007). Corpus tools as an affordance to learning in professional legal education. *Journal of English for academic purposes*, 6(4): 303-318.
- Hendrickson, J. M. (1978). Error correction in foreign language teaching: recent theory, research, and practice. *The modern language journal*, 62(8): 387- 398.
- Holec, H. (ed.) (1988). *Autonomy and self-directed learning: present fields of application*. Strasbourg: Council of Europe.
- Holec, H. (1980). Learner training: meeting needs in self-directed learning. In: Altman, H. B. & James, C. V. (eds.). *Foreign language learning: meeting individual needs*. Oxford: Pergamon, 30-45.
- Hyland, F. (2003). Focusing on form: Student engagement with teacher feedback. *System*, 31(2): 217-230.
- Jarvis, H. (2004). Investigating the classroom applications of computers on EFL courses at higher education institutions in the UK. *Journal of English for academic purposes*, 3(2): 111-137.
- Johansson, S. (2009). Some thoughts on corpora and second-language acquisition. In: Aijmer, K. (ed.). *Corpora and language teaching*. Amsterdam: John Benjamins, 33-44.
- Johns, T. (1988). Whence and whither classroom concordancing? In: Bongaerts, P., De Haan, P., Lobbe, S. & Wekker, H. (eds.), *Computer applications in language learning*. Dordrecht: Foris, 9-27.
- Johns, T. (1990). From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria*, 10: 14-34.
- Johns, T., Lee H. C. and Wang L. (2008). Integrating corpus-based CALL programs in teaching English through children's literature. *Computer Assisted Language Learning*, 2: 483 -506
- Johnson, A. (2004). Creating a writing course utilizing class and student blogs. *The internet TESL journal* 10(8).
- Kennedy, C. & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language Learning and Technology*, 5: 77-90.
- Kennedy, C. & Miceli, T. (2010). Corpus-assisted creative writing: introducing intermediate Italian learners to a corpus as a reference resource. *Language learning and technology*, 14(1): 28-44.
- Kenworthy, R. C. (2004). Developing writing skills in a foreign language via the internet. *The internet TESL journal*, 10(10).
- Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the development of second language writing skills. *The modern language journal*, 75(3): 305-313.
- Kilgariff, A. (2001). Web as corpus. In: Rayson, A., Wilson, T., McEnery, A., Hardie & Khoja, S. (eds.), *Proceedings of the corpus linguistics 2001 conference*. Lancaster: UCREL, 342-344.
- Krajka, J. (2000). Using the internet in ESL writing instruction. *The Internet TESL Journal*, 6(11).
- Lavid, J. (2007). Contrastive patterns of mental transitivity in English and Spanish: a student-centred corpus-based study. In: Hidalgo, E. Quereda, L. & Santana J. (eds.). *Corpora in the foreign language classroom*. Amsterdam: Rodopi, 237-252.

- Liu, D. & Jiang, P. (2009). Using a corpus-based lexicogrammatical approach to grammar instruction in EFL and ESL contexts. *The Modern Language Journal*, 93: 61-78.
- Ma, B. K. C. (1994). Learning strategies in ESP classroom concordancing: an initial investigation into data-driven learning. In Flowerdew, J. & Tong, A. (eds.). *Entering Texts*. Hong Kong: Language Centre, The Hong Kong University of Science and Technology, 197-214.
- Mansor, N. (2007). Collaborative learning via email discussion: strategies for ESL writing classroom. *The Internet TESL Journal*, 13(3).
- McCarthy, M. (2008). Accessing and interpreting corpus information in the teacher education context. *Language Teaching*, 41(4): 563-574.
- Miceli, T. & Kennedy, C. (2002). An Apprenticeship with the CWIC Corpus: a tool for learner writers in Italian. In: Kennedy, C. (ed.) *Proceedings of Workshop Innovations in Italian Teaching*. Brisbane: Griffith University, 83-94.
- Milton, J. (2006). Resource-rich web-based feedback: helping learners become independent writers. In: Hyland, K. and Hyland, F. (eds.), *Feedback in second language writing*. New York: Cambridge University Press, 123-139.
- Murray, L. J. (2008). Plagiarism and copyright infringement: the cost of confusion. In: Eisner, C. & Vicinus, M. (eds.), *Originality, imitation, and plagiarism: teaching writing in the digital age*. Michigan: The University of Michigan Press, 173-183.
- O'Keeffe, A., McCarthy, M. & Carter, R. (2007). *From corpus to classroom: language use and language teaching*. Cambridge: Cambridge University Press.
- O'Sullivan, Í. & Chambers, A. (2006). Learners' writing skills in French: corpus consultation and learner evaluation. *Journal of Second Language Writing*, 15: 49-68.
- Palfrey, J., & Gasser, U. (2008). *Born digital. Understanding the First Generation of Digital Natives*. New York: Basic Books.
- Renouf, A. (2003). WebCorp: providing a renewable data source for corpus linguists. *Language and computers*, 48: 39-58.
- Robb, T. (2003a). Google as a quick 'n dirty corpus tool. *TESL-EJ*, 7(2).
- Robb, T. (2003b). Google as a corpus tool? *ETJ Journal*, 4(1).
- Rundell, M. (2000). The biggest corpus of all. *Humanising language teaching*, 2(3).
- Schroeder, A., Minocha, S., & Schneider, C. (2010). The strengths, weaknesses, opportunities and threats of using social software in higher and further education teaching and learning. *Journal of Computer Assisted Learning*, 26: 159-174.
- Senders, S. (2008). Academic plagiarism and the limits of theft. In: Eisner, C. & Vicinus, M. (eds.), *Originality, imitation, and plagiarism: teaching writing in the digital age*. Michigan: The University of Michigan Press, 195-219.
- Shei, C. (2008a). Discovering the hidden treasure on the internet: using Google to uncover the veil of phraseology. *CALL*, 21(1): 67-85.
- Shei, C. (2008b). Web as corpus, Google, and TESOL: a new trilogy. *Taiwan Journal of TESOL*, 5(2): 1-28.
- Sun, Y. (2003). Learning process, strategies and web-based concordancers: a case study. *British journal of educational technology*, 34(5): 601-613.

Tribble, C. (1997). Improvising corpora for ELT: quick-and-dirty ways of developing corpora for language teaching. In: Melia, J. & Lewandowska-Tomaszczyk, B. (eds.) *PALC 97 Proceedings*, Lodz: Lodz University Press.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language learning*, 46(2): 327-369.

Widdowson, H. (2000). On the limitations of linguistics applied. *Applied Linguistics*, 21(1): 3-25.

Wu, S., Franken, M., & Witten, H. (2009). Refining the use of the web (and web search) as a language teaching and learning resource. *CALL*, 22(3): 249-268.

Yoon, H. (2008). More than a linguistic reference: the influence of corpus technology on L2 academic writing. *Language learning and technology*, 12(2): 31-48.

Yoon, H. & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of second language writing*, 13: 257-283.

Zamel, V. (1985). Responding to student writing. *TESOL Quarterly*, 19(1): 79-97.

## Article:

### ***Constructing an evidence-base for future CALL design with 'engineering power': The need for more basic research and instrumental replication***

Zöe Handley

Department of Education, University of York, UK

---

[zoe.handley@york.ac.uk](mailto:zoe.handley@york.ac.uk)

#### **Abstract**

This paper argues that the goal of Computer-Assisted Language Learning (CALL) research should be to construct a reliable evidence-base with 'engineering power' and generality upon which the design of future CALL software and activities can be based. In order to establish such an evidence base for future CALL design, it suggests that CALL research needs to move away from CALL versus non-CALL comparisons, and focus on investigating the differential impact of individual attributes and affordances, that is, specific features of a technology which might have an impact on learning. Further, in order to help researchers find possible explanations for the success or failure of CALL interventions and make appropriate adjustments to their design, it argues that these studies should be conducted within the framework of Second Language Acquisition (SLA) theory and research. Despite this, a recent review of research examining the effectiveness of CALL in primary and secondary English as a Foreign Language (EFL) found that CALL vs. non-CALL comparisons are still common and studies focusing on individual coding elements are rare. Further, few studies make links with SLA and few measure linguistic outcomes using measures developed in the field of SLA. One reason for this may be poor reporting of methods and difficulty in obtaining the instruments used in SLA research. Reporting guidelines and the use of the IRIS database ([www.iris-database.org](http://www.iris-database.org)) are introduced as possible solutions to these problems.

**Keywords:** Research methods, basic research, second language acquisition, replication, instruments.

#### **1. Introduction**

More basic Computer-Assisted Language Learning (CALL) research —and replications thereof— is required to permit researchers to construct a reliable evidence-base with 'engineering power' for the design of future CALL software and activities. An evidence-base with 'engineering power' is one that is sufficiently specific that it translates into CALL designs which work in practice (Burkhardt and Schoenfeld, 2003). Basic research refers to studies which provide insights into what specific features of digital environments create conditions and engage learners in processes that promote Second Language Acquisition (SLA), as well as what task variables promote SLA (Pederson, 1987).

My experience of synthesizing the literature in the field (Macaro, Walter & Handley, 2012), however, suggests that CALL research like educational research (Burkhardt & Schoenfeld, 2003) and SLA research (Porte, 2013) more broadly, is failing to achieve this and what we have instead is an accumulation of studies whose findings cannot easily be connected to those of other studies in the broader field of SLA or even within CALL itself.

Firstly, broad atheoretical CALL versus non-CALL comparisons of 'pen-and-paper' versus 'traditional' classroom activities are still common in the CALL evidence-base (Macaro et al., 2012) despite Pederson's (1987) call for them to "forever be abandoned" (p. 125). There are two reasons that this is a concern. First such studies do not have 'engineering power' because they fall into the trap of equating medium with method (ibid.). That is, they fail to acknowledge the fact that a particular technology might be used in a variety of different ways to support language learning and to implement a variety of different approaches to and methods of language teaching (Garrett, 1991) and fully exploit the added value of new technologies (Yildiz & Atkins, 1993): "technology is often used to change and expand the intended learning outcomes rather than to increase the level of performance in exactly the same areas as those targets by classroom instruction" (Chapelle, 2010, p. 70). With respect to the latter, in CALL research the possibility to engage in language learning activities 'anytime, anywhere' through the use of mobile technologies has been exploited to implement spaced vocabulary learning (Lu, 2008) and to contextualise vocabulary learning, that is, adapt it to the learners' immediate local environment (Chen & Li, 2010; Hwang & Chen, 2013; Gutiérrez-Colon et al., 2013).

Broad atheoretical CALL versus non-CALL comparisons also do not have explanatory power: the experimental condition often differs in multiple ways from the control condition and it is consequently not possible to determine to which feature of the software any observed differences should be attributed. O'Hara & Pritchard's (2008) evaluation of the impact of preparing a hyperlinked multimedia PowerPoint report on students' breadth of vocabulary knowledge illustrates this point well. In this study, production of PowerPoint reports with access to on-line resources was compared with production of pen-and-paper reports with access to paper-based classroom resources. The experimental condition, in other words, differed in two ways from the control condition: the medium in which the report was produced (PowerPoint vs. pen-and-paper) and access to resources (online vs. classroom). It is impossible therefore to know whether the higher levels of vocabulary knowledge observed in the experimental group should be attributed to the medium in which the report was produced or to access to online resources.

Secondly, the majority of CALL research is not grounded in SLA theory (Macaro et al., 2012). Grounding CALL research in SLA theory helps researchers to identify possible explanations for the effectiveness of particular manipulations of CALL environments and subsequently make appropriate adjustments to their design to better support language acquisition (Pederson, 1987).

Thirdly, the outcome measures employed in many CALL studies were developed for the specific purposes of the study in question and often differ from those commonly used in SLA research (Macaro et al., 2012). A combination of multiple-choice questions and ratings of learners' certainty in their choices was used as a measure of fluency of lexical recall in a study investigating the effects of different combinations of multimedia presentation on vocabulary learning (Kim & Gilman, 2008) rather than more widely accepted measures such as response latency, i.e. reaction time, for example. This is problematic because failure to engage in instrumental replication, i.e. to use the same outcome measures as employed in previous research, limits the comparability of studies

(Polio, 2012) and coherence of the discipline (Burkhardt & Schoenfeld, 2003), and is a barrier to meta-analysis (Oswald & Plonsky, 2010; Slavin, 1995). Aggregating the results of quantitative studies, meta-analyses are a key tool in the construction of an evidence-base within a discipline. When outcome measures are not consistently operationalised, they, however, produce less reliable estimates of effects and are challenging to interpret (ibid.). It has therefore been suggested that the principle inclusion criterion for a meta-analyses ought to be the construct validity of measures of the dependent variable: “a meta-analysis focusing on school achievement as a dependent measure must explicitly describe what is meant by school achievement and must only include studies that measure what is commonly understood as school achievement” (Slavin, 1995, p. 13), for example.

Finally, methods are frequently not adequately reported to permit replication (Macaro et al., 2012). In particular, instruments are often not provided (ibid.). Replication is, however, a cornerstone of scientific enquiry, necessary to ensure the construction of a reliable evidence-base (Polio, 2012) which has generality (Burkhardt & Schoenfeld, 2003). Reliability refers to the extent to which the individual findings have been validated through follow-up studies. Generality (or generalizability) refers to the extent to which individual findings have been demonstrated to hold in a wide range of contexts (ibid.). The demonstration of generality is perhaps the most important motivation for replication in CALL and SLA more broadly given the range of contextual variables that might have an impact on language learning (Chun, 2012).

In summary, current approaches to CALL research “are encouraging an *accumulation* of vaguely inter-connected research findings rather than the *construction* of knowledge across independent studies” (Porte, 2013, p. 12, original emphasis) which can be translated into designs for future CALL software and activities. In response to this, in the remainder of this paper, I introduce some of the different forms that basic research and replication might take within the field of CALL, and introduce IRIS ([www.iris-database.org](http://www.iris-database.org)), a digital repository of instruments, materials and stimuli used to elicit data in peer-reviewed research into second and foreign languages, as a resource to facilitate replication and promote the design of comparable studies. First, however, it is necessary to introduce the concept of ‘engineering power’. Where possible, as above, all ideas will be illustrated with examples drawn from Macaro et al.’s (2012) systematic review of research on the use of technology in primary and secondary English as a Foreign Language (EFL) teaching.

## 2. ‘Engineering power’

Like the automotive engineer designing and tuning a Formula 1 racing car, the early CALL researcher designing and optimising a learning environment was faced with a myriad of design options: “how and when to use graphics, sound feedback, branching from one learning task to the next based on learner response or request for new material, and how to display all these coding options accurately and efficiently” (Pederson, 1987, p. 100). To that list today we can add: how and when to provide interaction with other learners and the teacher (e.g. synchronously or asynchronously, one-to-one or many-to-one), how and when to personalise learning (e.g. based on attainment or context/location), and so on. The problem is that the theories that CALL researchers have to draw on such as socio-cultural theory are not sufficiently constrained —do not specify under what conditions the theory applies— and specific to translate into designs for CALL software and activities that work in practice (Burkhardt & Schoenfeld, 2003). In the same way that medium does not equate to method and there are many different ways in which a single technology might be employed to facilitate language learning (see above), there are often many different ways in which a particular theory might be translated into designs for CALL software and CALL activities.



In most studies within the field of CALL, socio-cultural theory has been translated into designs which exploit technology to provide learners access to more able partners (see for example Lund, 2008 and Sasaki & Takeuchi, 2010). It has, however, also been argued that support might be provided through access to appropriate resources as well as access to more able partners (Luckin & Clark, 2011; van Lier, 2004), for example. Grand theories such as socio-cultural theory are therefore not adequate to guide the design of CALL software and activities (Burkhardt & Schoenfeld, 2003). Highly specified 'local' theories which take into account the skill (reading, writing, speaking, or listening) or knowledge (vocabulary, grammar or pronunciation), the learner and the learning context, are rather what is required. In other words, like the 'craft' knowledge that practising teachers construct, such theories would be concrete, contextually rich and linked with practice (Hiebert, Gallimore & Stigler, 2002).

Further, to have "engineering power", the CALL evidence-base needs to have *generality*, that is, "go beyond the specific environment being examined, in order to make a contribution to knowledge of affordances of a technology or language learning processes" (Stockwell, 2012, p. 154). This will only be achieved if we abandon broad CALL versus non-CALL comparisons and focus our research efforts on attributes and affordances which transcend multiple specific technologies (Colpaert, 2010; Pederson, 1987). Attributes refer to features of the computer which have the potential to support and develop cognitive processing, such as symbol systems, multimedia and random access (Colpaert, 2010; Pederson, 1987). Affordances are features of the computer which enable learners to engage in processes that support language learning (Colpaert, 2010). These include the possibility to access authentic materials and interact with individuals and groups in the target language (ibid.). Kim and Gilman's (2008) systematic examination of the differential impact of different combinations of multimedia on learners' retention of vocabulary is a good example of a study with engineering power. It is specific and examines the impact of attributes which transcend a wide variety of technologies.

### 3. Basic CALL research

More basic CALL research is, however, required to allow us to *construct* an evidence-base upon which to design future CALL. Basic research refers to studies designed "to discover something about how students best learn a language", i.e. which "provid[es] explanatory data and add[s] to the theoretical bases for second language learning" (Pederson, 1987, p. 125). In other words, basic CALL research goes beyond evaluation and asks "Why did it work?" in addition to "Did it work?" (Levy & Stockwell, 2006, p. 42) and draws on and contributes to the development of SLA theory. Engaging in basic research, it has been suggested, has two benefits. First trials of complex health education interventions suggest that interventions grounded in appropriate theory are more likely to be effective (Campbell, Fitzpatrick, Haines et al. 2000). Second, where trials are unsuccessful, theory helps researchers identify possible explanations for failure to achieve learning goals and refine the design of interventions, in this case CALL software and activities (Pederson, 1987).

Basic CALL research has tended to take one of three forms: (1) exploratory research, (2) observational research, or (3) narrowly focused experimental research. Exploratory research is characterised by ethnographic studies in which researchers observe and interview students about their naturalistic use of CALL software with a view to generating theories regarding what features of digital environments create conditions and engage learners in processes that promote SLA (Pederson, 1987). An example of an informative ethnographic study is Gruber-Miller & Benton's (2001) examination of the *VRoma MOO (1)* for Latin. In observational studies the processes that students engage in during software use are logged and the relationship between software use and

learning gains is explored. An observational study with engineering power would resemble Proctor, Dalton and Grisham's (2007) investigation of native speakers' and English Language Learners' use of the *Universal Literacy Environment*, but track learners use of the different scaffolds provided at a more fine-grained level than overall frequency of use of scaffolds. Narrowly focused experimental studies isolate out the specific attributes and affordances of a technology which might have a differential impact on learning, and explore hypotheses grounded in SLA theory and research. In other words, narrowly focused experimental studies explore "the relative effectiveness of the pedagogical techniques that [a particular technology] implements, i.e., different types of feedback, online help, textual annotations, glossing formats, etc." (Burston, 2006, p. 258). Kim & Gilman's (2008) investigation of the differential impact of different combinations of multimedia on vocabulary knowledge is a good example of a narrowly focused experimental study. Another example is Dalton et al.'s (2011) comparison of different versions of a reading tutor integrating different forms of support, namely vocabulary versus reading support. It should, however, be noted that both vocabulary support and reading support could be realised in a number of different ways.

All of the above methods have the potential to make a significant contribution to our understanding of the conditions and processes which support SLA, as long as researchers engage with SLA theory and instrumental replication (see below). They are, however, not without their critiques. The value of narrowly focused experimental studies in particular has been questioned:

*The treatment method leads to a danger that all experiments with computers and learning will be failures: either they are trivial because very little happened or they are "unscientific" because something real did happen and too many factors changed at once.* (Papert, 1987, p. 26)

Two 'engineering' approaches to educational research are therefore beginning to attract attention in the field of CALL. These are design-based research (Barab & Squire, 2004; Burkhardt & Schoenfeld, 2003; Yutdhana, 2008) and educational engineering (Colpaert, 2006, 2010). In contrast, with the scientific approach to research upon which conventional methods draw, the engineering approach is transformative. That is, engineering research, like much educational research, is practice-oriented and aims to both understand "how the world works" and "help it to work better" (Burkhardt & Schoenfeld, 2003, p. 5). It achieves this by "us[ing] existing knowledge in experimental development to produce new or substantially improved materials, devices, products, and processes including design and construction" (Higher Education Research Funding Council, 1999, p. 4). Design-based research (also referred to as design experiments and design research) refers to an approach in which 'local' theories of learning and teaching are tested and refined through iterative cycles of design and evaluation in collaboration with end-users, i.e. learners and teachers, and gradually scaled up and rolled out for use in practice (Barab & Squire, 2004; Burkhardt & Schoenfeld, 2003; Gorard, Roberts & Taylor, 2004; Yutdhana, 2005). In other words, in addition to being transformative and practice-oriented, design-based research recognizes and values teacher cognition (see Borg, 2003; Kumaravadivelu, 1994) and is impact-oriented. A study adopting Bannan-Ritland's (2003) Integrative Learning Design (ILD) framework for design-based research, for example, would begin with informed exploration of the learning context and problem, i.e. needs analysis. This phase of the design process, which would also include a review of the literature and identification of appropriate learning theory, would result in a specification for the design of the CALL system. In the next phase of the design process, enactment, would involve translating the requirements to a design and developing a prototype. The local impact of the design would then be evaluated in the next phase, the results of which would lead to adjustments to the design and further

cycles of evaluation. Design-based research aims to produce 'shareable theories' (Design-Based Collective, 2003, p. 5). In the final stage, others would therefore be encouraged to adopt the design and theory to allow evaluation of broader impact. Pardo-Ballester & Rodriguez's (2009, 2010) development of online readings for elementary learners of Spanish for business and engineering, for example, is grounded in design-based research.

Educational engineering as conceived by Colpaert (2006, 2010) is also characterised by iterative cycles of development. The approach, however, is grounded in theories of motivation and the assumption that CALL ought to "support the learner in better achieving learning goals" (Colpaert, 2010, p. 273) and prioritises process over product as outcome measures: "Engineering does not focus on measurable significant differences on a product level, but rather on observable phenomena on a process level" (ibid., p. 262). The departure for design and research within this approach is therefore an examination of learner goals. Having identified learner goals through focus group discussions and compared them with other competing goals and in particular pedagogical goals, appropriate learning theories to operationalise the competing goals are identified and a design for the CALL software and tasks is elaborated. The resulting design and any design and theoretical questions that arise from it are then explored through iterative cycles of design and evaluation as in design-based research. Educational engineering has therefore been characterised as 'slow research' and all of the projects that have adopted this research to date are still on-going. It is therefore not possible to discuss any completed projects at this point. For a list of on-going projects see Colpaert (2010).

In summary, whatever methodology is adopted, drawing links with SLA theory and research is essential to drive the construction of an evidence-base for the design of future CALL software and activities forward. It will "lead to a stronger focus on the learning process rather than the technology" (Stockwell, 2012, p. 160) and, by providing insights into the reasons for the success and failure of CALL software and activities, it helps researchers and developers refine the design of future CALL software.

#### **4. Replication in CALL**

Replication is also required to *construct a reliable* evidence-base with *generality*. Exact replications, in which researchers attempt to copy the original study as closely as possible using identical subjects, conditions, and instruments, among other things, should be conducted where possible to allow the validation of findings (Polio, 2012; Porte & Richards, 2012). Instrumental replications, approximate replications in which the same outcome measures as used in previous research are employed, should be conducted in a range of different contexts to permit the demonstration of the generality of findings and also to permit comparisons and meta-analyses of studies within CALL and in the broader field of SLA (Polio, 2012). Further, conceptual replications in which findings are tested using a different study design, in particular different data collection procedures (e.g. observation versus self-report) are essential to demonstrate the validity of findings, i.e. to demonstrate that they are not artefacts of the original design (Polio, 2012; Porte & Richards, 2012).

Replication in CALL research, as in SLA research more broadly, has, however, largely been neglected, with the exception of a number of studies which have replicated findings of SLA research (Chun, 2012). In fact, some question whether replication is even possible in CALL given the pace of technological advances and the fact that older technologies quickly fall into obsolescence (Chun, 2012). This argument does not, however, hold if we move away from broad CALL versus non-CALL comparisons and

focus our research efforts on the exploration of the impact of attributes and affordances which transcend individual technologies, new and old, as discussed above.

A greater problem, however, is that, as in SLA research more broadly (Polio & Gass, 1997), CALL research is not adequately reported to permit instrumental replication, let alone exact replication (Macaro et al., 2012). Instruments, including background questionnaires, measures of proficiency, instruments for data elicitation and pre- and post-tests, and coding frameworks (Polio & Gass, 1997), are rarely provided in CALL studies, and often barely discussed in the methods sections of research articles (Macaro et al., 2012). While it is always possible to contact authors to request materials, researchers can be difficult to track –they move– and they may not always be able to easily locate materials within their archives (Marsden & King, 2013, Marsden & Mackey 2014).

One way to overcome these problems is to introduce reporting guidelines, as suggested by Polio & Gass (1997, p506):

*[E]xamples of what might ultimately be useful to researchers [include]: (a) Detailed guidelines and examples of coding categories, (b) A listing of examples that were excluded from consideration, (c) Measures of proficiency (descriptions of tests where security is a problem), (d) Instruments for data elicitation, including pre-tests and post-tests, (e) Experimental protocols and instructions to subjects, and (f) Demographic background of subjects.*

Experience in the health sciences has demonstrated that, in addition to permitting replication, the introduction of reporting guidelines has increased the quality of published research (Moher, Jones & Lepage, 2001). Researchers interested in building on Polio & Gass's (1997) suggestions are encouraged to consult the reporting guidelines for relevant forms of research in the health sciences, including CONSORT for randomized controlled trials (Moher, Schulz, & Altman, 2001; [www.consort-statement.org](http://www.consort-statement.org)), i.e. experimental research, STROBE for observational studies ([www.strobe-statement.org](http://www.strobe-statement.org)), and COREQ for qualitative interview-based research (Tong, Sainsbury, & Craig, 2007).

Whether or not reporting guidelines are introduced, barriers to replication will, however, remain. First it will remain difficult to replicate studies which have already been published. Second, it will remain difficult to locate instruments. Articles in electronic databases are typically indexed, that is, assigned thesaurus terms, on the basis of the title and abstract alone. Even if we were to adopt reporting guidelines, it would simply not be possible to provide sufficient information to index instruments in the abstracts of CALL and SLA research articles (see below for a list of the dimensions on which it would be desirable to index instruments for use in CALL and SLA research).

## **5. The IRIS database**

Instruments for Research into Second Language Learning and Teaching (IRIS) is an open access digital repository of materials used to collect data in research on second and foreign language acquisition developed and curated by the Digital Library at the University of York which might help researchers in the fields of CALL and SLA overcome those barriers. All instruments held on the database have been used to collect data for a peer-reviewed publication, i.e. a peer-reviewed journal or conference proceedings, an edited book or a successful doctoral thesis. The database is searchable along a number of dimensions including instrument type, linguistic feature, and learner proficiency, and materials can be downloaded and re-used, with most held under a Creative Commons derivatives allowed non-commercial share-alike licence. In other words researchers "can remix, tweak, and build upon this work non-commercially, as long as [they] credit the

creators of the instrument and license [their] new creations under identical terms” ([www.iris-database.org](http://www.iris-database.org)).

It is also possible for researchers to upload their own instruments to the database for use by other researchers. In fact, 30 top ranking journal editors are now encouraging uploads, including the editors of the following SLA journals: *Applied Linguistics*, *Language Learning*, *Language Teaching*, *Studies in Second Language Acquisition* and *The Modern Language Journal*, as well as *Computer Assisted Language Learning Journal* and *System*. IRIS currently holds over 850 documents bundled into approximately 280 instruments. The coverage of the database is wide, with over fifty instrument types represented, including language background questionnaires, cloze tests, grammaticality judgement tests, and elicitation tasks, and over forty research areas, including motivation, processing instruction, and task-based interaction.

As a research area CALL is currently under-represented with only two instruments in comparison with morphosyntax (grammar) for which over 100 instruments have been uploaded. In line with current interests in computer-mediated task-based language learning, however, a variety of tasks are held on the database which might be re-used and adapted in this area of research. These include tasks designed to:

- Investigate learners’ use of communication strategies -e.g. García Mayo’s (2005) decision-making task.
- Elicit specific morphosyntactic forms -e.g. Mifka Profozic’s (2012) picture description tasks for eliciting the French *passé composé* and *imparfait*.
- Examine the impact of task complexity on the extent to which focus on form or meaning -e.g. Révész’s (2011) argumentative tasks.

Moreover, if your area of interest in CALL or SLA is not represented and there is an instrument that you would like to examine or re-use, it is possible to get the IRIS team ([iris@iris-database.org](mailto:iris@iris-database.org)) to track down the materials for you by placing a request through the IRIS database.

## 6. Conclusion

Current CALL research which is dominated by broad media comparisons has resulted in “an accumulation of vaguely inter-connected research findings” (Porte, 2013, p. 12). In order to *construct a reliable evidence-base with ‘engineering power’* upon which to base future CALL design, more basic research —and replications thereof— is necessary. Instrumental replication is particularly important to permit researchers to build on the findings of previous research. In order to permit such comparisons, CALL researchers in the field are encouraged to contribute instruments from their peer-reviewed publications to the IRIS database. With nearly 5000 downloads to date, 15000 hits on the site, and references to the publications in which the instruments have been used, having materials on IRIS increases the visibility of individual researcher’s work. Integrating the option to request downloaders to leave their name and e-mail address, having materials on IRIS also permits researchers to track the impact of their research.

## Acknowledgements

IRIS is developed and curated by the Digital Library at the University of York, and directed by Emma Marsden (York, UK) and Alison Mackey (Georgetown, USA / Lancaster, UK). It is funded by the Economic and Social Research Council and the British Academy. The author would like to thank the IRIS team for supporting attendance at EUROCALL 2014 where an earlier version of this paper was presented, and in particular Emma Marsden for her suggestions on how to improve the paper.

## References

- Bannan-Ritland, B. (2003). The role of design in research: The integrative learning design framework. *Educational Researcher*, 32(1), 21-24.
- Barab, S. & Squire, K. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences*, 13(1), 1-14.
- Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe and do. *Language Teaching*, 36(2), 81-109.
- Burkhardt, H. & Schoenfeld, A. H. (2003). Improving educational research: Toward a more useful, more influential, and better-funded enterprise. *Educational Researcher*, 32(3), 3-14.
- Burston, J (2006). Working towards effective assessment of CALL. In Donaldson, P. R. & Haggstrom, M. A. (eds.). *Changing Language Education Through CALL*. London: Routledge, 249-270.
- Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, A. L., Sandercock, P., Spiegelhalter, D. & Tyrer, P. (2000). Framework for design and evaluation of complex interventions to improve health. *British Medical Journal*, 321, 694-696.
- Chapelle, C. (2010). The spread of computer-assisted language learning. *Language Teaching*, 43(1), 66-74.
- Chen, C.-M. & Li, Y. L. (2010). Personalised context-aware ubiquitous learning system for supporting effective English as a second language. *TESOL Quarterly*, 20(1), 27-46.
- Chun, D. (2012). Review article: Replication studies in CALL research. *CALICO Journal*, 29(4), 591-600.
- Colpaert, J. (2006). Pedagogy-driven design for online language teaching and learning. *CALICO Journal*, 23(3), 477-497.
- Colpaert, J. (2010). Elicitation of language learners' personal goals as design concepts. *Innovation in language learning and teaching*, 4(3) 259-274.
- Dalton, B., Proctor, C. P., Uccelli, P., Mo, E. & Snow, C. E. (2011). Designing for diversity: The role of reading strategies and interactive vocabulary in a digital reading environment for fifth-grade monolingual English and bilingual students. *Journal of Literacy Research*, 43, 68-100
- Design-Based Research Collective (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, 32(1), 5-8
- García Mayo, M. (2005). Interactional strategies for interlanguage communication: Do they provide evidence for attention to form? In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (Studies on Language Acquisition Series). Mouton de Gruyter.
- Garrett, N. (1991). Technology in the service of language learning: Trends and issues. *Modern Language Journal*, 75, 74-101
- Gorard, S., Roberts, K. & Taylor, C. (2004). What kind of creature is a design experiment? *British Educational Research Journal*, 30(4), 577-590.
- Gruber-Miller, J. & Benton, C. (2001). How do you say 'MOO' in Latin? Assessing student learning and motivation in beginning Latin. *CALICO Journal*, 18(2), 305-38.
- Gutiérrez-Colon, M.; Gibert, M.I.; Triana, I.; Gimeno, A.; Appel, C. & Hopkins, J. (2003). Improving learners' reading skills through instant short messages: a sample

study using WhatsApp. *WorldCALL 2013 – CALL: Sustainability and Computer-Assisted Language Learning Conference Proceedings*. University of Ulster, pp.80-84.

Hiebert, J., Gallimore, R. & Stigler, J. (2002). A knowledge base for the teaching profession: What would it look like and how can we get one? *Educational Researcher*, 31(3), 3-15.

Higher Education Research Funding Council (1999). *Guidance on submissions research assessment exercise*, Paragraph 1.12. London: Higher Education Funding Council for England and Wales 1999.

Hwang, W. Y. & Chen, H. S. L. (2013). Users' familiar situational context facilitate the practice of EFL in elementary schools with mobile devices. *Computer-Assisted Language Learning*, 26(2), 101-125

Kim, D.; Gilman, D. A. (2008). Effects of Text, Audio, and Graphic Aids in Multimedia Instruction for Vocabulary Learning. *Educational Technology & Society*. 11 (3): 114-126.

Kumaravadivelu, B. (1994). The post-method condition: (E)merging strategies for Second/Foreign language teaching. *TESOL Quarterly*, 28(1), 27-48.

Levy, M. & Stockwell, G. (2006). *CALL Dimensions: Options and Issues in Computer-Assisted Language Learning*. London: Lawrence Erlbaum.

Lu, M. (2008). Effectiveness of vocabulary learning via mobile phones. *Journal of Computer Assisted Learning*, 24(6), 515-525.

Luckin, R. & Clark, W. (2011). More than a game: The participatory Design of contextualised technology-rich learning experiences with the ecology of resources. *Journal of e-Learning and Knowledge Society*, 7(3), 33-50.

Lund, A. (2008). Wikis: A collective approach to language production. *ReCALL*, 20(1), 35-54.

Macaro, E., Handley, Z. L., & Walter, C. (2012). A systematic review of CALL in English as a second language: Focus on primary and secondary education. *Language Teaching*, 45(1), 1-43

Marsden, E. & King, J. (2013). The Instruments for Research into Second Languages (IRIS) digital repository. *The Language Teacher*, 37(2), 35-38.

Marsden, E. J., & Mackey, A. (2014). IRIS: a new resource for second language research. *Linguistic Approaches to Bilingualism*, 4(1), 125-130.

Mifka Profozic, N. (2012). Oral corrective feedback, individual differences and L2 acquisition of French past tenses. (Unpublished doctoral dissertation). University of Auckland.

Moher, D., Jones, A. & Lepage, L. (2001). Use of the CONSORT Statement and quality of reports of randomized trials. A comparative before-and-after evaluation. *The Journal of the American Medical Association*, 285, 1992-5.

Moher, D., Schulz, K. ., Altman, D. (2001). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *The Journal of the American Medical Association*, 285, 1987-91.

O'Hara, & Pritchard, (2008). Hypermedia authoring as a vehicle for vocabulary development in middle school English as a second language classrooms. *Clearing House: A Journal of Educational Strategies, Issues, and Ideas*, 82(2), 60-65.

Oswald, F. L. & Plonsky, L. (2010). Meta-analysis in second language research: Choices and challenges. *Annual Review of Applied Linguistics*, 30, 85-110.

- Pardo-Ballester, C. & Rodriguez, J. C. (2009). Using design-based research to guide the development of online instructional materials. In Chapelle, C. A., Jun, H. G., & Katz, I. (Eds.). *Developing and evaluating language learning materials* (pp. 86-102). Ames, IA: Iowa State University.
- Pardo-Ballester, C. & Rodriguez, J. C. (2010). Developing Spanish online readings using design-based research. *CALICO Journal*, 27(3), 540-553.
- Pederson, K (1987). Research on CALL. In Smith, W. F. (ed.). *Modern media in foreign language education: Theory and implementation*. Lincolnwood, Illinois: National Textbook Company, pp. 99-131.
- Polio, C. (2012). Replication in published applied linguistics research: A historical perspective. In Porte, G. (ed.). *Replication research in applied linguistics*. Cambridge: Cambridge University Press, pp. 47-91.
- Polio, C. & Gass, S. (1997). Replication and reporting. *Studies in Second Language Acquisition*, 19, 499-508.
- Porte, G. (2013). Who needs replication? *CALICO Journal*, 30(1), 10-15
- Porte, G. & Richards, K. (2012). Focus article: Replication in second language writing research. *Journal of Second Language Writing*, 21(2012), 284-193.
- Proctor, C. P., Dalton, B. & Grisham, D. L. (2007). Scaffolding English language learners and struggling readers in a universal literacy environment with embedded strategy instruction and vocabulary support. *Journal of Literacy Research*, 39(1), 71-93.
- Révész, A. (2011). Task complexity, focus on L2 constructions, and individual differences: A classroom-based study. *The Modern Language Journal*, 95(4).
- Sasaki, A. & Takeuchi, O. (2010). EFL students' vocabulary learning in NS-NNS e-mail interactions: Do they learn new words by imitation? *ReCALL*, 22(1): 70-82.
- Slavin, R. E. (1995). Best evidence synthesis: An intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology*, 48(1), 9-18.
- Stockwell, G. (2012). Diversity in research and practice. In Stockwell, G. (ed.). *Computer-Assisted Language Learning: Diversity in Research and Practice*. Cambridge: Cambridge University Press, pp. 147-163.
- Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus-groups. *International Journal for Quality in Health Care*, 19(6), 349-357.
- van Lier, L. (2004). *The Ecology and Semiotics of Language Learning. A Sociocultural Perspective*. Boston: Kluwer Academic
- Yildiz, R. & Atkins, M. (1993). Evaluating multimedia applications. *Computers in Education*, 21(1/2), 133-139.
- Yutdhana, S. (2008). Design-based research in CALL. In Egbert, J. L. & Petrie, G. M. (eds.). *CALL research perspectives*. Mahwah, NJ: Lawrence Erlbaum, pp. 169-178.

## Notes

[1] MOO stands for Multi-user Object Oriented domains. A MOO is an online 2D or 3D virtual world.



## Article

### ***Podcasts for Learning English Pronunciation in Igboland: Students' Experiences and Expectations***

**E.E. Mbah\*, B.M. Mbah\*\*, M.I. Iloene\*\*\***

**Department of Linguistics, Igbo and other Nigerian languages, University of Nigeria, Nsukka, Nigeria**

**G. Iloene\*\*\*\***

**Department of Linguistics, Ebonyi State University, Abakaliki, Nigeria**

---

\*ezyambah @ yahoo.co.uk | \*\*boniface.mbah @ unn.edu.ng | \*\*\*modesta.iloene @ unn.edu.ng | \*\*\*\*georgeiloene @ yahoo.com

#### **Abstract**

This paper studies students' experiences and expectations on the use of podcasts in learning English pronunciation in Igboland. The study is based on a survey where two universities were studied. A proportional sampling technique with the aid of a structured questionnaire was used to elicit information. The data gathered were analysed using mean, standard deviation, t-test, and ANOVA with the aid of Statistical tool for the Social sciences. The study concluded that the students agreed to the fact that podcasts improved their English pronunciation. The hypotheses tested showed that there was no significant difference in the use of podcasts with regard to the students' internet usage habits, language proficiency level, or gender. Thus, it was concluded that this technology was appropriate for second language learning.

**Keywords:** Podcasts, English pronunciation, Igboland.

#### **1. Introduction**

English is learned as a second language in Igboland (and all over Nigeria). Learning English in Nigeria is confronted with some challenges. One of the challenges according to Oluike (1978) is that English is not taught to solve the language problems relating to the linguistic peculiarities of Nigerian learners including the fact that rather than writing textbooks for commercial purposes as is customary in the country, they should be written for different language groups to solve their peculiar linguistic problems. Language teaching involves teaching pronunciation, vocabulary, grammar and the culture of the target language. The place of pronunciation in L2 teaching is often relegated when compared to grammar, vocabulary and culture (Lord, 2008). Part of the reason for the relegation is that many teachers assume that with more L2 input, students will learn pronunciation, or it will be acquired at a later stage.

Podcasts are among the new techniques and technologies which meet the learners' needs of having additional pronunciation input outside the classroom. The net-generation students are often very busy and involve themselves in multitasking (Tapscott, 2009) and many of them have devices for playing audio files (Rainie & Madden, 2005; Schmidt, 2008). These reasons combine to make it necessary for podcasts to be one of the tools used in delivering L2 materials to the students. Hence,

Craig, et al. (2007) and Windham (2007) agree that many L2 learners find the use of podcasts motivating since many of them study through distant learning programs, and may not have enough time to attend language laboratory and classrooms regularly. Podcasts are important for teaching and learning phonetics. Knight (2010) avers that podcasts help in alleviating the difficulties students encounter in phonetics since they provide the students with an alternative audio-based exercise material against the paper-based ones. The report of Thorne and Payne (2005) of Duke University's iPod projects reveals that podcast-based projects in language classrooms are necessary in developing oral skills. Lord (2008) engaged his students (undergraduate Spanish Phonetics class) in a collaborative podcasting project. The students were divided into small groups and were made to create and maintain their own podcasts channel where they uploaded recordings for their members to comment on. Pre- and post- semester attitudes and pronunciation abilities were tested. The results showed that there was an improvement on the students' pronunciation even though the factor(s) that influenced the improvement cannot be pinpointed. On the other hand, Ducate and Lomicka (2009) tested the effectiveness of podcasting in honing American English students' pronunciation in German and French and records insignificant improvement although the students appreciated the tool. Knight (2010) investigated students' perception of podcasts in phonetics. The results showed that the students perceived podcasts to be useful in their learning of phonetics. In the same vein, Li (2012) examined the students' perception of podcasts and reports that the year 6 secondary school students perceived podcasts to be a useful tool which has enhanced their language skills. Chan, Chen, and Döpel (2011) studied two podcast projects organised at a university in Singapore, aimed to aid classroom instruction for Chinese and Koreans as a foreign language. They used a semi-structured interview to determine their perceptions of the podcasts' quality and usefulness. They observed that respondents who used podcasts on the move or outside their abodes had significant positive attitudes towards podcasts and were also found to be interested in podcast-based learning after being exposed to the podcast course. Hasan and Hoon (2013) reviewed twenty journal articles to establish the effects of podcasts on ESL students' language skills and attitude levels. They found out that podcasts greatly facilitate L2 pronunciation among other language skills.

### *1.2. Purpose of the study*

The general objective of this study is to identify podcasts for learning English pronunciation by students in Igboland. Its specific objectives are contained in the subsequent research questions and hypotheses.

### *1.3. Research questions*

The following research questions will be answered by the study:

1. How does the students' background regarding internet use or computer-assisted language learning (CALL) affect the use of ESL podcasts?
2. To what extent do the types of gadget students use for podcasts affect their interest in listening to podcasts?
3. How did students come to learn about podcasts?
4. Why do students listen to podcasts?
5. In what manner does listening to podcasts influence students' performance in English phonetics-related courses and their spoken English?
6. What are the students' experiences in using podcasts to learn English pronunciation?
7. What are the students' expectations in using podcasts to learn English pronunciation?

#### 1.4. Research hypotheses

The following seven null hypotheses were formulated and tested at 0.05 level of significance.

1. There is no significant difference in the mean ratings of the responses of male and female students regarding their internet or computer-assisted language learning usage for language learning.
2. There is no significant difference in the mean ratings of the responses of male and female students on the types of gadget that affect one's active use of podcasts.
3. There is no significant difference in the mean ratings of the responses of 100, 200, 300 and 400 level students on the influences of students' background information on the knowledge of podcasts.
4. There is no significant difference in the mean ratings of the responses of male and female students on the students' reasons for listening to podcasts.
5. There is no significant difference in the mean ratings of the responses of 100, 200, 300 and 400 level students on the influences of students' listening to podcasts on performance in English phonetics.
6. There is no significant difference in the mean ratings of the responses of 100, 200, 300 and 400 level students on students' experiences in using podcasts to learning English pronunciation.
7. There is no significant difference in the mean ratings of the responses of male and female students' expectations in using podcasts to learn English pronunciation.

## 2. Methodology

This study adopted survey-based research. Five universities in the five major states that constitute Igboland were used for the study. One university from each state was taken. In a situation where there are two or more universities in a state, preference was given to a federal university. However, where there is no federal university of up to ten years of age, an older state university was chosen. The five universities selected were Abia State University, Umuahia; Ebonyi State University, Abakaliki; Imo State University, Owerri; Nnamdi Azikiwe University, Awka; and the University of Nigeria, Nsukka. A preliminary study carried out showed that only two universities made use of podcasts. The universities that made use of podcasts were the University of Nigeria, Nsukka, and Nnamdi Azikiwe University, Awka.

The population of this study was comprised of 4000 respondents, made up of students from each academic level. The proportionate sampling technique was used for the study. In using this technique, the 10% sample that was drawn from the population of 4000 students amounted to 400.

### 2.1. Data collection instrument

The instrument used for data collection was a structured questionnaire. The questionnaire was made up of two parts. Part one comprising nine items to collect information on personal data, while part two, comprising 34 items, was structured into six sections enquiring about students' experience and expectations in using podcasts for learning English pronunciation. The response scale was as follows:

Strongly Agree (SA)	5 points
Agree (A)	4 points
Undecided (UD)	3 points
Disagree (D)	2 points

Strongly Disagree (SD)	1point
Total	15
Mean for response scale	15/5
	= 3.00

### 2.2. Data collection method

The questionnaire was administered to the respondents and collected after one week. A total of 376 out of 400 copies of the questionnaire items were returned. It was these numbers that were analyzed to generate data used for answering the research questions and testing of the null hypotheses.

### 2.3. Data analysis method

The data collected from the respondents was analyzed using mean, standard deviation t-test and ANOVA statistics. SPSS was used also to ensure accuracy. The mean and standard deviation were used to answer the research questions. The mean for the response scale was 3.00. The lower limit of the mean was 2.50 while the upper limit was 3.50 with an interval scale of 0.05 from the mean. Any item with a mean rating of 3.50 and above was regarded as "agreed" while any item with a mean rating less than 3.50 was regarded as "disagreed". The standard deviation was used to determine the closeness or otherwise of the opinions of the respondents from the group mean. The t-test and ANOVA statistics were used to test the five null hypotheses at a probability of 0.05 level of significance for the calculation of t-test and ANOVA statistics. In any hypotheses whose significance levels were less than or equal to the stated 0.05 level of significance, the null hypothesis was rejected, but if the levels were greater than an 0.05 level of significance, the null hypotheses were accepted.

## 3. Results and discussion

The results of the research questions and their corresponding hypotheses are presented and subsequently discussed in this section.

### Research Question One

How does the students' background regarding internet use or computer-assisted language learning affect the use of ESL podcasts?

#### Hypothesis One

There is no significant difference in the mean ratings of the responses of male and female students regarding their internet or computer-assisted language learning usage for language learning.

The data for answering research question one and testing hypothesis one are presented in table 1 below.

SN	Students' use of the internet that affects the use of ESL podcasts	XM	XF	XG	p-values	Remarks	
						RQ	Ho
1	Searching internet information and resources	4.15	4.14	<b>4.15</b>	0.978	<b>Agree</b>	<b>NS</b>
2	Accessing SLN courses and class assignments	4.24	4.47	<b>4.36</b>	0.005	<b>Agree</b>	<b>S*</b>

3	Making PowerPoint slides	4.10	4.26	<b>4.18</b>	0.055	<b>Agree</b>	<b>NS</b>
4	Downloading online podcasts	3.97	4.08	<b>4.02</b>	0.178	<b>Agree</b>	<b>NS</b>
5	Listening to online podcasts	4.03	4.30	<b>4.17</b>	0.001	<b>Agree</b>	<b>S*</b>
6	Posting comments to online group and social networking	4.10	4.33	<b>4.22</b>	0.009	<b>Agree</b>	<b>S*</b>
7	Creating/working on webpage, journal and weblog	4.10	4.09	<b>4.10</b>	0.928	<b>Agree</b>	<b>NS</b>
8	Checking emails with different browsers	4.12	4.06	<b>4.09</b>	0.502	<b>Agree</b>	<b>NS</b>
9	Sharing ideas using e-learning forum platform	3.99	4.12	<b>4.05</b>	0.124	<b>Agree</b>	<b>NS</b>

Table 1. Mean ratings and t-test statistics regarding the students' internet or computer-assisted language learning usage (N = 376)(1) .

The results presented in table 1 show that the grand mean values of the 9 items in the table ranged from 4.02 to 4.36 which are all greater than the cut-off point value of 3.00 on the 5-point rating scale. This indicates that the respondents agreed that all 9 items in the table are ways through which the internet or computer-assisted language learning affect their use of ESL podcasts.

The results of t-test statistics in the table reveal that, the p-values of items 2, 5 and 6 were 0.005, 0.001 and 0.009 respectively which are less than a 0.05 level of significance. This implies that there are significant differences in the mean ratings of male and female students on the 3 items. On the other hand, the p-values of the remaining 6 items range from 0.55 to 0.978 which are greater than a 0.05 level of significance. This indicates that there are no significant differences in the mean ratings of the responses of male and female students on the six items.

Hence, the students' internet or CALL usage facilitates their motivation for using podcasts to improve their English pronunciation. One who does not know his/her way through the internet may find it difficult to search for suitable podcasts. This is not the case with these students since this buttresses Tapscott's (1999) view that this generation of students belong to the net-generation who involve themselves in multitasking and avail themselves of internet opportunities.

### Research Question Two

What are the types of gadget that affect one's active use of podcasts?

### Hypothesis Two

There is no significant difference in the mean ratings of the responses of male and female students on the types of gadget that affect one's active use of podcasts.

The data for answering research question two and testing hypothesis two are presented in table 2 below.

SN	Types of gadget that affect one's active use of podcasts include:	XM	XF	XG	p-values	Remarks	
						RQ	Ho
1	The use of ipod/Mp3 player	3.99	4.03	<b>4.06</b>	0.132	<b>Agree</b>	<b>NS</b>
2	The use of desktop computer	3.98	3.94	<b>3.96</b>	0.566	<b>Agree</b>	<b>NS</b>

3	The use of laptop/notebook computer	4.01	4.02	<b>4.01</b>	0.904	<b>Agree</b>	<b>NS</b>
4	The use of cell phone	4.04	4.06	<b>4.05</b>	0.810	<b>Agree</b>	<b>NS</b>
5	The use of Blackberry, Mp4 and other personal digital assistant	4.16	4.38	<b>4.27</b>	0.009	<b>Agree</b>	<b>S*</b>
6	The use of digital camera	4.27	4.33	<b>4.31</b>	0.073	<b>Agree</b>	<b>NS</b>
7	The use of video camera	4.34	4.55	<b>4.45</b>	0.010	<b>Agree</b>	<b>S*</b>
8	The use of webcam	4.39	4.26	<b>4.32</b>	0.130	<b>Agree</b>	<b>NS</b>

Table 2. Mean ratings and t-test statistics of the types of gadget that affect one's active use of podcasts (N = 376).

For abbreviations see Note 1 below.

These results reveal that the respondents agreed that ipod/Mp3 players, desktop computers, laptop/notebook computers, cell phones, Blackberry, Mp4 and the likes, digital cameras, and webcams are the types of gadgets that affect active use of podcasts. Thus, portables devices for playing audio files are good for downloading and listening to podcasts (Rainie & Madden, 2005; Schmidt, 2008). The extra input to support students in their learning can be accessed through the use of new mobile technologies (Ashby, Feguera-Clark, Seo, & Yanagisawa, 2005; Stenson. Downing, Smith & Smith 1992; Eskenazi 1999; Hardison 2004). thus helping students listen to their downloaded podcasts anywhere, anytime.

### Research Question Three

How does the students' prior knowledge on podcasts influence their motivation?

### Hypothesis Three

There is no significant difference in the mean ratings of the responses of 100, 200, 300 and 400 level students on the influences of students' background information on the knowledge of podcasts.

The data for answering research question three and testing hypothesis three are presented in table 3 below.

SN	Influences of students' background information on the knowledge of podcasts include:	Total Sum of Square	Mean Square	P-values	XG	Remarks	
						RQ	Ho
1	knowledge of podcasts was through my teacher	231.104	2.356	0.009	<b>4.16</b>	<b>Agree</b>	<b>S*</b>
2	knowledge of podcasts was through a friend	198.040	0.524	0.136	<b>3.95</b>	<b>Agree</b>	<b>NS</b>
3	knowledge of podcasts was through my interest in acquiring the native speaker's pronunciation	239.436	1.243	0.119	<b>4.19</b>	<b>Agree</b>	<b>NS</b>
4	learning about podcasts is necessary	222.202	1.534	0.100	<b>4.37</b>	<b>Agree</b>	<b>NS</b>

Table 3. Mean ratings and analysis of variance (ANOVA) on the influences of students' background information on the knowledge of podcasts (N = 376).

The results presented in table 3 show that the grand mean values of the 4 items in the table range from 3.95 to 4.37 which are all greater than the cut-off point value of 3.00

on the 5-point rating scale. This indicates that the respondents agreed that all 4 items in the table have influenced the students' background knowledge about podcasts.

The results of Analysis of variance (ANOVA) in the table reveal that, the p-values of items 1 was 0.009 which is less than a 0.05 level of significance. This shows that there are significant differences in the mean ratings of students at all levels of education on item 1. On the other hand, the p-values of the remaining 3 items ranged from 0.100 to 0.136 which are greater than a 0.05 level of significance. This indicates that there are no significant differences in the mean ratings of the responses of students at all levels of education on the 3 items. This finding shows that irrespective of the source of information, the students are willing to improve their pronunciation. Similar studies have shown that students desire to acquire native-like pronunciation irrespective of where the information is taken from (Derwing and Munro, 2003; Kang, 2010; Scales, Wennerstrom, Richard, & Wu, 2006; Timmis, 2002). This may be one of the reasons why they were willing to access podcasts and listen to them without being monitored.

#### Research Question Four

Why do students listen to podcasts?

#### Hypothesis Four

There is no significant difference in the mean ratings of the responses of male and female students on the students' reasons for listening to podcasts.

The data for answering research question four and testing hypothesis four are presented in table 4 below.

SN	Students' reasons for listening to podcasts are:	XM	XF	XG	p-values	Remarks	
						RQ	Ho
1	Vocabulary	4.37	4.28	4.33	0.280	Agree	NS
2	Pronunciation	4.25	4.20	4.23	0.582	Agree	NS
3	Composition	4.28	4.15	4.22	0.059	Agree	NS
4	Grammar	3.96	4.04	4.01	0.375	Agree	NS
5	Logical reasoning	4.42	4.47	4.45	0.483	Agree	NS
6	Socialisation	4.04	3.88	3.96	0.068	Agree	NS
7	Lectures	4.09	3.94	4.00	0.207	Agree	NS
8	Entertainment	4.29	4.18	4.23	0.064	Agree	NS

Table 4. Mean ratings and t-test statistics regarding the students' reasons for listening to podcasts (N = 376).

The results presented in table 4 show that the grand mean values of the 8 items in the table range from 3.96 to 4.45 which are all greater than the cut-off point value of 3.00 on the 5-point rating scale. This indicates that the respondents agreed that all the 8 items in the table are reasons for listening to podcasts.

The results of Analysis of variance (ANOVA) in the table reveal that, the p-values of the 8 items range from 0.059 to 0.483 which are greater than a 0.05 level of significance. This indicates that there are no significant differences in the mean ratings of the responses of students' reasons for listening to podcasts at all levels of education on the

8 items. The results of research question four show that although many students were introduced to podcasts first by their phonetics teachers, as revealed in the results of research question three, the primary essence of a phonetics teacher introducing his students to podcasts may be to improve their pronunciation, but as the students started listening to podcasts, they discovered that in addition to good pronunciation, podcasts are also useful in areas of vocabulary development, composition, grammar, logical reasoning, socialisation, lectures and entertainment.

### Research Question Five

How does listening to podcasts influence students' performance in English phonetics-related courses and their spoken English?

### Hypothesis Five

There is no significant difference in the mean ratings of the responses of 100, 200, 300 and 400 level students on the influences of students' listening to podcasts on performance in English phonetics.

The data for answering research question five and testing hypothesis five are presented in table 5 below.

SN	Influence of listening to podcasts on performance in phonetics include:	Total Sum of Square	Mean Square	P-values	XG	Remarks	
						RQ	Ho
1	Podcasts have positively affected scores in English Phonetics courses	193.660	0.915	0.310	<b>4.15</b>	<b>Agree</b>	<b>NS</b>
2	Podcasts have improved oral English performance	196.359	0.545	0.035	<b>4.35</b>	<b>Agree</b>	<b>S*</b>
3	Podcasts have improved English pronunciation more than vocabulary	206.231	1.111	0.108	<b>4.05</b>	<b>Agree</b>	<b>NS</b>
4	Podcasts have improved English pronunciation more than grammar	186.245	0.464	0.425	<b>4.24</b>	<b>Agree</b>	<b>NS</b>
5	Podcasts have improved pronunciation more than logical reasoning	226.926	0.823	0.254	<b>3.89</b>	<b>Agree</b>	<b>NS</b>

Table 5. Mean ratings and Analysis of variance (ANOVA) on the influences of students' listening to podcasts on performance in English phonetics (N = 376).

The results presented in table 5 show that the grand mean values of the 5 items in the table range from 3.89 to 4.24 which are all greater than the cut-off point value of 3.00 on 5-point rating scale. This indicates that the respondents agreed that all 5 items in the table influence them toward listening to podcasts on performance in English phonetics.

The ANOVA analysis reveals that the p-value of item 2 was 0.035 which is less than a 0.05 level of significance. This indicates that there are significant differences in the mean ratings of students at all levels of education on item 2. On the other hand, the p-values of the remaining 4 items range from 0.108 to 0.425 which are greater than a 0.05 level of significance. This range indicates that there are no significant differences in the mean ratings of the responses of students at all levels of education on the 4 items.



The results above show that the teachers' primary aim of exposing them to podcasts was achieved. Hence they acknowledged the fact that they had gained more knowledge in pronunciation than in any other aspect of language acquisition.

### Research Question Six

What are the students' experiences in using podcasts to learn English pronunciation?

### Hypothesis Six

There is no significant difference in the mean ratings of the responses of 100, 200, 300 and 400 level students on students' experiences in using podcasts to learning English pronunciation.

The data for answering research question six and testing hypothesis six are presented in table 6 below.

SN	Students' experiences in using podcasts to learning English pronunciation include:	Total Sum of Square	Mean Square	p-values	XG	Remarks	
						RQ	Ho
1	It is convenient to listen to podcasts at any place any time	223.777	1.058	0.150	4.16	Agree	NS
2	Ability to download and save podcasts to Computer/mobile devices conveniently / easily	241.277	1.674	0.060	4.09	Agree	NS
3	Listening to podcasts on computer instead of iPod or mp3 players	245.989	0.633	0.327	4.00	Agree	NS
4	Listening to podcasts is interesting	214.614	0.370	0.587	4.13	Agree	NS
5	Podcasts are not good enough because the presenters cannot be asked questions	302.870	1.179	0.083	4.20	Agree	NS

Table 6. Mean ratings and analysis of variance (ANOVA) on students' experiences in using podcasts to learning English pronunciation (N = 376).

The results presented in table 6 show that the grand mean values of the 5 items in the table ranged from 4.00 to 4.20 which are all greater than the cut-off point value of 3.00 on the 5-point rating scale. This indicates that the respondents agreed that all 5 items in the table comprise the students' experiences in using podcasts while learning English pronunciation. The results of the analysis of variance (ANOVA) reveals that the p-values of the 5 items range from 0.060 to 0.587 which are greater than a 0.05 level of significance. This indicates that there are no significant differences in the mean ratings of the responses of students' experiences in using podcasts to learn English pronunciation at all levels of education on the 5 items.

The general view of the students on their experiences reveals that they enjoy listening to podcasts despite the fact that most of the podcasts are not interactive where they can ask the presenter questions. Craig, et al. (2007) and Windham (2007) also discovered that many L2 learners find the use of podcasts motivating. This is also in line with Knight (2010) and Thorne and Payne's (2005) findings. Newnham and Miller (2007) discovered that their respondents had a positive attitude toward using podcasts to learn.

### Research Question Seven

What are the students' expectations in using podcasts to learn English pronunciation?

### Hypothesis Seven

There is no significant difference in the mean ratings of the responses of male and female students' expectations in using podcasts to learn English pronunciation.

The data for answering research question seven and testing hypothesis seven are presented in table 7 below.

SN	Students' expectations in using podcasts to learn English pronunciation include:	XM	XF	XG	p-values	Remarks	
						RQ	Ho
1	The presenter's voice be clear	4.17	4.05	<b>4.16</b>	0.110	<b>Agree</b>	<b>NS</b>
2	Podcasts be interactive	4.15	4.22	<b>4.23</b>	0.204	<b>Agree</b>	<b>NS</b>
3	Free/cheap internet access be provided by the University administration	4.23	4.27	<b>4.35</b>	0.074	<b>Agree</b>	<b>NS</b>
4	Teachers of English phonetics and other related courses be abreast of new technologies in learning pronunciation	4.09	4.05	<b>4.07</b>	0.652	<b>Agree</b>	<b>NS</b>

Table 7. Mean ratings and t-test statistics on students' expectations in using podcasts to learn English pronunciation (N = 376).

The results presented in table 7 show that the grand mean values of the 4 items in the table range from 4.07 to 4.35 which are all greater than the cut-off point value of 3.00 on the 5-point rating scale. This indicates that the respondents agreed that all the 4 items in the table comprise the students' expectations in using podcasts to learn English pronunciation. The results of Analysis of variance (ANOVA) reveal that the p-values of the 4 items range from 0.074 to 0.652 which are greater than a 0.05 level of significance. This indicates that there are no significant differences in the mean ratings of the responses of students' expectations in using podcasts to learn English pronunciation at all levels of education on the 4 items.

Access to the internet is the first step in the use of podcasts. With the high cost of internet facilities in Nigeria, many students who would wish to download podcasts may not be able to do so. Hence they desire that their respective university administration to provide them with cheap internet facilities to enable them to make effective use of podcasts. Due to financial constraints a very limited number of students would be able to download podcasts on a regular basis. This may be one of the reasons why they expect their teachers to be abreast with new technologies for language learning since their teachers are presumably financially better off than they are.

### 4. Conclusion

This study sought to investigate students' experience and expectations regarding the use of podcasts to learn English pronunciation in the Igbo speech community. Most students testified to the fact that although their first knowledge of podcasts was independent of their teacher, they see podcasts as an effective tool that has reasonably improved their oral performance in English phonetics-related courses through the use of mobile gadgets. For effective use of podcasts in learning English pronunciation in Igboland, the following recommendations are made: 1. the presenter's voice should be as clear as possible; 2. as far as possible podcasts should be interactive; 3. free/cheap internet access should be provided by the University administration, and 4. teachers of English phonetics and other related courses should be abreast of new technologies to aid in learning pronunciation. Podcasts are therefore a pedagogic instrument that

learners of English language in the Igbo speech community embrace in learning English as a second language in all levels of undergraduate education irrespective of gender.

## Notes

[1]

<b>XM</b>	Mean of Male Students
<b>XF</b>	Mean of Female Students
<b>XG</b>	Grand mean
<b>RQ</b>	Research Question
<b>Ho</b>	Hypothesis
<b>NS</b>	Not Significant
Level of significance	$p \leq 0.05$
<b>S*</b>	Significant

## References

- Ashby, M., Figueroa-Clark, M., Seo, E., & Yanagisawa, K. (2005). Innovations in practical phonetics teaching and learning. *Proceedings of the Phonetics Teaching and Learning Conference*, UCL.
- Chan, W.M., Chen, I.R., & Döpel, M. (2011). Podcasting in foreign language learning: Insights for podcast design from a developmental research project. In M. Levy, F. Blin, C. Bradin Siskin, & O. Takeuchi (Eds.), *WorldCALL: Global perspectives on computer-assisted language learning*. New York & London: Routledge, pp. 19-37.
- Craig, D., Paraiso, J., & Patten, K.B. (2007). *eLiteracy and literacy: Using ipods in the ESL classroom*. Paper presented at the Society for Information Technology and Teacher Education International Conference, Chesapeake, VA.
- Derwing, T.M., & Rossiter, M.J. (2002). ESL learners' perceptions of their pronunciation needs and strategies. *System*, 30, 155-166.
- Ducate, L. & Lomicka, L. (2009). Podcasting: An effective tool for honing language students' pronunciation? *Language Learning & Technology*, 13, 66-86.
- Eskenazi, M. (1999). Using automatic speech processing for foreign language pronunciation tutoring: Some issues and a prototype. *Language Learning and Technology*, 2(2), 62-76.
- Hardison, D. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology*, 8, 34-52.
- Hasan, M.M. & Hoon, T.B. (2013). Podcast applications in language learning: A review of recent studies. *English Language Teaching*, 6. doi:10.5539/elt.v6n2p128 URL: <http://ccsenet.org/journal/index.php/elt/article/view/23820>.

- Kang, O. (2010). ESL learners' attitudes toward pronunciation instruction. In J.M. Levis & K. LeVelle (Eds.), *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference*. Ames, IA: Iowa State University, pp. 105–118.
- Knight, R-A. (2010). Sounds for study: Speech and language therapy students' use and perception of exercise podcasts for phonetics. *International Journal of Teaching and Learning in Higher Education*, 22, 269–276.
- Lord, G. (2008). Podcasting communities and second language pronunciation. *Foreign Language Annals*, 41, 364–379.
- Scales, J., Wennerstrom, A., Richard, D., & Wu, S.H. (2006). Language learners' perceptions of accent. *TESOL Quarterly*, 40, 715–738.
- Schmidt, J. (2008). Podcasting as a learning tool: German language and culture every day. *Die Unterrichtspraxis/Teaching German*, 41, 186–194.
- Stenson, N., Downing, B., Smith, J., & Smith, K. (1992). The effectiveness of computer assisted pronunciation training. *CALICO Journal*, 9, 5–18.
- Tapscott, D. (2009). *Grown up digital: How the net generation is changing your world*. New York: McGraw-Hill.
- Timmis, I. (2002). Native-speaker norms and international English: A classroom view. *ELT Journal*, 56, 240–249.
- Thorne, S. & Payne, S. (2005). Evolutionary trajectories, internet-mediated expression, and language education. The Pennsylvania State University [http://language.la.psu.edu/~thorne/thorne\\_payne\\_calico2005.pdf](http://language.la.psu.edu/~thorne/thorne_payne_calico2005.pdf)
- Windham, C. (2007). Confessions of a podcast junkie. *EDUCAUSE Review*, 42, 50–65.

## Editorial Board:

Dr. Christine Appel (Universitat Oberta de Catalunya, ES) | Dr. Paul Bangs (London City University, UK) | Dr. Françoise Blin (Dublin City University, IE) | Dr. Angela Chambers (University of Limerick, IE) | Dr. Thierry Chanier ( Université Clermont 2, FR ) | Dr. Jozef Colpaert (University of Antwerp, BE) | Dr. Robert Fischer (Texas State University, USA) | Mr. Anthony Fitzpatrick (Consultant, UK) | Dr. John Gillespie (University of Ulster at Coleraine, UK) | Dr. Trude Heift (Simon Fraser University, Canada) | Dr. Phil Hubbard (Stanford University, USA) | Dr. Kristi Jauregui (University of Utrecht, NL) | Dr. Mike Levy (University of Queensland, Australia) | Dr. Peter Liddell (University of Victoria, Canada) | Dr. Robert O'Dowd (Universidad de León, ES) | Dr. Pascual Pérez-Paredes (Universidad de Murcia, ES) | Dr. Joan Tomàs Pujolà (Universitat de Barcelona, ES) | Dr. German Ruiperez (Universidad Nacional de Educación a Distancia, ES) | Dr. Bernd Rüschoff (University of Duisburg-Essen, DE) | Dr. Mathias Schulze (University of Waterloo, Canada) | Dr. Bryan Smith (Arizona State University, USA) | Dr. Peppi Taalas (University of Jyväskylä, FI) | Mrs. June Thompson (University of Hull, UK) | Dr. Steven Thorne (Pennsylvania State University, USA) | Dr. George S. Ypsilandis (Aristotle University, GR) |

---

*The EUROCALL Review* is published online biannually by the European Association for Computer-Assisted Language Learning (EUROCALL) and is hosted online at <http://www.eurocall-languages.org/review/index.html> and at the Universitat Politècnica de València (Spain). Regular sections include:

- *Reports on EUROCALL Special Interest Groups*: up-to-date information on SIG activities.
- *Projects*: reports on on-going CALL or CALL-related R&D projects in which EUROCALL members participate.
- *Recommended websites*: reports and reviews of examples of good practice in language learning website development.
- *Articles*: research, reflective-practice or research & development articles relating to CALL and TELL.
- *Publications by EUROCALL members*: reviews of new books, CALL software, etc. published or edited by members.

*The EUROCALL Review* is indexed in ERIC (Education Resources Information Center - an internet-based digital library of education research and information sponsored by the Institute of Education Sciences of the U.S. Department of Education.)

Articles submitted to *The EUROCALL Review* are peer-reviewed by members of the Editorial Board.

If you can provide information on any topic of interest to the CALL community, please send this directly to Ana Gimeno, email [agimeno@upvnet.upv.es](mailto:agimeno@upvnet.upv.es)



*The EUROCALL Review*, edited by Ana Gimeno, is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License](https://creativecommons.org/licenses/by-nc-nd/3.0/).