e-TEALS

An **e**-journal of
Teacher **E**ducation and
Applied **L**anguage **S**tudies

Analysing English; Le
English; Listening to E
Reading English; Res
English; Speaking En
Studying English; Tea
English; Translating E
Using English; Writing
English¸ Analysing En
Learning English; List
English; Reading Eng
Researching English;
Speaking English; Stu
English; Teaching Eng
Translating English; U
English; Writing Engli
Analysing English; Le
English; Listening to E
Reading English; Res
English; Speaking En
Studying English List
English Reading Engl
Researching English;
Speaking English; Stu
English; Teaching Eng
Translating English; U
English; Writing Engli
Analysing English; Le
English; Listening to E
Reading English; Res
English; Speaking En
Studying English; Tea
English; Translating E
Using English; Writing
English¸ Analysing En
Learning English; List

e-TEALS

An **e**-journal of
Teacher **E**ducation and **A**pplied
Language **S**tudies

No. 5 | 2014

e-TEALS

An e-journal of Teacher Education and Applied Language Studies

CETAPS

Scientific committee

Manuela Guilherme (Marie Curie Research Fellow at the Centre for Social Studies, University of Coimbra)

Judith Hanks (The Language Centre, University of Leeds)

María Luisa Pérez Cañado (Departamento de Filología Inglesa, Universidad de Jaén)

Víctor Pavón Vázquez (Dpto. Filologías Inglesa y Alemana, Facultad de Filosofía y Letras, Universidad de Córdoba)

Siân Preece (Department of Culture, Communication and Media, Institute of Education, University of London)

Erwin Gierlinger (Institut Ausbildung Allgemeine Pflichtschulen (APS), Department of Modern Languages, University of Education Upper Austria)

Publication Ethics and Publication Malpractice Statement

It is necessary to agree upon standards of expected ethical behaviour for all parties involved in the act of publishing at *E-TEALS: An e-journal of Teacher Education and Applied Language Studies*: the author, the journal editor, the peer reviewer and the publisher.
Our ethics statements are based on COPE's Best Practice Guidelines for Journal Editors.

Publication decisions

The editor of the E-TEALS is responsible for deciding which of the articles submitted to the journal should be published.

The editor may be guided by the policies of the journal's editorial board and constrained by such legal requirements as shall then be in force regarding libel, copyright infringement and plagiarism. The editor may confer with other editors or reviewers in making this decision.

Fair play

An editor at any time evaluate manuscripts for their intellectual content without regard to race, gender, sexual orientation, religious belief, ethnic origin, citizenship, or political philosophy of the authors.

Confidentiality

The editor and any editorial staff must not disclose any information about a submitted manuscript to anyone other than the corresponding author, reviewers, potential reviewers, other editorial advisers, and the publisher, as appropriate.

CETAPS

Disclosure and conflicts of interest

Unpublished materials disclosed in a submitted manuscript must not be used in an editor's own research without the express written consent of the author.

DUTIES OF REVIEWERS

Contribution to Editorial Decisions

Peer review assists the editor in making editorial decisions and through the editorial communications with the author may also assist the author in improving the paper.

Promptness

Any selected referee who feels unqualified to review the research reported in a manuscript or knows that its prompt review will be impossible should notify the editor and excuse himself from the review process.

Confidentiality

Any manuscripts received for review must be treated as confidential documents. They must not be shown to or discussed with others except as authorized by the editor.

Standards of Objectivity

Reviews should be conducted objectively. Personal criticism of the author is inappropriate. Referees should express their views clearly with supporting arguments.

Acknowledgement of Sources

Reviewers should identify relevant published work that has not been cited by the authors. Any statement that an observation, derivation, or argument had been previously reported should be accompanied by the relevant citation. A reviewer should also call to the editor's attention any substantial similarity or overlap between the manuscript under consideration and any other published paper of which they have personal knowledge.

Disclosure and Conflict of Interest

Privileged information or ideas obtained through peer review must be kept confidential and not used for personal advantage. Reviewers should not consider manuscripts in which they have conflicts of interest resulting from competitive, collaborative, or other relationships or connections with any of the authors, companies, or institutions connected to the papers.

Reporting standards

Authors of reports of original research should present an accurate account of the work performed as well as an objective discussion of its significance. Underlying data should be represented accurately in the paper. A paper should contain sufficient detail and references to permit others to replicate the work. Fraudulent or knowingly inaccurate statements constitute unethical behaviour and are unacceptable.

Originality and Plagiarism

The authors should ensure that they have written entirely original works, and if the authors have used the work and/or words of others that this has been appropriately cited or quoted.

Multiple, Redundant or Concurrent Publication

An author should not in general publish manuscripts describing essentially the same research in more than one journal or primary publication. Submitting the same manuscript to more than one journal concurrently constitutes unethical publishing behaviour and is unacceptable.

Acknowledgement of Sources

Proper acknowledgment of the work of others must always be given. Authors should cite publications that have been influential in determining the nature of the reported work.

Authorship of the Paper

Authorship should be limited to those who have made a significant contribution to the conception, design, execution, or interpretation of the reported study. All those who have made significant contributions should be listed as co-authors. Where there are others who have participated in certain substantive aspects of the research project, they should be acknowledged or listed as contributors.

The corresponding author should ensure that all appropriate co-authors and no inappropriate co-authors are included on the paper, and that all co-authors have seen and approved the final version of the paper and have agreed to its submission for publication.

Disclosure and Conflicts of Interest

All authors should disclose in their manuscript any financial or other substantive conflict of interest that might be construed to influence the results or interpretation of their manuscript. All sources of financial support for the project should be disclosed.

CETAPS

Fundamental errors in published works

When an author discovers a significant error or inaccuracy in his/her own published work, it is the author's obligation to promptly notify the journal editor or publisher and cooperate with the editor to retract or correct the paper.

Consider the Candidate: Using Test-taker
Feedback to Enhance Quality and
Validity in Language Testing

David Ewing RYAN
| University of Veracruz

Abstract | This article discusses the importance for language test developers of taking candidate feedback into consideration in order to improve the overall quality and validity of their language tests. It reports on a study that was carried out at the Universidad Veracruzana, Mexico, between July and December 2010. The main objective of the study was to try and ascertain the positive and negative consequences for 245 candidates as a result of preparing for and taking a language test. The test was given in May 2010 during the spring application of EXAVER, a tiered-suite of English language certification tests developed and administered by the Universidad Veracruzana. A mixed methods approach was used for the study, incorporating both quantitative and qualitative data. The quantitative data came from the responses of a web-based questionnaire survey. The qualitative data came from the author's Research Journal, spanning a period of eight months, and from a series of semi-structured interviews. The findings of the study suggest that language test candidates not only have strong opinions (both positive and negative) about the tests they take, but they also have a strong desire to share those opinions with test developers. This type of feedback can then be used to substantially improve future tests, thereby helping to enhance the validity of the test system. The research provides a new perspective on a relatively unexplored area of language testing and has implications for language testing practitioners who welcome a more transparent and democratic form of assessment.

Key words | candidate feedback, candidate questionnaires, language test consequences, language test impact, consequential validity, collaborative language assessment, democratic language assessment, critical language assessment

Introduction

In his 2004 article in *Language Assessment Quarterly*, advocating the need to "broaden, deepen and consolidate" many of our ideas about language testing, Cumming makes the convincing argument that more research is needed on the role of stakeholders in language testing contexts that have traditionally been overlooked (3). It can be successfully argued that one of these neglected areas is Mexico, and, indeed, Latin America in general. Mexico seems to be in the paradoxical situation of many Latin American countries that, on the one hand, has seen a pronounced increase over the past several decades in demand for English language instruction (and the assessment that accompanies it) while, on the other, has seen a scarcity of studies investigating the specific variables that help define the uniqueness of Mexico's context. Without doubt, one of these variables would be the candidates who actually take language tests in Mexico.

The story of these candidates, not just in Mexico but in many counties and geographic areas throughout the world, is, to a large extent, an untold one. Indeed, in the language testing literature of the last ten to fifteen years, it would be difficult to find an issue that more scholars seem to agree on than the idea that candidates are one of the most important stakeholders in language testing and yet, paradoxically, one of the most neglected ones.

Hamp-Lyons, for example, notes that "many more studies are needed of students' views and their accounts of the effects on their lives of test preparation, test-taking and the scores they have received on tests" (299). Shohamy perceives that "it is through the voices of test takers who report on the testing experiences and consequences that the features of the use of tests can be identified. Yet, in the testing literature, test takers are often kept silent; their personal experiences are not heard or shared" (*The Power of Tests* 7). And Cumming maintains that "serious consideration of the uses of language assessment requires adopting research methods that investigate people's attitudes, beliefs, cultural values, and ways of interacting . . . . Such inquiry is indispensable for understanding why people perform the ways they do in language assessment, and thus necessary for validation" (9).

The purpose of this study, therefore, was to give free rein to what can be considered as the neglected voices of test candidates in one specific test project, in one specific place, and at one specific time. The principle reason for doing this is grounded in the concept of professional responsibility that comes with being a language test developer. As McNamara and Roever insist, "language testing has a real impact on real people's lives" (8). This impact starts with the stakeholders who are immediately affected by the test, such as the test candidates and test developers, and extends outward to society at large. This impact, in turn, implies a significant amount of responsibility on the part of test developers to ensure that the tests they write and administer are as valid and reliable as possible.

One of the most valuable techniques for helping test developers to measure the validity of their tests is, precisely, by listening to the voices of candidates. Candidate perceptions, feelings, points of view, attitudes, opinions and suggestions, taken together, can serve as evidence of the positive and negative consequences of tests. In addition, feedback from candidates can serve as the impetus for discussions that can, and should, be happening not just among several stakeholders, but among many (Madaus, qtd. in Shohamy, *The Power of Tests* 149). Enlarging the dialogue in this way can help further promote not just the validity of individual tests, but the validity of the language test system as a whole, which, according to Shohamy, needs to continually "encourage testers, teachers, test takers, and the public at large to question the uses of tests, the materials they are based on and to critique the values and beliefs inherent in them" (*The Power of Tests* 131).

The article is divided into seven parts. Part 1 contains a literature review and theoretical overview of consequential validity, or the concept of looking at language tests in terms of the positive and negative consequences that are sustained by the candidates who take such tests. Part 2 gives a brief summary of the goal of the study. Part 3 contains a summary of the EXAVER English language certification tests, which served as the practical context of the study. Part 4 explains the methodology used in the study. Parts 5 and 6 offer, respectively, an overview of the findings and a discussion of those findings. Finally, Part 7 offers some general conclusions about the topic at hand.

1. Theoretical Context of Study: Consequential Validity or the Social Side of Language Testing

1.1. Scholarly Interpretations of Consequential Validity

Traditionally, consequential validity in language testing has been seen as a type of validity theory that attempts to measure the consequences of the way in which the scores from language tests are interpreted and used (Davies et al. 131). These consequences are also known as effects or impact. One of the first scholars to discuss the concept was Spolsky who reminded language testers about the important consequences their decisions may have on the lives of test candidates. He therefore urged testers to ensure that the evidence they present about the inferences they make regarding candidates' test scores is as credible and compelling as possible (Spolsky, qtd. in Bachman, "Building and Supporting a Case for Test Use" 5).

Cronbach was one of the first scholars to enlarge the concept of consequential validity to embrace society at large. He felt that the judgments that language testers make about the positive or negative consequences of a test were highly influenced by society's views of what is desirable or undesirable in the specific cultural context in which the assessment takes place. He also felt that these societal views are not stagnant but change over time (Cronbach, qtd. in McNamara and Roever 11). Cronbach's belief in the importance of social and cultural values in language testing is an essential one and was taken up by other theorists such as Messick, and Bachman and Palmer, as will be seen subsequently.

Along with Spolsky and Cronbach, another scholar who contributed immensely to the current understanding of test consequences was Messick, who defined consequential validity not as a theory in and of itself, but rather as one of the six defining aspects of construct validity (Messick, "Validity"; "The Interplay of Evidence and Consequences in the Validation of Performance Assessments"; and "Validity and Washback in Language Testing"). Messick's motivation for describing consequential validity in this way might be best explained by his definition of test constructs as being the embodiment of social values (McNamara 334).

Following Messick's important contributions to the understanding of consequential validity, the next major treatment of the topic was by Bachman, in *Fundamental Considerations in*

Language Testing. This contains a detailed section devoted to "the consequential or ethical basis of validity" (279) in which the author echoes Cronbach's and Messick's thoughts above on the important role that values play in test impact and that these values (or considerations) "are essentially political, that they change over time and they will differ from one society to another" (280).

Bachman and Palmer include an entire section in their book exploring the impact that language tests have on test candidates in particular. The authors identify three ways that language tests have a direct impact on candidates: 1) as a result of candidates' preparing for and taking the test;[1] 2) from the type of feedback that candidates receive about their performance on the test; and 3) from the decisions that testers make based on candidate scores (31).

More recent scholars, such as McNamara and Roever, and Shohamy, have also written extensively on the topic of consequential validity. McNamara and Roever wrote an entire volume that is devoted to exploring the various social considerations involved in language testing and that develops the authors' belief that a language test with positive psychometrical qualities does not necessarily mean that it will have positive social consequences (2). Shohamy is undoubtedly one of the most passionate supporters of fairness and ethics in language testing; regarding consequential validity, she stresses the need for researchers to carefully describe the nature of test consequences. She also stresses, however, the challenges and complexities that researchers may encounter in doing so, since these consequences are often invisible due to the fact that they tend to take place "outside the domains which the researcher examines" (*The Power of Tests* 49).

Shohamy also joins a fairly long list of other scholars (see Alderson and Wall; Bailey; Cheng, Watanabe and Curtis; Hamp-Lyons; Hughes; Messick, *Validity and Washback*; Wall, *Impact and Washback*; and Wall, *The Impact of High-Stakes Examinations* to name just a few) who argue that another important area of test consequences is washback, which can be defined as "the effect of testing on teaching and learning" (Hughes, qtd. in Bachman and Palmer 30). In language testing, a good example of positive washback is that identified above by Bachman and Palmer when they discuss the feedback that testers and candidates should, ideally, receive from

each other about the test, and this idea of reciprocal or mutual feedback shows up repeatedly throughout Shohamy's work, notably in her idea of "democratic" or "collaborative" assessment.

Finally, O'Sullivan echoes the concerns of earlier scholars (notably those of Cronbach, Messick, Bachman, and McNamara and Roever) who focused on the social values that are implicit in the constructs that inform language tests. According to O'Sullivan, from the very beginning of the test development process onward, test developers need to play close attention to the particular characteristics (whether they be individual, linguistic, cultural, social, or a combination of these) of the population that will be using the test, and in so doing test developers "are actually taking into consideration aspects of test consequence" (6).

For O'Sullivan, as shall be seen in Section 2.3., language tests that exhibit a high level of concern for both the local test context, and for the particular needs or realities of the candidates within that context, can be considered examples of the phenomenon he defines as language test "localization" (6).

1.2. The Business Approach

During the course of the study, the researcher decided to adopt his own approach to observing the relationship between language tests and the candidates who take them, and this was done through the lens of business administration.

It can be argued that the schools, universities, and other institutions that employ language teachers and testers, are businesses[2] in the sense that they offer:

1. a product (e.g. knowledge of a language or certification of that knowledge)
2. people who might be seen as selling the product (e.g. the teachers who work at a language school, or the test developers who write a language test)
3. people who purchase the product (e.g. the students at a language school, or the candidates who take a language test)

It also seems fair to assume that in order for any business to be successful, it needs to be aware of two key variables: the quality of the product it is trying to sell, and knowledge about the client

(or customer) base that the product is designed for. In other words, a successful business usually needs to have a well-designed and properly functioning product, but, just as important, it also needs to be familiar with and have an understanding of the requirements of the people who will eventually purchase the product.

In keeping, therefore, with the way that scholars such as Bachman and Palmer have thought about consequential validity in terms of the impact, or consequences, that language tests have on candidates, consequential validity also has to do with testers taking the time to get to know their candidates, and, more importantly, taking the time to familiarize themselves with what candidates think of the product that testers are selling them, namely language tests.

One of the best ways that testers have of ascertaining this information is through candidate feedback questionnaires that can be distributed immediately following the test administration, or, in some instances, prior to the test administration, or in still further instances, both prior to *and* following the administration, which would yield a more complete picture of candidate attitudes about the test. This information can then be used as a valuable set of qualitative data that can help compliment the quantitative data that testers receive from such elements as item analysis and descriptive statistics. While statistics are indispensable for helping to inform testers about the overall psychometric quality of their tests, candidate feed-back questionnaires can provide language testers with another way of measuring overall quality, as seen through the positive and negative consequences that candidates experience as a result of preparing for and taking a language test.

The reason of course that test developers should desire this information is so that they can identify: a) aspects of the test and the test system that seem to be working well for candidates, b) aspects of the test and the test system that seem to *not* be working well for candidates, and c) suggestions that candidates might have for improving the test and the test system. Information gleaned from any or all of these areas can then be used to substantially improve the overall quality of the test and the test system.

The topic of candidate feedback questionnaires will be discussed in greater depth in Parts 4, 5, and 6.

1.3. Importance of the Language Test System

Loosely defined, a language test system can be seen as the collective entity of all the elements that are exterior to the test *per se*, but are still indirectly related to the test. These elements might also be referred to as the "peripheral aspects" of the test, and they include, among other things, such variables as:

- **the test registration process** or how candidates are able to enrol for the test. In the case of the EXAVER tests to be discussed in Part 2, as well as for many other language certification tests, the registration process is completed online via a website.

- **the test orientation and preparation process**, which includes all the information *about* the test, including practice tests that candidates can, and should, take in preparation for the test. Again, in the case of EXAVER, all or most of this process is online.

- **the test "reception" process**, or the way that candidates are physically greeted and treated both prior to and during the test by the examiners (or invigilators) who administer the test. It is important for testers to know, for example, whether the behaviour of the examiners was calm, welcoming, and impartial, or nervous, rude, and biased.

- An example illustrating why these different elements of the test system are important can be seen in the work of Bachman and Palmer. They suggest that one way of promoting positive test impact is by involving candidates at various stages of the test development process and by soliciting candidates' feedback on such things as their overall impression of the test, as well as their impression of more detailed aspects such as the tasks that appear in the test. Bachman and Palmer argue that if candidates are involved in this way, it is very possible that they will have a more positive perception of the test, will have more confidence and motivation when taking it, and, as a result, will very likely perform better (32).

The same argument might be made for why candidates should be highly encouraged to take a practice test, which, as noted above, can be considered to be part of the "test orientation and preparation process". If candidates perceive the tasks in the practice test as being fair and of good quality (or, in the words of Bachman and Palmer, as being "authentic" and "interactive"), then, according to Bachman and Palmer's hypothesis, this would likely serve to increase candidates' sense of confidence and motivation regarding the practice test, which could, likewise, help them to perform better on a live test. By the same token, but related to the "test reception process", if candidates are treated with respect and courtesy by invigilators on the day of the test, this could help them to feel more at ease, and to concentrate better, which, once again, might help them to perform better. The implications for test validity are clear: testers obviously want candidates to perform to the best of their ability, so that the scores calculated based on candidate performance are as fair, valid, and authentic as possible.

To conclude, therefore, elements of the test system can be considered to be related to the validity of the test, in the sense that if these elements are sufficiently attended to, this could help candidates to form a positive impression of the test and to perform to the best of their ability, which would, in turn, enhance the overall validity of the test. But if these elements are not attended to, then the opposite scenario might occur and candidates could form a negative impression of the test which could, then, interfere in their performing to the best of their ability, which would likewise diminish the overall validity of the test.

2. Goal of the Study

The study focused specifically on what Bachman and Palmer considered to be one of the three ways that language tests have a direct impact on test candidates, namely, in terms of the consequences that candidates experience as a result of preparing for and taking these tests (31). In order to measure this impact, it was necessary to liberate the voices of the test candidates who participated in the study, and this, then, became the primary goal of the study. This was accomplished, first, by soliciting candidates' opinions both about the test they took and about the

test system, and second, by soliciting their suggestions of ways to improve the test and the test system.

3. Practical Context of Study: The EXAVER English Language Certification Tests

3.1. General Description

EXAVER is the name of the tests that were used as the basis of the study, and refers to a tiered-suite of English language certification tests that are administered by the Universidad Veracruzana (UV), in the south-eastern Mexican state of Veracruz. The suite was developed in the year 2000 by a small group of English language teachers from the UV, as well as representatives from the following international organizations: The British Council, Cambridge Assessment, and Roehampton University's Centre for Language Assessment and Research (CLARe).

The construct behind the EXAVER tests is to measure three language proficiency levels that are informally linked to the Council of Europe's *Common European Framework of Reference for Languages* (CEFR) in that each EXAVER level is based on the Council of Europe content specification that inspired the corresponding CEFR level (Council of Europe 23). This is summarized in Table 1 below. The EXAVER tests are administered twice a year, once in the spring and once in the fall, at 11 separate language centres throughout the state of Veracruz.

EXAVER	CEFR	Council of Europe
1 Upper Beginner	A2	Waystage
2 Lower Intermediate	B1	Threshold
3 Upper Intermediate	B2	Vantage

Table 1: *Levels of EXAVER tests and their corresponding CEFR L* (adapted from Abad et al.)

3.2. Test Structure

Each EXAVER test contains three separate sections, or papers, and the structure of each of these sections is described in Table 2.

Paper 1 Reading and Writing	Paper 2 Listening	Paper 3 Speaking
• 5 parts • Variety of tasks: matching, multiple choice, modified cloze text • Indirect measure of writing	• 4 parts • Range from comprehension of relatively short informal conversations to comprehension of more formal and substantially longer conversations	• 3 parts • Combine some type of interview task (interlocutor to candidate), discussion task (between a pair of candidates) and a long-turn task (interlocutor to candidate)

Table 2: *EXAVER test structure* (after Dunne)

3.3. Test Localization

According to O'Sullivan, one of the defining characteristics of the EXAVER examinations is that they represent "the first systematic attempt to create a 'local', affordable, and sustainable language test system" (10). In focusing their attention on the local geographic context where the examinations take place (e.g. south-eastern Mexico), and on the particular needs of the test candidates within that context (students, primarily, of the Universidad Veracruzana), EXAVER's test developers helped create a process now known as "localization". O'Sullivan defines this as "the practice of taking into account those learning-focused factors that can impact on linguistic performance . . . [and] the recognition of the importance of test context on test development . . ." (6).

Economic affordability was one of the first local variables that EXAVER's test developers took into consideration. Due to the fact that the majority of EXAVER's candidates were (and continue to be) unable to afford the cost of more reputable international English language certification tests, EXAVER's test developers decided to create a suite of economically affordable tests, more in line with median to lower income brackets based on the Mexican minimum wage.[3]

Table 3 shows the current costs (as of September 2014) of taking an EXAVER test, with their approximate equivalents in Euros.[4]

LEVEL	Cost in MX Pesos	Cost in Euros
EXAVER 1	350	Approx 21
EXAVER 2	400	Approx 23
EXAVER 3	450	Approx 25

Table 3: *Comparative cost of taking an EXAVER test*

4. Methodology

4.1. Type of Data and Participants

The study included both quantitative and qualitative data. The quantitative data came from the responses of 245 EXAVER candidates who completed a web-based questionnaire survey, administered in the summer of 2010, following the spring 2010 test administration of EXAVER's three levels. The qualitative data came from the author's Research Journal, spanning a period of eight months, from March to October 2010, and from a series of semi-structured interviews conducted in October 2010 with four of the questionnaire's respondents.

4.2. Type of Method

A mixed methods approach for data collection and analysis was used for the study, starting with a quantitative investigation in the form of a web-based questionnaire and then followed by a qualitative study, in the form of semi-structured interviews. In practical terms this meant that at the end of the questionnaire, participants were able to tick a box and include their name and email address, signifying their desire to be contacted by the researcher for the second phase of the study.

4.3. Data Collection & Analysis of Quantitative Phase[5]

The web-based questionnaire included 44 questions, comprising 42 closed-format, multiple-choice questions, and two open-ended questions. Of the closed-format questions, 10 employed a Likert Scale, with options spanning from 1 to 5, as a way of ascertaining candidates' opinions or feelings about a variety of topics related to the test they took and to the test system. Excel Version 2003 was used to analyse the data.

5. Findings

Out of the total 964 candidates who took an EXAVER test in May 2010, 245 of them (or 25%) responded to the survey. Of these 245 candidates, 99 (40%) ticked the box at the end of the survey demonstrating their desire to participate in the semi-structured interviews that constituted the second phase of the study. This, therefore, was the first finding of significance, namely, that such a large percentage of candidates wished to participate in the second phase of the study and further elaborate on their opinions about the process of preparing for and taking the test. This indicated an apparent high level of interest among EXAVER's candidates to have their voices heard.

Web-based questionnaire survey

As research instruments, questionnaires, like all forms of data collection, have their own distinct advantages and disadvantages. In terms of the latter, researchers sometimes complain about the lack of depth and richness that is found in multiple-choice responses (Dörnyei 115). For this reason, the researcher chose to include two open-ended questions in the survey, along with the overwhelming majority of 42 multiple-choice questions. While the responses to all of the survey's questions provided important feedback, the responses to the two open-ended questions (identified in the survey as numbers 17 and 30) are noteworthy, due both to the high number of candidates who responded to them (well over half of the total 245 candidates who responded to the survey), as well as to the diversity of their answers. Summaries of these responses follow.

Question 17

The text for question 17 read:

> "Do you feel that there is anything we could include on the EXAVER website that might help future candidates to feel less anxious and/or more confident before taking the test? If so, please write your comment(s) below, taking all the space that is necessary."

Question 17 yielded 144 total responses, including:

- 23 positive comments, such as:

  - "The teachers who administered the test were excellent, and they worked well together as a team."

  - "I didn't hire a tutor or use any books to prepare for the test, but I found the information on the website very useful."

  - "Everything on the website is very clear – congratulations!"

  - "EXAVER is an excellent alternative for certifying your level of English, and it's great that it was developed here at our university."

- 24 negative comments, such as:

  - "The waiting time to get your grade is too long... you really need to find a way to make it go faster."

  - "The noise from traffic in the street at the language centre where I took the test made it very difficult to hear the CD during the listening section."

  - "My chances of passing the test would have been much better if the preparation materials had been more accessible; I received an 'error' message when I tried to open the Sample Tests on the website."

  - "I would have benefited from a greater variety, and greater scale of difficulty, of test preparation materials – the Sample Tests on the website were really easy and not very helpful."

- 97 suggestions, such as:

  - "Include a video on the website of a sample Speaking Test".

- *"Include testimonies or opinions on the website from past candidates who had a positive experience taking the test."*

- *"Include a bibliography on the website of literature to consult for helping candidates to prepare for the test."*

- *"Include a description (either on the website or included with the diploma) of how grades are calculated."*

Question 30

The text for question 30 read:

"Do you have any other comments (positive or negative) and/or suggestions that you'd like to add regarding the EXAVER test you took or about the EXAVER Project in general? If so, please write them below, taking all the space that is necessary."

Question 30 yielded 127 responses, including:

- 38 positive comments, such as:

  - *"I liked the test – everything seemed very clear and precise. Thanks."*

  - *"The EXAVER staff appeared to be very knowledgeable and when they gave the instructions in English, it was very clear, which set me at ease and made me feel more confident."*

  - *"I feel lucky for having had the opportunity of taking an EXAVER test and it was a great experience."*

  - *"Everything was fine, and the level seemed very appropriate."*

- 61 negative comments, such as:

  - *"It was very tedious waiting so long to take the Speaking Test."*

  - *"It was frustrating having to physically go to the language centre where I took the test in order to get my diploma – they only give them out in the mornings, and I disagree with this policy."*

○ *"While waiting in line to enter the test centre, I was told that my name was not on the list even though I had my registration receipt. In the end I was able to take the test, but I felt very nervous."*

○ *"The pencil they gave me for filling in the Answer Sheet was of really poor quality."*

- 28 suggestions, such as:

  ○ "*It would be nice to have a more detailed report on how I fared on the test, such as knowing how I performed on each part of the test maybe in terms of percentages."*

  ○ *"In order to accommodate the needs of students, there should be more applications of the tests than just twice a year."*

  ○ *"You should design a course of several months duration that students could take for helping them to prepare for the test."*

  ○ *"There should be more publicity for the tests, especially for those of us who are not students of the Universidad Veracruzana, but rather from the community at large."*

6. Discussion

6.1. Specific Concerns

The phrasing of Question 17 in the web-based survey, with special emphasis on the words "more confident" and "less anxious", was intentional as a way of reflecting the researcher's premise[6] that the less anxious and more confident candidates feel before taking a language test, the more likely they are to perform better. The relatively long list of suggestions (97 in total) that candidates gave in response to this question have, therefore, proven quite useful in helping EXAVER's test developers and administrators to improve the quality of the preparation materials on the exam board's website so that candidates can, indeed, feel more confident and less anxious before taking a live test.

Question 30 in the web-based survey should seem familiar to researchers, since it is the classic *"Do you have anything else to say?"* type of query that is usually included as the final

question in an oral interview. It was considered necessary to use it as an open-ended question in the survey as a type of "safety net" in order to ensure that candidates were given the opportunity of stating anything, and everything, they wished to state about the process of preparing for and taking an EXAVER test.

One of the negative responses to Question 30, referring to a candidate's sense of anxiety over their name not being found on the official list of registrants for the test, relates to the theme of Question 17. It should serve to remind testers of the importance of taking measures to avoid circumstances that might create unnecessary stress or anxiety for candidates on the day of the test. One way of doing this is for test examiners and administrators to meet together and develop a list of all the things that could feasibly go wrong on the day of the test, and then to come up with an effective way of dealing with each of them. Each potential problem and its corresponding solution could then appear on a printed sheet of paper that could be given to invigilators on the day of the test.

By contrast, one of the positive responses to Question 30 illustrates how what might be interpreted as a rather routine, mundane task (reading the initial instructions once candidates are seated) can actually serve to minimize stress and anxiety, and to boost candidates' sense of confidence, provided the instructions are read calmly and clearly. Both of these examples serve to reinforce the importance of ensuring that the test "reception" process (see Section 1.3.) is as smooth and professional as possible.

6.2. General Concerns

Candidate responses from both the questionnaire survey and the semi-structured interviews provided a rich representation of the diversity of opinions, feelings, perceptions, and attitudes that EXAVER candidates have about the tests they take. They also provided EXAVER's test developers with important insight regarding some of the positive and negative consequences for test candidates as a result of preparing for and taking a language test. With particular regard to the questionnaire survey, the quantity and variety of responses bring to mind Shohamy's

observation that the overwhelming majority of test candidates not only have a strong need and desire to express their feelings about the test they took, but they also have the inherent *right* to do so, and it is the responsibility of language teachers and testers to *enable* them to do so (*The Power of Tests* 156). By providing for this, she feels that testers can help democratize the act of taking a test so that the experience becomes more of a collaborative, horizontal process, rather than an authoritarian, top-down one (136-37).

It can be argued, however, that the most important step that takes place in the overall process of soliciting candidate feedback is what testers finally end up doing with this feedback after receiving it. For this reason, one might correctly refer to the "final consequences" of consequential validity, for it is the final actions that test developers take regarding candidate feedback that could serve to increase the likelihood of positive consequences occurring for future candidates and, accordingly, could serve to decrease the likelihood of negative consequences occurring for those candidates.

As a way of illustrating how a language test board can convert candidate feedback into concrete actions that will hopefully generate positive impact for future candidates, the following is a list of actions that EXAVER has already undertaken or is currently undertaking based on candidate feedback from this and other studies:

- Streamlined registration process, making it much easier for current and future candidates to register for the tests.
- New online grade allocation process to substantially reduce the waiting time for receiving grades.
- Sample Speaking Test for each of EXAVER's three levels, uploaded to the EXAVER website so that potential candidates have an idea of the format of the test, as well as the type of tasks they can expect to encounter. These tests serve to compliment the Sample Reading, Writing, and Listening Tests that have appeared on the website since EXAVER's inception.

- Drafting of a document with a list of administrative procedures that can potentially be problematic for examiners and invigilators on the day of the test, along with their corresponding solutions.

- Dissemination of candidate feedback questionnaires in order to continue to monitor the positive and negative consequences for the candidates who take the tests.

- Analysis and discussion of appropriate action(s) to take based on candidate responses to the questionnaires.

- Follow-through to confirm that appropriate action was in fact taken.

7. Conclusion

By now it has perhaps become apparent to the reader that what candidate feedback and consequential validity in language testing actually relate to is a type of assessment that is more inclusive and democratic in nature than the traditional, authoritarian type of model that was prevalent in so many assessment contexts throughout the world during much of the twentieth century and, indeed, prior to that.[7]

When test developers refuse to solicit candidate feedback, or do so without following through on it, this only serves to reinforce the undemocratic nature of the assessment, and the power and control that testers often exert over the candidates who take their tests. Conversely, when test developers solicit candidate feedback and then take positive actions based on that feedback, this serves to strengthen the overall democratic nature of the assessment by revealing a horizontal, collaborative process. Moreover, this process encourages the participation of not merely a few, but a wide variety of stakeholders, thereby strengthening even further the democratic nature of the process.

Another important point that language test developers should consider when judging the validity of their assessments is that language testing, like any type of testing, is, at best, an inexact science. There is an innumerable amount of things that can go wrong on the day of the test and that can interfere in its validity. The air conditioning in a hot and humid room could suddenly stop

working, thereby forcing candidates to finish the remainder of the test in uncomfortable physical conditions. Or an oral examiner could ask a candidate what s/he did on her last vacation, without knowing that someone in the candidate's immediate family died at that time. In both of these not overly extraordinary cases, the candidate's concentration could feasibly be thrown off, thereby negatively affecting his/her performance on the test, which would also mean that the score the candidate receives on the test is not a true reflection of his or her ability.

The above examples represent real situations that have taken place during previous EXAVER test administrations. Due to the fact that language testers work with real people in the real world, real problems are bound to occur, and there is very little that testers can do to ensure that these problems will no longer occur in the future. There are, however, many things that language testers are in fact able to control when it comes to designing and administering their tests. These include the following:

- Concern for the test's most important stakeholder: the candidate.
- Collective elements of the test system such as the test registration process, the test orientation and preparation process, and the test reception process.
- The overall quality of the test *per se*, e.g. its reliability, and validity of construct.
- Being responsible and effective examiners, e.g. giving fair and non-partial treatment to all candidates and following-up after the test by writing a "post-exam" report with a list of things that went right and wrong during the test application.

By concerning themselves with these and other important variables, language testers can help safeguard the overall fairness and integrity of the test and the test system. In so doing, they also help to underscore the difference between assessments that, on the one hand, are moving towards a more dynamic, responsible, and democratic model, and on the other hand, ones that continue to remain more stagnant and conventional in nature.

Notes

[1] As shall be seen in Part 2, the focus of the study was on this first type of impact described by Bachman and Palmer.

[2] The use of "business" here stems *not* from an interpretation of the word focused on such variables as volume and profitability, but rather, in a more general sense, as a synonym for a place providing an exchange of goods.

[3] As of December 2014, the Mexican minimum wage was approximately 61 pesos per day.

[4] For more details on the EXAVER examinations and the EXAVER test system, especially as they relate to localization, see Abad et al. "Developing affordable, 'local' tests: the EXAVER Project," in *Language Testing: Theories and Practices*. Ed. Barry O'Sullivan. Palgrave Macmillan, 2011. 228-43.

[5] Due to space considerations, information related to the methodology and findings from the second (qualitative) phase of the study could not be included here, but is available by contacting the author at <dewing@uv.mx>.

[6] This premise was itself based on Bachman and Palmer's similar hypothesis. See Section 1.3.

[7] The traditional or authoritarian model of education and assessment is of course still prevalent in many parts of the world today, including in many educational contexts in Mexico.

Works Cited

Abad, Adriana Florescano, B. O'Sullivan, C. Sanchez Chavez, D. Ewing Ryan, E. Zamora Lara, L. A. Santana Martinez, M. I. Gonzalez Macias, M. Maxwell Hart, P. Evelyn Grounds, P. Reidy Ryan, R. A. Dunne, and T. de Jesus Romero Barradas. "Developing Affordable 'Local' Tests: the EXAVER Project." *Language Testing: Theories and Practices*. Ed. Barry O'Sullivan. Oxford: Palgrave Macmillan, 2011. 228-43.

Alderson, Charles, and Dianne Wall. "Does Washback Exist?" *Applied Linguistics* 14.2. (1993): 115-29.

Bachman, Lyle. *Fundamental Considerations in Language Testing.* Oxford: Oxford UP, 1990.

- - - . "Building and Supporting a Case for Test Use." *Language Assessment Quarterly* 2 (2005): 1-34.

Bachman, Lyle, and Adrian Palmer. *Language Testing in Practice.* Oxford: Oxford UP, 1996.

Bailey, Kathleen. "Washback in Language Testing." *TOEFL Monograph* 15. Princeton, NJ: Educational Testing Service, 1999.

Cheng, Liying, Yoshinori Watanabe, and Andy Curtis. *Washback in Language Testing: Research Contexts and Methods.* Mahwah, N.J.: Lawrence Erlbaum & Associates, 2004.

Council of Europe. *Common European Framework of Reference for Languages: Learning, Teaching and Assessment.* Cambridge: Cambridge UP, 2001.

Cumming, Alister. "Broadening, Deepening and Consolidating." *Language Assessment Quarterly,* 1 (2004): 5-18.

Davies, Alan, A. Brown, C. Elder, K. Hill, T. Lumley, and T. McNamara (1999). "Dictionary of Language Testing." *Studies in Language Testing.* Vol. 7, Cambridge: University of Cambridge Local Examinations Syndicate, 1999. 118-19.

Dörnyei, Zoltán. *Research Methods in Applied Linguistics.* Oxford: Oxford UP, 2007.

Dunne, Roger. "The EXAVER Project: Conception and Development." *MEXTESOL Journal,* 31.7 (2007): 23-30.

Hamp-Lyons, Liz. "Washback, Impact and Validity: Ethical Concerns." *Language Testing* 14 (1997): 295-303.

Hughes, Arthur. *Testing for Language Teachers.* Cambridge: Cambridge UP, 2003.

McNamara, Tim. (2001). "Language Assessment as Social Practice: Challenges for Research." *Language Testing* 18 (2001): 333-49.

McNamara, Tim, and Carl Roever. *Language Testing: The Social Dimension*. Oxford: Blackwell Publishing, 2006.

Messick, Samuel. "Validity." *Educational Measurement*. Ed. R.L. Linn**.** 3rd ed. New York: American Council on Education, 1989. 13-103.

---. "The Interplay of Evidence and Consequences in the Validation of Performance Assessments." *Educational Researcher* 23 (1994): 13-23.

---. "Validity and Washback in Language Testing." *Language Testing* 13 (1996): 241-56.

O'Sullivan, Barry, ed. *Language Testing: Theories and Practices.* Oxford: Palgrave Macmillan, 2011.

Shohamy, Elana. "Testing Methods, Testing Consequences: Are They Ethical? Are They Fair?" *Language Testin,* 14 (1997): 340-49.

---. "Democratic Assessment as an Alternative." *Language Testing* 18.4 (2001): 373-91.

---. *The Power of Tests: A Critical Perspective on the Uses of Language Tests.* London: Longman, 2001.

Wall, Dianne. "Impact and Washback in Language Testing." *Encyclopedia of Language and Education*. Eds. Caroline Clapham and David Corson. Vol.7: Language Testing and Assessment. Dordrecht. The Netherlands: Kluwer Academic Publishers, 1997. 291-302.

---. *The Impact of High-Stakes Examinations on Classroom Teaching, a Case Study Using Insights from Testing and Innovation Theory.* Cambridge: Cambridge UP. 2005.

Using Translation to Teach Native and Non-native Varieties of International English

Luís GUERRA
| University of Évora - ULICES

Abstract | The term "New Englishes" attempts to cover the large number of varieties of English, far from uniform among themselves in their features and use and different from the historically and culturally established British and American standards. Over the past years, these New Englishes have been more acknowledged in the foreign language class. Linguists have called attention to the importance of increasing the learner's linguistic awareness by covering topics of "linguistic variation and varieties of many types: national, regional, social, functional, international" (Gnutzmann 167). This paper aims at discussing the advantages and possibilities of teaching native and non-native English varieties in the foreign language class. It presents some data included in a Foreign Language and Translation course which attempted to integrate linguistics and translation by analysing the features of African American Vernacular English, Singapore English, Indian English and Australian English in terms of their phonological, morphological, lexical and syntactic levels.

Key words | World Englishes, native/non-native varieties

1. Introduction

Applied linguists and language educators have been promoting lively debates over how globalization has been affecting the English language and how English has been influencing globalization. The increasing numbers of non-native speakers, the emergence of New Englishes, the use of English for intercultural communications, the intelligibility of standard and non-standard varieties of English, are just a few of the most talked about topics.

Discussing the worldwide development of English, Crystal (1997) proposed that the English language has achieved its present global status due to the two ways in which it has been employed by countries all over the world. First, where English has some kind of special status, it has been made the official language of several countries and used in diverse contexts such as the government, the legal system, commerce, the media, and the educational system. In such countries, English is characterised as a "second language", as a complement to the speaker's native language. Second, in other countries English has no official status and it is learnt in schools as a "foreign language". Though Crystal makes use of the distinction between second or foreign use of the language to explain the worldwide importance of English, he points out that such distinction has lost some of the relevance it may have had. It is argued that one may find more use of English in some countries where it is learnt as a foreign language than in some of the countries where it has been described as a second language.

In an attempt to describe the cultural and linguistic developments of the English language at the turn of the century, McArthur was able to identify three different backgrounds as far as the existence of standard varieties of English are concerned. First, that at the end of the twentieth century two standard varieties, British English and American English, were long-established and broadly accepted, and for many the only legitimate varieties of English. Second, that some Anglophone nations (Australia, Canada, New Zealand and South Africa) had already developed their "clear-cut national standards" (5) becoming less dependent on the British and American norms. Finally, though not as consensual as in the previous context, that standardizing processes

have begun in some countries such as India, Nigeria, Singapore and Malaysia, leading to the development of the so-called World Englishes.

2. *New* Englishes or *World* Englishes

According to Jenkins, "the term 'New Englishes' covers a large number of varieties of English which are far from uniform in their characteristics and current use" (22) although they might share some features. Platt, Weber and Ho define New English as a variety which fulfils the following criteria: 1) it has developed through the education system; 2) it has developed in an area where a native variety of English was *not* the language spoken by most of the population; 3) it is used for a range of functions *among* those who speak or write it in the region where it is used; and 4) it has become "localized" or "nativized" by adopting some language features of its own.

Currently, there are a considerable number of texts on World Englishes. Platt Platt, Weber and Ho highlight the importance of linguistic features (i.e. sounds, sentence structures and special expressions) which make it possible to define a particular New English and provide an extensive list of the functions of the New Englishes in several parts of the world, their similarities and differences in terms of accents and stress patterns, vocabulary, morphological, syntactic and pragmatic features.

Mesthrie and Bhatt carried out a similar analysis of World Englishes. However, they take a broad view of the term, including prototypical varieties like Indian English or Nigerian English, but also varieties like Black South African English or individual varieties of native American Indian English, varieties which have undergone language shift (i.e. Indian South African English) or are in the process of doing (i.e. Singapore English). Essentially, Mesthrie and Bhatt identify the "regular and widely, informally accepted features" (47) of World Englishes in the realm of phonetics and phonology, syntax, lexis and pragmatics.

Equally, Melchers and Shaw presented an overview of the varieties of English around the globe in their phonological, lexical, grammatical and pragmatic dimensions. Alternatively, Melchers and Shaw's description of world Englishes follow Kachru's concentric circles (the *Inner Circle of*

English, made up of "norm-providing varieties", the *Outer Circle of English*, including "norm-developing varieties" and the *Expanding Circle of English*, with "norm-dependent varieties") and on the geographical location of the varieties. Their description focuses mainly on varieties of the Inner Circle (England, Wales, Scotland, Ireland, the USA, Canada, Australia, New Zealand, South Africa, Liberia and the Caribbean) and the Outer Circle (South Asia, Africa and South East Asia).

Significantly, Kirkpatrick offers a thorough description of a wide range of native and non-native varieties and then discusses the implications of these varieties for English language learning and teaching. In essence, Kirkpatrick's description of selected varieties of World Englishes takes into consideration the pedagogical consequences of adopting an alternative approach to English varieties in the language classroom.

3. World Englishes and Teaching and Learning English as a Foreign Language

As early as the 1980s, some linguists attempted to establish a framework for teaching English as an international language (EIL). In order to identify the major features of EIL, Campbell et al. introduced three principles of teaching international English: 1) knowledge of the different social and cultural patterns and rules present in communicative exchanges involving speakers of more than one country or culture; 2) training native speakers in the use of English in international contexts; and 3) training non-native speakers in the use of language with native as well as with non-native speakers.

Several other authors have reported on significant changes to be introduced in teaching the language. If we are to accept English as an international language of communication and incorporate these characteristics into the classroom, educators in the field of English language teaching will have to take on some responsibilities. Trifonovitch pointed out some aspects that need to be emphasised in the classroom. Among those, he suggests that as speakers of English will be contacting a variety of cultures – native and non-native – teachers should not concentrate on the cultures of the native speakers.

Modiano identified two major areas in the teaching of EIL and their scope: language varieties and culture. Modiano believes that when teachers only emphasize AmE or BrE, students tend to perceive other varieties as less valued. Such approach to teaching "presents English as the property of a specified faction of the native-speaker contingency" ("Linguistic Imperialism, Cultural Integrity, and EIL" 340). Modiano also stresses that when students need to learn English as a tool for intercultural communication seeking competence in an international perspective on the language, they are supposed

> to develop the ability to comprehend a wide range of varieties but also strive to utilize language which has a high likelihood of being comprehensible among a broad cross-section of the peoples who comprise the English-using world. ("Ideology and the ELT Practitioner" 162)

In Modiano's opinion, teaching and learning English based on an international frame of reference aiming at developing such competence is superior "when compared to the conventional integration-orientated practices associated with the learning of culture-specific varieties such as British English", what he calls a "nation-state centred view" ("Linguistic Imperialism" 340).

According to Modiano, in order to promote cultural equality, "a multiplicity of teaching practices, and a view of the language as belonging to a broad range of peoples and cultures, is the best that language instructors can do" (340). Baxter seems to share the same viewpoint when he says that "teaching materials should be drawn from all the various English-using communities, not only L1 communities, so as to introduce students to the different manners of speaking English and to build an attitudinal base of acceptance" (67).

Kirkpatrick suggests that "courses in World Englishes are becoming ever more popular and are seen, especially among ELT practitioners and professionals, as relevant for those who plan to become English language teachers" (1). Kirkpatrick acknowledges that "the model of English that should be used in the classrooms in expanding (EFL) circle countries has been a subject of discussion for some time" (3). Moreover, he believes that "the curriculum should

comprise the cultures of the people using the language for cross-cultural communication rather than Anglo-American cultures" (3), thus questioning the appropriateness of native speaker models and their cultures.

Moreover, Kirkpatrick adds that the current model for the language classrooms in outer and expanding circle countries may follow one of these alternatives: 1) adoption of an exonormative (Inner Circle) native speaker model, or 2) adoption of an endonormative nativised model. Regardless of the model adopted, but especially in the case of the expanding circle countries which have usually followed the native speaker model, ELT should incorporate activities which allow students to develop awareness of the multiple forms of English. As Kirkpatrick puts forward,

> in aiming to teach and learn English in ways that would allow for effective communication across linguistic and cultural boundaries the focus of the classroom moves from the acquisition of the norms associated with a standard model to a focus on learning linguistic features, cultural information and communication strategies that will facilitate communication. (194)

Brown suggests ways to reconceptualise ELT when using the sociolinguistic features of the international varieties of English. For her,

> attention to a world Englishes perspective in choice of methodology and curriculum design will result in an ecologically sound approach to language education, one that is attentive to the role that shifts in context bring to language education. (689)

Conveniently, Baumgardner and Friedrich provide some examples to include Outer- and Expanding-Circle Englishes into ELT classrooms. Friedrich suggests that "by bringing awareness to the different varieties of English that the students will encounter and by teaching them to view these varieties as legitimate expressions of a language in constant change and spread, a World Englishes approach can greatly facilitate learning" (444). Moroever, Baumgardner proposes that

"whether in Inner-, Outer-, or Expanding-Circle classrooms students' sensitivity towards the unprecedented spread and diversification of the English language should be one of all teachers' goals" (668).

Finally, Medgyes reinforces this idea by saying that teaching English as an international language is basically "teaching a large stock of native and non-native varieties of English" (185).

In these circumstances, it is vital that teachers integrate activities emphasizing the linguistic and cultural diversity of the English language. However, in a study to identify the representations of native and non-native varieties and cultures on Portuguese basic and secondary textbooks, Guerra concluded that few of the materials analysed made references to or used non-native varieties or native varieties other than British and American English. Obviously, there is a gap between the proposed theory of World Englishes in the EFL classroom and the actual use of representations of international English varieties.

In view of this, one effective way of filling this gap can be through the use of translation activities.

4. Using Native and Non-native Varieties in Translation Activities in the EFL Class

Due to the type of distinctive features of native and non-native varieties of English, employing translation activities in the language class might prove to be a motivating and enriching learning experience.

The following are some examples of activities which can be done in advanced EFL classes. As students are usually familiar with a standard variety of English (American or British), the translation activity should take two main steps: first, "translating" the original text into Standard English, and second, translating the Standard English text into the students' native language.

The translation activities are structured as follows. Each variety of English is introduced separately (African American Vernacular English, Singapore English, Indian English and Australian English) and some of the main features of the variety (phonological, morphological, syntactic, vocabulary) are identified. Then, excerpts of the materials used in class (texts in the

variety and the Standard English counterpart) for analysis and translation are provided. The texts in Singapore English, Indian English and Australian English were created to be used in the activity attempting to include several features of the variety.

4.1. African-American Vernacular English (AAVE)

According to Sidnell, although some features of AAVE might be unique to this variety, it shares many similarities in its structure with other standard and nonstandard varieties of English spoken in the US and the Caribbean. Some linguists believe that AAVE developed out of the contact between speakers of West African languages and speakers of vernacular English varieties. For these scholars, West Africans learned English on plantations in the southern states of the US leading to the development of a pidgin which was later expanded through a process of creolization. For this reason, many linguists would argue that, because AAVE shares so many characteristics with Creole dialects all over the world, AAVE itself is a creole.

The pronunciation of AAVE is in many ways similar to varieties of Southern American English. However, there is little regional variation among speakers of AAVE. Some research has tried to suggest that AAVE has grammatical structures in common with West African languages but this is disputed. Similar to what happens with any language variety, topic, status, age, and setting influence the usage of AAVE. Remarkably, there are many literary uses of AAVE, especially in African-American literature.

Phonology:

- Word-initially, /θ/ is normally as in SE (so *thin* is [θɪn])

- Word-initially, /ð/ is [d] (so *this* is [dɪs])

- Word-medially and -finally, /θ/ is realized as either [f] or [t] (so [mʌmf] or [mʌnt] for *month*); /ð/ as either [v] or [d] (so [smuv] for *smooth*)

- /r/ is usually dropped if not followed by a vowel. Intervocalic /r/ may also be dropped, e.g. SE *story* ([stɔri]) can be pronounced [stɔ.i]

- /l/ is often deleted in patterns similar to that of /r/ and can make homophones of *toll* and *toe*, *fault* and *fought*, and *tool* and *too*.

- /ɪŋ/ ⟶ /æŋ/, *thang*

- Front-shifting of stress, *PO-lice*

- Consonant cluster simplification, one *tes*, two *tesses*

- Unstressed initial syllables are deleted, *'bout it*

- Word-final devoicing of /b/, /d/, and /g/, *cub* sounds like *cup*

- Use of metathesized forms like *aks* for *ask* or *graps* for *grasp*

Morphology and Syntax:

- Third person -<s> deletion: *I walk, he walk*

- Genitive ending: *Rosemary house*

- Nominative forms of some pronouns rather than genitive ones: *They house*

- Deletion of copula: *He here* (He's here)*, She a teacher* (She's a teacher)

- No copula deletion where it cannot be reduced in SE: *I ain' the one did it, he is. Allah is God. He ain' home, is he?*

- *He workin'.* (He is working [right now])

- *He be workin'*. (He works frequently or habitually. Better illustrated with "He be workin' Tuesdays.")

- *He stay workin'*. (He is always working)

- 'Been': *She been married.* (She married a long time ago and still is married), *He been workin'.* (He has been working)

- Completive 'done': *He done worked.* (He has worked). Syntactically, "He worked" is valid, but "done" is used to emphasize the completed nature of the action.

- Future perfective *be done* (combination of future tense and completive aspect): *She be done graduated by June* (spoken in December of previous year)

- Use of *ain't* as a general negative indicator (SE *am not*, *isn't*, *aren't*, *haven't* and *hasn't*). Some speakers of AAVE also use *ain't* instead of *don't*, *doesn't*, or *didn't* (e.g. *I ain't know that*)

- Multiple negation: *I don't know nothing about no one no more.*

- *It* or *is* denotes the existence of something (SE "there is" or "there are"): *Is a doughnut in the cabinet.* (There's a doughnut in the cabinet), *It ain't no spoon.* (There isn't a spoon)

- Altered syntax in questions: *Why they ain't growing?* (Why aren't they growing?), *Who the hell she think she is?* (Who the hell does she think she is?)

Lexicon:

bogus 'fake/fraudulent', **hep, hip** 'well informed, up-to-date', **cat** 'a friend, a fellow', **cool** 'calm, controlled', **dig** 'to understand, appreciate, pay attention', **bad** 'really good', **bad-eye** 'nasty look', **big-eye** 'greedy', **bread** 'money', **crib** 'house', **mojo** 'personal magic', **wigga**, **wigger** 'white youth who identify with AA culture', **Mickey D's** 'McDonald's', **dig** 'to understand/appreciate', **gray** 'an adjective for whites' (gray dude), **The Man** 'white authority', **Miss Ann** 'white woman', **bogus** 'fake/fraudulent', **hep, hip** 'well informed, up-to-date', **cool** 'calm, controlled', **bad** 'really good'.

AAVE text:

The text used in the translation activity was taken from Alice Walker's *The Color Purple* due to the diversity of AAVE features (phonological, morphological and syntactic) it presents.

Dear God,

He beat me today cause he say I winked at a boy in church. I may have got somethin in my eye but I didn't wink. I don't even look at mens. That's the truth. I look at women, tho, cause I'm not scared of them. Maybe cause my mama cuss me you think I kept mad at her. But I ain't. I felt sorry for mama. Trying to believe his story kilt her.

Sometime he still be looking at Nettie, but I always git in his light. Now I tell her to marry Mr. _____. I don't tell her why.

I say Marry him, Nettie, an try to have one good year out your life. After that, I know she be big.

But me, never again. A girl at church say you git big if you bleed every month. I don't bleed no more.

Standard English:

The following table identifies the features of AAVE which appear in the text and provides their standard English form and an explanation of these differences.

AAVE	SE	Explanation
cause	because	deletion of unstressed initial syllable
say	said	tense (Pres ⟶ Past)
somethin	something	/ŋ/ deletion ⟶ /n/
mens	men	use of regular plural
tho	though	simplified speling
cuss	(used to) cuss (at)	(old fash.) to swear at somebody; Pres ⟶ 'used to'
think	would think	pres ⟶ modal 'would'
ain't	didn't	use of *ain't*
kilt	killed	/d/ ⟶ /t/ (devoicing)
sometime	sometimes	-s deletion
(still) be looking	is usually looking/ usually looks	'be' + -ing
git	get	/ɪ/ vs /ɛ/
an	and	final -d deletion (/d/)
out your life	… of …	preposition
be 'big'	will be	future tense (will be)
big	pregnant	lexicon
don't … no	don't … any	double negative

Table 1. *Comparison of AAVE features and Standard English*

4.2. Singapore English (Singlish)

According to Brown, Singapore English can be seen as a legitimate variety of English in its own right, as it helps to convey Singaporean identity and can be taken as a sign of a growing national self-confidence. And for that reason, it should be analysed as a unique variety, not as an imitation

of Standard English. Such attitude to Singapore English can be found in the Introduction to *The Coxford Singlish Dictionary*, whose unknown author states that Singlish is a "mish-mash of various languages and dialects . . . not merely badly spoken English" (viii) which should be celebrated as a cultural phenomenon unique to Singapore. At the other end of the scale, Shelley remarks that "Singlish is an amorphous body of deviations from Standard English; and Standard English is something equally wooly, vague and lumpy" (5-6) or, as A. Brown puts it, "deficient in certain aspects and in need of correction" (vi). However, Gartshore provides a clear definition of Singlish, underlining its basic linguistic features:

> Singlish is the informal, spoken Asian English indigenous to Singapore – a language academics call 'Singapore Colloquial English'. Most Singaporeans are multi-lingual and speak Singlish as a second language to Chinese and Malay dialects, Tamil, or Standard English. Standard English grammar rarely applies to Singlish. Grammatical endings, tenses, plurals, and the definite article are ignored for the most part, allowing for a more direct rhythmic discourse. Particles and sentence endings feature in Singlish and can be heard in most conversations. 'Okay lah' is one example in which 'lah' lends emphasis and conveys a sense of agreement. (8)

Basically, the features that distinguish Singapore English from British or American Standard English can be found in phonology and in the use of some culturally-based lexical items. It is also important to note that Singapore English has been influenced by many varieties throughout its history: British English, for historical reasons, and more recently, Indian English and American English.

Phonology:
- word stress: *Ar-RAB-ic, cal-LEN-dar, in-DUS-try, pur-CHASE, PHO-tography*
- lack of distinction between initial /p/ and /d/, /t/ and /d/, /k/ and /g/, /t/ and / θ/, /d/ and /ð/: *bark (=park), dan (=tan), gum (=come), though (=taught), they (=day)*

- omission of final sound /l/ and /n/: *drawl (=draw), brown (=brow)*

- simplification of final consonant clusters (loss of /t/, /d/, /s/ and /z/): *based on (=base on), difference (=different)*

- mid-word voicing: *December*, /z/

- lack of distinction between long and short vowels: *bead (=bid), pool (=pull)*

Morphology and Syntax:

- 'got' to express location (as in there is/are): *Singapore got two universities*

- elimination of uncountable nouns: *furnitures, clothings, equipments, informations*

- no distinction between 'very' and 'too'

- 'yes' and 'no' are often not given as the reply to yes-no questions: *Can you swim? Can; Do you like hot food? I like; Can I take this road? Cannot*

- deletion of 'it': *Don't miss! You can't resist! You'll regret if you miss it.*

- omission of preposition: *to participate* [in] *a game, to pick someone* [up] *in a car*

- use of a different preposition: *in campus, to take out shoes, to hand up homework*

- inclusion of a preposition: *to consider about something, to tolerate with someone*

- use of 'just' at the end of a sentence: *She was here just; When did the train leave? Just.*

Lexicon:

cager 'basketball player', **air-con** 'air conditining', **to chop** 'to stamp', **to call** 'to ask, to tell', **blanco** 'correction fluid', **wash (photos)** 'develop and print', **to use** 'to wear', **stylo(-mylo)** 'stylish', **stoned** 'tired', **to stay** 'to live', **medical hall** 'chemist's, drugstore'

Singlish text:

1. – Tom, you need to relak lah.

   – Very stoned lah. Now must koon. I'll just open my shoes. Can you diam? Off the radio please.

2.   That char bor is very the chio but she always like to action and she's so hiao.

3.   – Borrow me twenty sing? I want to buy new stylo mylo shoes.

       – Can.

4.   – You join me for makan?

       – I no mood. I've just taken my dinner. You go head, later I catch up.

       – Ok, but can you send me to the bus stop?

5.   – I cannot tahan my job. I'll open a medical hall.

       – What talking you?

6.   – Nice to meet you. What you doing now?

       – I'm a hawker.

       – And where do you stay?

       – In a maisonette.

Standard English text:

1.   – Tom, you need to RELAX.

       – I'M very TIRED. I must SLEEP NOW. I'll just TAKE OFF my shoes. Can you BE
       QUIET? TURN OFF the radio please.

2.   That FEMALE/GIRL is very PRETTY but she always likes to SHOW OFF and she's
       so VAIN.

3.   – CAN YOU LEND me twenty (SINGAPORE) DOLLARS? I want to buy new
       STYLISH shoes.

       – YES, NO PROBLEM.

4.   – WOULD YOU LIKE TO JOIN ME TO EAT?

       – I AM NOT IN THE MOOD. I've just EATEN my dinner. You GO AHEAD, I WILL
       catch YOU up LATER.

       – Ok, but can you DROP ME OFF BY the bus stop?

5.   – I cannot TOLERATE my job. I'll open a CHEMIST'S/DRUGSTORE.

       – What ARE YOU TALKING ABOUT?

6.   – Nice to meet you. What DO YOU DO FOR A LIVING?

– I'm a HAWKER (FOOD SELLER AT 'HAWKER CENTRES' = OUTDOOR FOOD

COURTS)

– And where do you LIVE?

– In a MAISONETTE (a flat on two floors).

4.3. Indian English

According to Mehrotra, for historical reasons users of English in India have looked upon British

English as a model and a point of reference. However, due to language contact and the distinct

socio-cultural reality of the South Asian subcontinent, the English used in India has acquired

features distinct from the native British variety.

Jacob has also pointed out that when English interacted with a number of regional

languages, a new variety of English, often labelled Indian English, has developed. He adds that

as a non-native language for most Indian bilinguals, Indian English "is bound to have certain

characteristics of its own in grammar, pronunciation and vocabulary" (15).

Phonology:

- diphthongs /eɪ/ and /ou/ tend to to be monophthongal /eː/ and /oː/

- (southern India) word-initial front vowels tend to receive a preceding /j/ and back vowels

a preceding /w/: *eight* /jeːt/, *own* /woːn/

- (northern India) word-initial /sk/, /st/ or /sp/ tend to receive a preceding /i/: *speak* /ispiːk/

- non-rhotic for most educated speakers

- /r/ tends to be flap [ɾ]

- /p/, /t/ and /k/ tend to be unaspirated

- in some varieties, /v/ and /w/ are not distinguished; similarly /p/ and /f/; /t/ and /θ/; /d/

and /ð/

Morphology and Syntax:

- pluralization of mass nouns: *fruits*, *furnitures*

- use of nouns alone: *clothes* (*I have bought two clothes today* = items of clothing), *toasts* (*I'd like two toasts, please* = slices of toast)

- compound formation: *chalk-piece* (piece of chalk), *key-bunch* (bunch of keys), *schoolgoer* (one who goes to school)

- nominal forms of adjectives: *colour pencils* (coloured), *schedule flight* (scheduled)

- use of preposition: *to fear of*, *to return back*, *to pay attention on*, *to get down (from a vehicle)*

- use of *itself* and *only* to emphasize time or place: *Can I meet with you tomorrow itself?*, *We arrived today only*

- use of adverbial *there* for 'dummy' *there*: *I'm sure an explanation is there*

- use of present tense with durational phrases: *I am here since two o'clock*

- use of progressive aspect with habitual action: *I am doing it often*

- use of progressive aspect with stative verbs: *Are you wanting anything?*

- use of perfect aspect instead of simple past: *We have already finished it last week*

- direct question with no subject-verb inversion: *What this is made of?*

- indirect question with inversion: *I asked him where does he work*

- use of *isn't it* as a universal undifferentiated tag question: *They said they will be here, isn't it?*

Lexicon:

biodata 'curriculum vitae', **co-brother** 'wife's sister's husband', **cousin-sister** 'female cousin', **to half-fry** 'to fry (an egg) on one side', **hotel** 'restaurant, cafe' (not necessarily with lodgings), **stepney** 'a spare wheel', **tiffin** 'lunch', **the needful** 'whatever is necessary', **batch-mate** 'a person in the same class at school or college', **head-bath** 'a complete bath'.

Indian English text:

Sanjay is awake since five o'clock. He was too excited to stay in bed that morning. He couldn't even have his bed-tea. He took a head-bath in five minutes so he could spend more time choosing what to wear from his three favourite clothes. Sanjay has already prepared his school materials two days ago: colour pencils, rule paper, textbooks, among other items. He is a dedicated student and is having excellent grades in school. So it was no surprise when he got freeship this year. Sanjay is no back-bencher and will do the needful to be doctor.

Sanjay went to the convent in a tempo and as he got down he ran to meet his cousin-sister who was going to be his batch-mate this year. He didn't want to waste time chatting so he rushed into his classroom. It was a spacious room with lots of furnitures. The teacher, Ms. Prasad, stood by the door to greet each and every student.

"Are you wanting anything, Sanjay?" asked the teacher.

"No, thanks, Ms. Prasad", replied Sanjay.

Before sitting, Sanjay wiped off the blackboard as he wanted to make a good impression on Ms. Prasad.

Although Sanjay paid attention on the teacher, he was having difficulties to solve a Maths problem.

"I'm sure an explanation is there!"

Ms. Prasad approached Sanjay and said:

"You could finish it tomorrow, isn't it?"

"No, Ms. Prasad, I'll keep trying."

After some minutes struggling with the problem, Sanjay seemed to have arrived at a solution:

"Now only I have understood the problem!"

Standard English text:

Sanjay HAS BEEN awake since five o'clock. He was too excited to stay in bed that morning. He couldn't even have his EARLY MORNING CUP OF TEA. He took a BATH in five minutes so he could spend more time choosing what to wear from his THREE favourite ITEMS OF CLOTHING. Sanjay already PREPARED his school materials two days ago: COLOURED pencils, RULED paper, textbooks, among other items. He is a dedicated student and IS HAS excellent grades in school. So it was no surprise when he got FREESHIP (EXEMPTION FROM THE PAYMENT OF SCHOOL FEE) this year. Sanjay is NOT SOMEONE UNINTERESTED IN HIS STUDIES and will DO WHATEVER IS NECESSARY to be A doctor.

Sanjay went to the CONVENT (ENGLISH-MEDIUM SCHOOL, ESP. ONE RUN BY A CHRISTIAN MISSION) in a TEMPO (A THREE-WHEELED MOTORIZED VEHICLE) and as he GOT OFF he ran to meet his COUSIN who was going to be his CLASSMATE this year. He didn't want to waste time chatting so he rushed into his classroom. It was a spacious room with lots of PIECES OF FURNITURE. The teacher, Ms. Prasad, stood by the door to greet each and every student.

"DO YOU WANT anything, Sanjay?" asked the teacher.

"No, thanks, Ms. Prasad", replied Sanjay.

Before sitting, Sanjay CLEANED the blackboard as he wanted to make a good impression on Ms. Prasad.

Although Sanjay paid attention TO the teacher, he was having difficulties to solve a Maths problem.

"I'm sure THERE IS AN EXPLANATION!"

Ms. Prasad approached Sanjay and said:

"You could finish it tomorrow, COULDN'T YOU?"

"No, Ms. Prasad, I'll keep trying."

After some minutes struggling with the problem, Sanjay seemed to have arrived at a solution:

"JUST NOW I have understood the problem!"

4.4. Australian English

In a brief analysis of the features of Australian English, Todd and Hancock (1990) state that this variety is less differentiated than the varieties of any other English-speaking society of comparable size "because of the small population, the relative classlessness of Australian society and the homogeneity of the original settlers" (65) although one can clearly identify regional and class variation in Australia.

Australian English began to diverge from British English after the Colony of New South Wales was founded in 1788. Later on, in the 1850s, the wave of immigration as a result of the first Australian gold rush also had a significant influence on Australian English. Since then, Australian English has borrowed increasingly from non-British sources.

Phonology:

- a non-rhotic variety with a reasonably standard consonant inventory

- the long monophthong /i/ tends to be diphthongised: *beat* /bəɪt/

- some diphthongs tend to be monophthongized: *here*, *there*, *sure*

- the diphthong /eɪ/ is lowered and realized by many as in /aɪ/: *tail* (= RP, tile)

Morphology, Syntax and Spelling:

- the exposure to the different spellings of British and American English leads to a certain amount of spelling variation such as *organise/organize*. British spelling is generally preferred, although some words are usually written in the American form, such as *program* and *jail* rather than *programme* and *gaol* (although commonly one could be 'jailed' in a 'gaol')

- both _ise and _ize are accepted, as in British English, but '_ise' is the preferred form in Australian English by a ratio of about 3:1 according to the Macquarie's Australian Corpus of English

- tendency to excise the 'u'

- use of *but* at the end of a sentence as a modifier (equivalent to *though*): *I didn't do it but*

- use of *thanks* in requests: *Can I have a cup of tea, thanks?*

- use of feminine pronoun *she* to refer to inanimate nouns and in impersonal constructions: *She'll be all right* (Everything will be all right), *She's a stinker today* (The weather is excessively hot today)

Lexicon:

- use of diminutives:

* adding *-o* or *-ie* to the ends of abbreviated words. They can be used to indicate familiarity, although in many speech communities the diminutive form is more common than the original word or phrase: *arvo* (afternoon), *docco* (documentary), *servo* (service station, known in other countries as a "petrol station" or "gas station"), *bottle-o* (bottle-shop or liquor store), *rego* (still pronounced with a /ʤ/) (annual motor vehicle registration), *traino* (train station), *compo* (compensation), *lebo* (Lebanese), *lezzo* (lesbian) or *ambo* (ambulance officer). The same applies to names: *Jono* (John), *Freo* (Fremantle), and *The Salvos* (The Salvation Army); *barbie* (barbecue), *bikkie* (biscuit), *bikie* (member of a motorcycle club), *brekkie* (breakfast), *blowie* (blowfly or occasionally meaning oral sex), *brickie* (brick layer), *mozzie* (mosquito), *pollie* (politician), *c*hippie (carpenter) and *sparkie* (electrician).

* occasionally, a *-za* diminutive is used, usually for personal names where the first of multiple syllables ends in an "r": *Kazza* (Karen), *Jezza* (Jeremy).

* first syllable plus "-s": *turps,* turpentine (usually referring to drinking alcohol, e.g. "a night on the turps") or Ian Turpie; Gabs, pet form of Gabrielle.

a cuppa 'a cup of tea', **take a load off** 'sit down', **in the hols** 'during the holidays', **fair dinkum** 'genuine', **yapping** 'talking a lot', **take one for Ron** 'take another one for later on', **good on ya** 'good for you', **Pommy** 'an English person (unflattering), **Kiwi** 'a New Zealander', **footpath** 'pavement', **picture theatre** 'cinema'.

Australian English text:

– G'day, cobber! What are you doing this arvo?

– Not sure. I might meet Sue for a cuppa and some bikkies.

– We're meeting at Bob's for a barbie. And you know there's no barbie without some tinnies.

– Is Amanda going to be there?

– Yeah, why?

– She's always yapping and she's so ugly she looks like a bitzer, fair dinkum, mate!

– Don't worry, there'll be plenty of people to talk to, a bunch of Pommies, Kiwis, Aussies…

– All right, then. I better leave now before the travel agency closes. I need to pick up my ticket to Paris. I'm visiting my sister in the hols.

– Good on ya, mate!

– Take care.

Standard English:

– Hello, friend! What are you doing this afternoon?

– Not sure. I might meet Sue for a cup of tea and some biscuits.

– We're meeting at Bob's for a barbecue. And you know there's no barbecue without some cans of beer.

– Is Amanda going to be there?

– Yeah, why?

– She's always talking a lot and she's so ugly she looks like a mongrel dog, really, man!

– Don't worry, there'll be plenty of people to talk to, a bunch of English people, New Zealanders, Australians…

– All right, then. I better leave now before the travel agency closes. I need to pick up my ticket to Paris. I'm visiting my sister in the holidays.

– Good for you, man!

– Take care.

5. Conclusions

In a study conducted with Portuguese EFL university students and teachers, Guerra found out that there is an overall positive attitude towards learning and teaching about native and non-native varieties of English although they hold more positive attitudes toward the British and American Standard varieties. However, few subjects reported being familiar with other native and non-native varieties of English which might be explained by the little or no contact they had with them in their English language education.

All in all, the vast majority of subjects believed it is very important to learn about international features of English, and a high percentage of teachers reacted positively to incorporating non-native varieties in class.

Fundamentally, it is up to the EFL teachers to establish and fulfil these objectives. Dealing with native and non-native varieties has proven to be an essential tool to develop the students' awareness of the interdependent relationship between language and culture and to illustrate the linguistic, cultural and intercultural diversity of English. Using translation activities seem to be an effective approach to achieve this purpose.

Works Cited

The Coxford Singlish Dictionary. Singapore: Angsana Books, 2002.

Baumgardner, Robert J. "Utilizing Pakistani Newspaper English to Teach Grammar." *World Englishes* 6.3 (1987): 241-53.

Baxter, James. "How Should I Speak English? American-ly, Japanese-ly, or internationally?" *Teaching English Pronunciation: a Book of Readings*. Ed. Adam Brown. London: Routledge, 1991. 53-71.

Brown, Adam. *Singapore English in a Nutshell*. Singapore: Federal Publications, 1999.

Brown, Kimberley. "Models, Methods and Curriculum for ELT Preparation." *The Handbook of World Englishes*. Ed. Braj Kachru, Yamuna Kachru, and Cecil L. Nelson. Oxford: Wiley-Blackwell, 2009. 680-93.

Campbell, Donald, Peansiri Ekniyom, Anjum Haque, and Larry Smith. "English in International Settings: Problems and Their Causes." *Readings in English as an International Language*. Ed. Larry E. Smith. Oxford: Pergamon Press, 1983. 35-47.

Crystal, David. *English as a Global Language*. Cambridge: CUP, 1997.

Friedrich, Patricia. "Teaching World Englishes in Two South American Countries." *World Englishes* 21.3 (2002): 441-44.

Gartshore, William. *An Essential Guide to Singlish*. Singapore: Gartbooks, 2003.

Gnutzmann, Claus. "English as a Global Language: Perspectives for English Language Teaching and for Teacher Education in Germany." *Teaching and Learning English as a Global Language: Native and Non-native Perspectives*. Ed. Claus Gnutzmann. Tübingen: Stauffenburg-Verlag, 1999. 157-69.

Guerra, Luis. *Teaching and Learning English as an International Language in Portugal: Policy, Practice and Perceptions*. Saarbrücken: VDM Verlag, 2009.

Jacob, George. "Indian English: Certain Lexical and Grammatical Variations." *Modern English Teaching* 7.4 (1998): 15-19.

Jenkins, Jennifer. *World Englishes: A Resource Book for Students*. London: Routledge, 2003.

Kachru, Braj. "Standards, Codification and Sociolinguistic Realism: The English Language in the Outer Circle." *English in the World: Teaching and Learning the Language and Literatures*. Ed. Randolph Quirk and Henry Widdowson. Cambridge: Cambridge University Press, 1985. 11-30.

Kirkpatrick, Andy. *World Englishes: Implications for International Communication and English Language Teaching*. Cambridge: CUP, 2007.

McArthur, Tom. *The English Languages*. Cambridge: CUP, 1998.

Medgyes, Peter. "Language Training: A Neglected Area in Teacher Education." *Non-native Educators in English Language Teaching*. Ed. George Braine. Mahwah: Lawrence Erlbaum Associates, 1999. 159-76.

Mehrotra, Raja Ram. "A British Response to Some Indian English Usages." *English Today* 19.3 (2003): 19-25.

Melchers, Gunner, and Phillip Shaw. *World Englishes*. London: Arnold Publishers, 2003.

Mesthrie, Rajend, and Rakesh M. Bhatt. *World Englishes: The Study of New Linguistic Varieties*. Cambridge: CUP, 2008.

Modiano, Marko. "Ideology and the ELT Practitioner." *International Journal of Applied Linguistics* 11.2 (2001): 159-73.

---. "Linguistic Imperialism, Cultural Integrity, and EIL." *ELT Journal* 55.4 (2001): 339-46.

Mufwene, Salikoko, John R. Rickford, Guy Bailey, and John Baugh. *African-American English: Structure, History, and Use*. New York: Routledge, 1998.

Platt, John Talbot, Heidi Weber, and Ho Mian Lianet. *The New Englishes*. London: Routledge and Kegan Paul, 1984.

Shelley, Rex. *Sounds and Sins of Singlish*. Singapore: Times Books International, 2000.

Sidnell, Jack. "African American Vernacular English." *Language Varieties*. Web. n. p., 2000. Web. 28 March 2010 <http://www.une.edu.au/langnet/definitions/aave.html>.

Todd, Loreto, and Ian Hancock. *International English Usage*. London: Routledge, 1990.

Trifonovitch, Gregory. "English as an International Language: An Attitudinal Approach." *English for Cross-Cultural Communication*. Ed. Larry E. Smith. Hong Kong: MacMillan Press, 1981. 211-24.

Trudgill, Peter, and Jean Hannah. *International English: A Guide to Varieties of Standard English*. London: Arnold, 2002.

Walker, Alice. *The Color Purple*. New York: Mariner Books, 2006.

Spaced Learning: Making Space for Neuroscience in the Classroom

Ana Catarina VAZ
| University of Porto

Abstract | The paper offers an insight into the scientific research on the use of neuroscience in today's educational systems and classrooms. It also presents Spaced Learning as an example of practical teaching techniques which apply neuroscience in the classroom. A brief account of the potential relevance of neuroscience in education looks at how neuroscience and education are exploring the potential for scalable solutions, such as Teensleep and Spaced Learning. A closer examination of two different experiments in England and Portugal demonstrates how Spaced Learning has been successfully deployed and raised achievement. Finally, it considers making *space* for neuroscience in the classroom, and aims to explain how the classroom should be prepared for a neuroscientific approach.

Cognitive neuroscience – with its concern about perception, action, memory, language and selective attention – will increasingly come to represent the central focus of all neurosciences in the 21st century.

Eric Kandel

Slowly but surely, neuroscience has lately imposed itself in the most significant fields of study, namely in educational research studies and showing its importance in terms of more accurate and effective results regarding students' performance. For the first time ever, The Wellcome Trust and the Education Endowment Foundation will trial neuroscience discoveries in real schools marking thus the beginning of a new era for both education and neuroscience (cf. Philippou). However, preparing the formal classroom for neuroscience should be seen not only as a major investment for a successful report and data collection, but also as a crucial element for a rewarding outcome.

I. Cognitive Neuroscience

It is not by chance that the field of study which focuses on perception, action, memory, language and selective attention, known as neuroscience (neuroscience is the study of the brain and its processes, and these can impact directly on learning) has caught the attention of researchers in the last few years. In an ever growing, technologically dependent society, it is imperative that both educators and educational systems move forward towards an educational system that enhances and promotes not only interdependent and multidisciplinary thinking but also the accurate use of contents and selective attention. With all the resources and memory storage devices available nowadays, it does not seem productive, nor effective, to expect from students few basic facts learnt by heart.

> We spend millions of dollars and years of children's lifetimes teaching them to do sums as well as a $5 calculator . . . , children spend thousands of hours learning basic exact calculations, making those children the most expensive and least reliable pocket calculators on earth. (Kelley, *Making Minds* 146)

This round assertion challenges, firstly, the categorisation of all learning contents. Should we expect learners to operate as readily and skilfully as any electronic device, or should we, on the other hand, train their selective attention and help to fully develop their perception, action, memory and language skills? If we take into account the fact that Kelley, one of the leading educators in the UK, has defined learning as "the development of adaptive skills" (157), then the role of educational systems should be to select more accurately the most pertinent contents to be taught, namely those with the most explanatory potential and practical procedural skills. This is also where selective attention and memory should soon start to converge; inciting innovative neural systems in the brain and fully understand their contribution to a renewed and more powerful educational system.

II. Overview

Even though several aspects of formal learning and understanding may still not be fully understood and explained by neuroscience, many educators have experimented to improve learning. Some of these educators, though little known, made very perceptive analyses of the learning process. For example, Aiken, an American pedagogue, first published the work *Methods of Mind-Training, Concentrated Attention and Memory* in 1895, in which she explained that voluntary attention was an acquired skill or mental behaviour usually attained through habit and, in a more permanent state, aided by external support (Aiken 23). In spite of having her findings only empirically proven, she suggested three essential stages for attaining voluntary attention: 1) positive emotions and rewards; 2) emulation, ambition, practice (artificial attention); and 3) the habit of organizing past contents for more efficient and lasting learning (80-81).

In her work, Aiken also developed the idea that attention should be trained every day for at least 20 minutes. This would then result in better comprehension, greater satisfaction and less study hours for students. However, accurate recall, which should lead to an objective reasoning, should not be mistaken with the concept of learning by heart, but rather understood as accurately remembering and using valuable data. A good memory should always enhance attention, and

attention should always be sustained through (un)pleasant or more complex experiences (80). Most of her theories will be proven right almost two centuries later, as we will see later on with spaced learning.

Moreover, the importance of neuroscience (time patterns in long-term memory) and psychology (retrieval practice) shown empirically by Aiken strongly suggest an important use in education of evidence-based time patterns from both research traditions:

> Although the spacing effect in retrieval has been demonstrated in many subjects and educational contexts to be effective, it has rarely been systematically implemented in education (Dempster, 1988; Seabrook et al., 2005). Despite recent careful analysis of the temporal patterns demonstrating effective recall of word pairs and other tasks (Cepeda et al., 2006; Pavlik and Anderson, 2008; Cepeda et al., 2009) and despite specific programmes based on the approach (Carpenter et al., 2009, 2012; Sobel et al., 2011), this remains the case. (Kelley and Whatson 2)

More recently, neuroscience will be put to test in real contexts throughout schools within the UK. The need of a more challenging teaching environment with more satisfying results has led some of the most important entities in education to investigate a variety of ways neuroscience might improve teaching and learning. There are six projects in total, which will benefit from grants of almost £4 million. These series of randomized controlled trials will be conducted by the Wellcome Trust and the Education Endowment Foundation across England and they will test thousands of pupils after the identified a need for stronger evidence about how neuroscience relates to learning in order to support teachers and schools keen to make use of neuroscientific findings in the classroom. In some of the trials non-invasive bio-telemetric devices and brain imaging will be used to provide additional physiological data (cf. Philippou).

The six chosen projects are *Teensleep*, which will assess the impact of later school start times on teenagers' educational achievement; *Learning counterintuitive concepts*, which will test the benefit of suspending pre-existing beliefs when solving mathematical or scientific questions; *Fit to study*, which will look at the effect of medium to high cardiovascular activity on academic

attainment; *Engaging the brain's reward system*, which will examine the effect of uncertain reward on attainment); *GraphoGame Rime*, which will investigate the benefits of phonological awareness through "rhyme analogy", and finally, and perhaps the most complex and challenging, *Spaced learning*, which will check the effectiveness of repetition and spaced learning, a method that delivers a unit of work three times interspersed with different activities.

III. Spaced Learning

Spaced learning was first put into practice in a classroom in 2005 in, England. The project was initially led by Kelley, and the teaching technique developed by Bradley, a Science teacher. The method behind Spaced Learning is based on Fields' findings, published in the article "Making Memories Stick" in the *Scientific American Magazine* three years earlier, in 2005. Fields, an American researcher at *The Douglas Fields Lab at the National Institutes of Health*, US, found out that any content needed repetition and intervals of inactivity between them in order to be successfully stored as long-term memories:

> Both long- and short-term memories arise from the connections between neurons, at points of contact called synapses . . . . When a short-term memory is created, stimulation of the synapse is enough to temporarily "strengthen", or sensitize, it to subsequent signals. For a long-term memory the synapse strengthening becomes permanent. . . . Long-term memory often requires some kind of repetition . . . . If stimulation is applied repeatedly – three times in the reported experiments – the synapse becomes strengthened permanently. . . . each stimulus must be spaced by sufficient intervals of inactivity (10 minutes in our experiments). (76)

Spaced Learning, based on the findings by Fields, is a teaching method that creates neural pathways at the start of a unit of work and revisits them throughout the same lesson or session. "When we hear, see or do something once, it can be stored in our short-term memory. If we hear, see, or do it repeatedly, it can enter our long-term memory" (Bradley and Patton 3-4).

In short, any spaced learning lesson follows the structure of three main inputs separated by two breaks of a different character. In the first input, the teacher presents a variety of key selected contents in no more than 20 minutes. This should be done at a steady and comprehensive speed to stimulate the synaptic pathways. After the first input, students have a break of 10 minutes at least to perform a different task. This task can be of a playful or entertaining tone, such as dribbling, origami folding or singing, for example. The breaks allow the synaptic ways to rest before being strengthened again during the second input. In the second input, students are expected to revisit the contents presented during the first input and to execute basic tasks (e.g. recalling information, filling in the gaps, solving simple exercises). After the second input comes another pause in order to let the synaptic ways rest once more before being revisited again in the third input. The last and final input, or output, encourages the students to practice fully the contents learnt during the session (e.g. oral presentation of contents, solving more complex tasks) and detect eventual failures in the learning process.

In the first spaced learning experience inside a real classroom, which was led in England, three different groups were used. The total population of 55 students (between 15 and 16 years old) was divided by: experimental group, with whom spaced learning was used for 24 minutes; control group, which learnt the contents for 60 minutes without spaced learning; and a mixed group, with whom both approaches were used for also 60 minutes (24 minutes as in the experimental group; 36 minutes conventional social-based methods) (Kelley, *Making Minds* 154).

In spite of the fact that the final results (Academic Potencial Yellis)[1] were very similar (both experimental and mixed group reached a 4.9 grade and the control group 5.0), the learning efficiency (learning per 60 minutes as measured by the test)[2] was over twice as much in the experimental group when comparing with both the control and mixed groups (173 in the experimental group and 61.1 and 75.2 in control and mixed group, correspondingly), which may represent a substantial time saving inside the classroom, especially if we take into account the usefulness of the breaks for developing other skills (Vaz, *O Spaced Learning Enquanto Abordagem Pedagógica* 80-2, 111-12).

Ermesinde experiment

On a later experiment in Ermesinde, Portugal, with secondary school students, spaced learning was used to teach modern foreign languages.

The first lesson was taught to a 10th form group of nine students (German as a foreign language: A1 level) in a 90-minute session. The theme selected was "Wohnen" (inhabiting), following the national syllabus (cf. Lapa, Mota, and Vilela). Students were already familiar with the different rooms and house types from previous lessons, but all furniture related vocabulary was unknown (cf. Vaz, "Spaced Learning").

At the end of the lesson the experimental group was able to write an average of 54.2 correct words (cf. Table 1). The highest number was 76 words and the lowest 40. Four weeks later the same group of students wrote an average of 38.2 correct words, only 16 less than a month before.

The same final task was conducted in a control group, made of sixteen secondary school students learning German at a higher level (A2), under the exact same conditions (e.g. time limit 20') but without any previous spaced learning lesson. The average number of correct words written went down to 21.6 correct words with values ranging from 10 to 44.

Average number of correct words written in 20'		
Experimental group (A1 level)	**54.2** (min 40 – max 76) → *at the end of the SL lesson* **38.2** (min 25 – max 66) → *four weeks later*	
Control group (A2 level)	**21.6** (min 10 – max 44)	

Table 1. *Average number of correct words written in 20'* (Vaz, "Spaced Learning in the MFL Classroom" 41)

What is more, two weeks after the original spaced learning lesson, students from the experimental group were able to identify and recall all the requested words (pieces of furniture) at a written examination, achieving an outstanding 100% efficiency at the vocabulary task. In the writing task

of the same examination, students got an average of 89%, being highest and lowest scores 100% and 82%, correspondingly. No dictionaries were used at any stage.

There seems to be great potential lying behind spaced learning, especially if we take into account the evidence from students, teacher, observers and test scores. In order to grade some opinions from students and teachers involved, we used a scale of 1 to 5, where 1 stood for *I totally agree* and 5 for *I totally disagree*.

In the questionnaire answered individually by the students of the experimental group, 88.9% of them said that they were in total agreement (1) with the fact that spaced learning helped their learning process after only one lesson. 11.1% of the students agreed only (2) with this statement (Vaz, "Spaced Learning in the MFL Classroom" 74).

When rating the lesson as a whole, 100% of the students rated the spaced learning session as *more interesting and productive than a regular lesson*. The same percentage agreed totally (1) that they enjoyed learning the lesson contents through spaced learning.

Both observers were also in total agreement (1) that spaced learning was important for students' acquisition (namely the use of different colours and highlighted words) and that both pauses were interesting for students without interfering with the upcoming input stages.

At the end of the fifth spaced learning lesson, both the observer and students (100%) agreed totally (1) on the importance of the breaks for a more efficient learning (Vaz, *O Spaced Learning Enquanto Abordagem Pedagógica* 110).

In terms of assessing attention needs, students from the experimental group were also analysed through an ADHD (*attention deficit hyperactivity disorder*) table, specifically designed to detect and label any eventual lack of attention before, during and after spaced learning. During this particular 90-minute lesson, there were no visible cases of difficulties in *sustaining attention and following instructions* or *concluding tasks,* which had been detected in previous regular lessons. On the other hand, there were occurrences of *abrupt answers*, which means students formulated their answers before they heard the whole question(s) at six different times. At the end of the first lesson there were 12 ADHD occurrences and five sessions later only 3. This means

students were four times more engaged and focusing on their tasks after being prepared to do so (through spaced learning training) (Vaz, "O Spaced Learning" 34, 47).

It is also interesting to mention that before any spaced learning lesson 21 ADHD occurrences were registered and that in a regular lesson (after spaced learning intervention) only 10 were measured in the same experimental group. The effect of spaced learning could be seen even when it was not being used directly. In the control group, in the same period of time, and without any intervention the ADHD occurrences went from 37 to 60 (Vaz, "O Spaced Learning" 47).

Regarding test scores, students from the experimental group went from 74.5% in the first term (without spaced learning) to 77% overall in just five sessions. It is important to note though that these tests covered subjects taught without spaced learning as well, so it might account for other relevant purposes: spaced learning as a self-disciplinary tool (47). In one of the tests all students got 100% in the vocabulary exercise (first spaced learning lesson).

It seemed nonetheless relevant to explore other significant issues. In addition to the improvement in self-discipline, there was an improvement in working habits and motivation, for example. With spaced learning, students had the chance to receive a quantitative classification to check their acquisition at the end of each lesson, which was done by the students themselves, collaboratively (working in pairs or small groups) or by the teacher, and it was used as an assessment tool.

The Spaced Learning developments at Monkseaton proved to constantly increase the rapidity, accuracy and educational impact, that is, good high-stakes test scores. In addition to it, the Ermesinde experiment demonstrated that a series of spaced learning lessons could help less able students outperform more able students, which hardly happens in Education. Both of these suggest spaced learning has huge and unprecedented potential.

Taking this evidence into account, great efforts have been taken both by public and private organisms in the UK and worldwide to prepare and train educators for the challenges of recent findings about spaced learning and further projects. Osiris Educational, for example, one of the leading independent training providers for teachers in the UK has presented different spaced

learning training sessions on Science, Modern Foreign Languages and the method itself. In one of the courses, teachers were expected, among other things, to acquire useful knowledge on how to prepare themselves and their classroom for spaced learning; on how to make *space* for it.

IV. Spaced Learning: Making *Space* for Neuroscience in the Classroom

Before initiating any major project or change in the teaching approach or classroom dynamics, whether neuroscientific related or not, both educators and students should be fully aware of the implications that those choices may hold.

In the case of spaced learning, for example, teachers of a specific subject and target group, should list the reasons why this specific group may be in need of an intervention of this kind. Therefore, one of the first questions they should ask is *why* the students might need spaced learning. For example, it has been proven that this method enhances low and/or heterogeneous results within a group. In other words, weaker students tend to participate more (quantitatively and qualitatively), getting better results (especially) in the last task, which leads to a lower heterogeneity within the group: from 4.4 to 3.5 standard deviation (the amount of variation from the average results) after only five spaced learning sessions (47).

Also, due to its mutually engaging aspect, namely during the breaks which are usually spent in team work/games, spaced learning may encourage students to work towards a more collaborative environment as well.

The attention deficit disorder, whether with or without hyperactivity, (AD[H]D) has been classified as one of the most common neural conditions in childhood, with a prevalence of 3% in girls and 8% in boys, and also teenage years (Skokauskasa et al. 291). AD(H)D is the term used to indicate that an individual has a significant problem maintaining attention during a specific expected task (cf. Greenberg). As difficult as it is to diagnose and control this disorder, students with a higher number of inattentive reactions have shown a significant decrease in disruptive occurrences after only five spaced learning sessions (from 12 to 3 occurrences) (Vaz, "O Spaced Learning" 34, 43). Their performance after kinaesthetic activities during the pauses also produced

better results in the upcoming input/outcome (Vaz, *O Spaced Learning Enquanto Abordagem Pedagógica* 81). Both these conclusions come to show that spaced learning may help to control and discipline sustained attention, the ability to renew attention which allows people to "pay attention" to things that last longer than a few minutes.

Spaced learning also allows teachers to focus on a specific chosen skill or skills. This is especially important in the foreign language learning context. During the experiment in Portugal, for example, in which selected groups of students learnt a foreign language through spaced learning, a main skill was always in focus and allowed, therefore, students to acknowledge and overcome specific problems within that skill.

The students take took part in the experiment in Portugal acknowledged vocabulary acquisition and grammar as the most suitable and efficient for spaced learning during the experiment (112). Different students and different learning (dis)abilities may lead to different conclusions.

 Since spaced learning follows the same basic structure and the activities chosen for the pauses are usually based on either ready-made materials or games (e.g. play-dough moulding, Simon says) that need less preparation, it can be concluded that the method may be useful when there is a lack of time to teach the book and to prepare materials, as it acts as a mean of compressing information and boost imagination. Furthermore, since a single lesson may comprehend an extensive amount of contents, it may be also suggested that spaced learning is used as a revision tool (60).

Other aspects such as boredom or general lack of interest may also be valid reasons for opting for spaced learning. These two conditions were diagnosed during a pre-spaced learning phase and they were both overcame after its implementation in the classroom; the method produced a very positive reaction; becoming itself a motivation for the lesson (34).

In short, it may be said that spaced learning, when thoroughly prepared, may 1) be easily adapted to any skill/subject; 2) improve learning and studying methods; 3) entertain students during the pauses, thus motivating them for the upcoming input/output phases and allowing them

to develop new skills (e.g. juggling); 4) enhance collaborative learning; 5) facilitate the assessment of quantitative and qualitative output; 6) become a motivational and disciplinary tool.

On the other hand, a few weaker points should be taken into account before deciding for a spaced learning approach. Firstly, it is imperative that the pre-established times are respected, so students (and educators) who are not used to firm self-discipline should be the object of a more disciplined attitude before engaging in the method. Secondly, since there might be a strong teacher's presence throughout the whole session, teachers should opt for a monitoring attitude rather than risking to lose the students' autonomy. Finally, and acutely related to this last issue, the prevalence of the *teacher talking time* (TTT), the time that teachers spend talking in class, should be, at an early stage, regulated and previously drilled outside the teaching context so that it does not nullify the students' performance and levels of attention.

Another crucial aspect that should be taken into account before opting for a spaced learning approach, and perhaps the feature which makes every spaced learning trial unique, is the type of spaced learning which each group or individual demands. In other terms, it has been proved fundamental to get to know the specificity of a teaching target before effectively choose the method through which students will learn (Alberts and Wulf 55).

Some suggested approaches to do this may include preference and/or personality tests of different sorts (e.g. VARK test), previously observed needs, learners' suggestions and requests (e.g. the pause as reward: play a favourite game during the break). Whichever approach is taken, pleasing everyone, even if at different times, will be a sure way to bring the group together and implement democratic teaching and learning styles.

Despite the specific reason behind the use of spaced learning, least favourite activities, weaker skills and cooperation enhancement can be overcome even when the main issues are different (e.g. ADHD) (Vaz, "Spaced Learning in the MFL Classroom" 36).

Conclusion

It seems significant that teachers anticipate the impact, and even potential interference, of spaced learning, or any other approach, on the pre-established teaching aims. So the answer to the question on whether spaced learning may interfere with set-up teaching aims is *yes*, it can, but only if there is no preparation. This means that teachers should be able to adapt spaced learning to the syllabus, classroom environment and their own teaching style and aims – and never the other way round.

In summary, before setting up a lesson scheme or any short/medium/long-term plan for implementing spaced learning in the classroom, it is necessary to first make *space* for four essential questions: 1) Why use spaced learning?; 2) What is the target group type of spaced learning?; 3) What are the teaching aims?; and finally 4) How will spaced learning interfere with pre-established contexts, if at all? Keeping a record of all events (e.g. teaching log, output results) as well as taking into account previous feedback from students will surely lead, as it has already done, to outstanding results both in educational and neuroscientific contexts.

Further reading

Aiken, Catharine. *Methods of Mind-training: Concentrated Attention and Memory*. New York: Read Books, 2009.

Fields, R. Douglas. "Making Memories Stick." *Scientific American* 292.2 (2005): 75-81.

Kelley, Paul. *Making Minds: What's Wrong with Education, and What Should We Do about It?* London: Routledge, 2008.

Vaz, Ana. *O Spaced Learning Enquanto Abordagem Pedagógica Potenciadora da Atenção, Retenção e Memorização Exata nas Aulas de Língua Estrangeira do Ensino Secundário*. Diss. [Master's Degree in Education]. Universidade do Porto, 2013.

Further information

- spacedlearningteam@gmail.com
- www.facebook.com/spaced.learning
- http://spacedlearning.wix.com/spacedlearning
- http://makingminds.net/

Notes

[1] The UK's best measure of learning potential: higher scores being better (Kelley 154).

[2] A form of examination introduced in 2006, reducing the chance of prior learning in the groups (Kelley, *Making Minds* 154).

Works Cited

Aiken, Catharine. *Methods of Mind-training: Concentrated Attention and Memory*. New York: Read Books, 2009 [1895].

Alberts, Bruce, and William Wulf. *Science Teaching Reconsidered: A Handbook*. Washington DC: National Academy, 1997.

Bradley, Angela, and Alec Patton. "Spaced Learning: Making Memories Stick." *Innovation Unit*. 2008. Web. 19 Feb. 2014. <http://www.innovationunit.org/sites/default/files/Spaced_Learning-downloadable_1.pdf>.

Fields, R. Douglas. "Making Memories Stick." *Scientific American* 292.2 (2005): 75-81.

Greenberg, Lawrence M. "Attention Deficit Disorders and the Test of Variables of Attention (T.O.V.A.®)." *About Attention*. The TOVA Company. Web. 19 Feb. 2014. <http://www.tovatest.com/adhd/#index8h2>.

Kelley, Paul, and Terry Whatson. "Making Long-term Memories in Minutes: A Spaced Learning Pattern from Memory Research in Education." *Frontiers*. 25 Sept. 2013. Web. 19 Feb. 2014. <http://journal.frontiersin.org/article/10.3389/fnhum.2013.00589/full#h4>.

Kelley, Paul. *Making Minds: What's Wrong with Education, and What Should We Do about It?* London: Routledge, 2008.

- - - . "Neuroscience In Education Trials Announced." *Making Minds*. 9 Oct. 2014. Web. 19 Nov. 2014. <http://makingminds.net/2014/10/neuroscience-education-trials-announced/>.

Lapa, Carlos, Lídia Silva Mota, and Margarida Vilela. "Programa de Alemão 10.º, 11.º e 12.º Anos." *Ministério Da Educação*. Apr. 2001. Web. 19 Feb. 2014. <www.dgidc.min-edu.pt/data/ensinosecundario/Programas/alemao_iniciacao_10_11_12_anos.pdf>.

Philippou, Emily. "New Research to Investigate If Neuroscience Can Improve Teaching and Learning in Schools." *Wellcome Trust*. 9 Oct. 2014. Web. 19 Nov. 2014. <http://www.wellcome.ac.uk/News/Media-office/Press-releases/2014/WTP057594.htm>.

Reilly, Randall C., and Yuko Munakata. *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. Cambridge, Mass.: MIT, 2000.

Skokauskasa, Norbert, Fiona McNicholas, Tawfik Masaudc, and Thomas Frodld. "Complementary Medicine for Children and Young People Who Have Attention Deficit Hyperactivity Disorder." *Current Opinion in Psychiatry* 24 (2011): 291–300.

Vaz, Ana. "Spaced Learning." *Infoappa*. InfoAppa, Dec. 2012. Web. 19 Feb. 2014. <http://www.appalemao.pt/downloads/revista45.pdf>.

- - -. "O Spaced Learning Enquanto Abordagem Pedagógica Potenciadora da Atenção, Retenção e Memorização Exata nas Aulas de LE do Ensino Secundário." Presentation. Universidade do Porto. 24 July 2013a. Viva Voce.

- - -. "Spaced Learning in the MFL Classroom: Does Your Classroom Speak SL?." Training Course. Osiris Educational, London. 26 June 2013b. Lecture.

- - -. *O Spaced Learning Enquanto Abordagem Pedagógica Potenciadora da Atenção, Retenção e Memorização Exata nas Aulas De Língua Estrangeira do Ensino Secundário.* Diss. [Master's Degree in Education]. Universidade do Porto, 2013c. Print.

Method-effect on Test-takers' Performance and Confidence in Language Tests

Anna L. MOUTI and George S. YPSILANDIS
| Aristotle University of Thessaloniki

Abstract | In a language test, the number of correct answers is commonly used as the only source of information that defines the result. The question raised is the following: do all task types accommodate all test-takers in the same way? This paper attempts to investigate this relationship by relating correctness and confidence in the answers provided with the task typology. Two achievement tests in two versions (same language content – different task type) were used while test-takers were asked to mark their degree of confidence for each item on both versions. Data obtained through this research were: 1) number of correct answers, and 2) degree of confidence in different task types.

Through the correlations of the variables examined (task typology, correctness and confidence) several findings were registered concerning the relationship between accuracy and confidence in different task types. The task type typology may be seen as a moderating factor which could be responsible for possible variations in accuracy and confidence scores.

Key words | language test performance, confidence, method effect, test question formats, task types

1. Introduction

The driving force behind this research is the need for a more precise and fair language measurement. When it comes to a language test, it seems that the number of correct answers is the only source of information that defines the result. This approach has only recently been challenged. In Schraw's words: "Effective Test-taking depends on two important skills: selecting correct responses to test questions and monitoring one's performance accurately" ("The Effect of Generalized Metacognitive Knowledge" 135). It may be possible to suggest that the process and measurement of self-monitoring is enhanced through the measurement of confidence. Indeed, Stankov and Lee consider confidence as an individual difference which could be defined as "a systematic tendency that leads one to act in a particular way because it reflects a belief, a faith in oneself" (962). Registering confidence may be another variable (adding to the test-taker profile) which affects test performance by raising metacognitive strategy awareness but also by providing metacognitive information for a more fair and precise scoring.

In the field of Applied Linguistics it was Yule who first traced the notion and the need for the study of confidence. Yule, Yanz and Tsuda "set out to quantify, in a limited way, the effects of that one important affective variable called confidence in one area of the language learner's performance" (475). Their experiment involved the following procedure: each time the participants chose an answer, they had to indicate how confident they were about the correctness of their answer. This approach is adopted in the framework of this study setting language test performance and confidence as the two dependent variables during the test-taking process.

In the discussion for a more precise and fair language testing procedure Bachman claimed that "a major concern in the design and development of language tests is to minimize the effects of the factors that are not part of the language ability" (*Fundamental Considerations* 166). Further, Bachman and Palmer stated that "[a] number of language testers have indicated that we should attempt to design our tests to elicit test-takers' best performance. We believe that one way to do this is to design the characteristics of the test task so as to promote feelings of comfort or safety in test-takers that will in turn facilitate the flexibility of response on their part" (66). In order to

guarantee a precise and fair measurement of the two variables stated above, throughout the study but also in real-life situations and authentic test-taking settings, it was attempted to examine another key independent variable: the question format. Possible positive correlations between the "question format" and language test performance and confidence could place under consideration fairness of language test scores but also the degree to which task-types "accommodate" the test-takers.
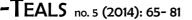
2. Question Formats

The independent variable of our research is the task typology and question format used in language tests. More specifically the study concentrates on closed-type tasks, where a fixed response is required. Some of the most frequent closed-ended task types used in language tests are the following:

Closed-ended Test Question Formats	
Selected-response	True/False
	Multiple-Choice
	Matching
Constructed-response	Gap-Filling
	Short answer

Table 1. *Close-ended Test Question Formats*

Douglas sees a test as "a measuring device, no different in principle from a ruler, a weighing scale or a thermometer" (2). For Douglas, "a language test is an instrument for measuring language ability" although the author expresses the following doubt: "In what sense can we actually measure a concept as abstract as *language ability*?" (2). Although it would require multiple papers to provide an answer to this question, it is customary to measure language ability though a number of close-ended questions which are accepted to be used to elicit test-takers' responses. Any estimate of language ability is based on the provided answers; to locate the degree of knowledge in a rather conventional and to a certain extent arbitrary manner. As presented in the

above table these responses can be either selected or constructed. Another term used for constructed-response is provided by Purpura, who makes the distinction between selected-response tasks and limited production tasks which "elicit the response embodying a limited amount of language production" (134). Our research hypotheses will rotate around multiple-choice questions and gap-filling ones as these two are used as representative task types of the selected and constructed response tasks.

Multiple-choice (MC) questions are the most commonly used to measure language competence because they are quick, economical and straightforward to score. A typical MC test item consists of two basic parts: the stem (a question or a problem to be solved) and a list of possible answers, which usually contains one correct (or "best") answer and a number of incorrect options (distractors). Burton et al. give six types of MC items (12-15):

(a)   Items of the *single-correct-answer* variety

(b)   Items of the *best-answer* variety

(c)   Items of the *negative* variety

(d)   Items of *multiple-response* variety

(e)   Items of the *combined-response* variety, and

(f)   Items of *multiple true or false* variety.

Some of the constructed response questions are gap-filling items which are also frequently used in testing. There is a variety of these questions like the cued (or guided) gap-filling questions. In these "the gaps are preceded by one or more lexical items or cues which must be transformed in order to fill the gap correctly" as indicated by Purpura (136). Purpura further suggests another type of gap-filling task: the cloze. In this task type every fifth, sixth or seventh word is deleted and the test-takers are asked to fill in the gap. It was Wilson Taylor in 1953 who first introduced the cloze procedure but there have been a number of variations since then: the standard cloze test, the modified cloze test, the multiple-choice cloze test, the c-test and the cloze elide. The modified

cloze test might be otherwise expressed as selective-deletion gap-filling, setting the "selective" as against random (Weir, Vidakovic, and Galaczi 158).

3. "Method-effect" on Performance and Confidence

Task types have been thoroughly examined in the past and there is a lot of research concerning the "method effect" on language test performance. Alderson, Clapham and Wall offer a working definition of the method effect in that "the method used for testing language ability may itself affect the student's score" and further claim that "its influence should be reduced as much as possible" (44).

Bachman initially discussed a framework for the facets of test methods that may affect language test performance (*Fundamental Considerations* 119). These facets were grouped into five sets: the facets of the testing environment, the facets of the test rubric, the facets of the input the test taker receives, the facets of the expected response, and the relationship between input and response. This was further developed in Bachman and Palmer (47-57) in a framework of language test task characteristics (setting, rubrics, input, expected response and relationship between input and response) where a suggestion is made in that test creators are expected to "design the characteristics of the test task so as to promote feelings of comfort or safety in test-takers that will in turn facilitate the flexibility of response on their part" (66). Shohamy and Wolf found that test methods influenced language test performance and that multiple-choice questions were easier than open-ended questions. This claim was investigated by several researchers. Tsagari investigated the above claim empirically and compared the effects of two test formats (free response and multiple choice) on two reading comprehension tests with identical content. Their findings revealed that method effects can be a source of variation of scores and may even measure different abilities. Cheng in research on listening comprehension came to a similar conclusion in that multiple-choice cloze was easier than regular (distinct) multiple choice and that the open-ended questions were found to be more difficult. In the research conducted by Zheng, Cheng and Klinger, participants achieved a higher percentage of correct responses in multiple-

choice questions and lower in constructed-response questions in reading comprehension. Liu, in 1998, used multiple-choice questions, true or false questions and short answer questions and the results showed that there were significant differences among the scores elicited by the three different test methods with short answer questions being the most difficult (147). Later, Liu conducted a similar research with reading comprehension and concluded that different test methods affect students' scores in this receptive skill. In this research, the gap-filling test was considered the most difficult, while multiple-choice questions and short answer questions were easier. In'nami and Koizumi conducted a meta-analysis on the effects of multiple-choice and open-ended formats on L1 reading, L2 reading, and L2 listening test performance. Fifty-six data sources were located and the results indicated that multiple-choice formats are easier than open-ended formats in L1 reading and L2 listening, with the degree of format effect ranging from small to large in L1 reading and medium to large in L2 listening. Currie and Chiramanee investigated the effect of multiple-choice format, compared with a constructed-response test on equivalent language test items. The scores of the two tests were found to be highly correlated but only 26% of the answers (either correct or wrong) were the same, with the multiple-choice questions being easier. In a recent research conducted by M. Salehi and H. Bagheri Sanjareh, response format was also found to be responsible for variations in language test performance. In general, researchers have proved that in cases that question format is found to be responsible for score variations, selected response type is considered easier than the constructed response type.

Although the impact of question format on language test performance was investigated by a great number of researchers in the last twenty-five years, the literature review about the impact of question/test format/method on language test confidence came up with very limited results. Findings of Pallier et al. show that "when one is answering a question, one's level of confidence is sensitive to question format" (262). Pallier et al. were based on previous research conducted by Kohler, who reached the conclusion that questions which ask for a hypothesis to be expressed generate less confidence than those which provide some alternatives for evaluation. Of course Kohler's research was not implemented in a language test setting but was part of a psychology

experiment. However, it may be possible to conclude that hypothesis generation is nearer to open-ended, cloze or short-answer questions while the alternatives could be considered to be linked to multiple-choice. Thus, Kohler's results could be implemented in a language testing setting. As is implied by the limited number of sources on this issue, further research would need to be conducted.

4. Research Method

4.1. Hypotheses-Design

In this study there are three variables under investigation: Test Question format as an independent variable in relation to two dependent variables, language test performance and language test confidence. More specifically, two research hypotheses were tested:

1. There is a relationship between task-types/test question formats and language test performance.
2. There is a relationship between task-types/test question formats and language test confidence.

In order to investigate the above two hypotheses a quasi-experimental realistic research was designed. The subjects were asked to complete two sets of test items, with the same language content and different task type/question formats. In addition the participants had to mark their level of confidence on a 100mm bar, indicating how confident they were feeling about the correctness of their answer on each item.

4.2. Instruments

The instruments used for measuring language test performance were 6 sets of language test items of constructed response and 6 sets of language test items of selected response. The sets of items were used to test linguistic knowledge relevant to grammar and vocabulary and were part of a longer test used as the final exam of two EAP language courses. There were 6 sets of

items – presented in the first stage of the research – in a constructed response format. More specifically there were 5 cued gap-filling tasks and 1 gap-filling task (selective deletion gap-filling/modified cloze), checking meaningful grammatical and lexical items. During the second phase of the research, the subjects were asked to complete the same sets of items in a different test format, i.e. multiple choice. Based on the studies by Rodriguez and In'nami and Koizumi, test items were constructed with stem equivalency. This was used to examine the impact of task typology on performance (format effects). A similar study using a stem equivalent format was used by Currie and Chiramanee. All the test items were statistically analysed and the facility index was between acceptable borderlines, between 0.48 and 0.82. The test items were scored dichotomously, both the multiple-choice and the gap-filling ones by "scoring for the contextual appropriateness i.e. to count as correct any word that fully fits the total surrounding both syntactically and semantically (Weir, Vidakovic, and Galaczi 70).

The process of confidence marking was implemented in 4 sets of items of constructed response and 4 sets of language test items of selected response. For the confidence measurement, a 100mm confidence bar was adopted for each item on which the subjects could mark "*how confident did they feel about the correctness of their answer*". The subjects were asked to mark their confidence using the bar (ravdos) suggested by P. Kambaki-Vougioukli and T. Vougiouklis. The adoption of the Vougioukli and Vougiouklis bar (ravdos) is claimed to be simpler and easier for both the subjects of an empirical research study and the researcher. Its advantages lie in both the design stage and the processing of the results as it is more flexible than the typical Likert scale.[1] In Kambaki-Vougioukli et al. (82) and Vougiouklis (20) it is clearly stated that:

> In every question, substitute the Likert scale with "the bar" whose poles are defined with "0" on the left and "1" on the right:
>
> 0_____1
>
> The subjects/participants are asked, instead of deciding and checking a specific grade on the scale, to cut the bar at any point they feel best expresses their answer to the specific question.

### 4.3.	Participants

All the participants were first and second year students, of Greek nationality and Greek mother tongue, aged 18-20. Ninety-eight of those were male and 102 female. The majority hold a B2 certificate and shared common educational characteristics, concerning the same educational background and similar EAP experience in their university department. They were all familiarized with language testing procedures and they have all attended English language courses during their primary, secondary and tertiary education.

Four groups of a total of 200 participants (separated according to the EAP course they were attending and who were assessed)[2] completed 12 sets of items in total, 6 of constructed-response and 6 of selected-response. The first group (55 students) completed 2 sets of items, the second group (48 students) 6 sets of items, the third group (65 students) completed 2 sets of items and the fourth group (32 students) completed 2 sets of items during the two stages of the research.

The first two groups (a total of 103 students) completed the 8 sets of items and they also marked their degree of confidence on the 100mm bar after each response they gave on the test.

5. Results

The research product was two series of items, completed and marked in terms of the degree of confidence for each item. Thus two sets of scores – the performance/accuracy scores (degree of accuracy) and the confidence score (degree of confidence) – were obtained and all of them were converted into percentages to create homogeneity in the dataset.[3] These two series of scores were further divided into four sets of scores. Two sets of scores (accuracy and confidence) for the constructed-response items and two sets for the selected-response items of the 8 sets of items examining both accuracy and confidence are presented in the following table. As expected, MC presents, overall, higher performance and confidence scores/percentages than the constructed-response, leading to a first conclusion that task types/test item format is responsible for variations on accuracy and confidence scores.

	Constructed-response (Mean)	Selected-response (Mean)
Performance	62.45%	74.49%
Confidence	65%	72 %

Table 2. *Performance & Confidence – Mean*

5.1. Test Question format * Language Test Performance

In the following table we can see the descriptive statistics of all the tasks used in the research procedure. It is obvious that all the multiple-choice tasks may be considered easier than the gap-filling ones as they provide a higher score. Although it was not included in our research hypotheses to examine the variations of gap-filling exercises in relation to language test performance, we can assume that there is no impact of the gap-filling variation on performance (cued gap-filling and selective gap-filling).

PERFORMANCE		N	Mean	Std. Deviation
P1	MC[4]	55	79.04	18.965
P2	CGF	53	56.98	25.178
P3	MC	55	77.53	18.989
P4	CGF	55	67.25	21.696
P5	MC	55	79.05	18.154
P6	SGF	55	67.93	17.513
P7	MC	48	81.88	19.388
P8	CGF	48	57.62	27.919
P9	MC	65	60.80	18.924
P10	CGF	72	55.00	21.189
P11	MC	32	74.38	28.391
P12	CGF	32	47.50	32.429

Table 3. *Language Test Performance Scores – Descriptive Statistics*

The scores were further analysed in each pair of the sets of items examined. As shown in the following table there is a correlation between all pairs showing high reliability between the two

formats, which implies that the test-items are measuring the same construct. It should be noted that the correlations (although they are considered statistically significant) do not reach very strong associations (0.679 at the highest). Further, strong differences in the scores obtained were registered, showing that the differences in scores would be very important when criterion-referenced scores are concerned. That means that the students' ranking (norm-referenced situations) would be the same but when it comes to a decision taken based on the total score according to a pass/fail cut-off point a strong variation would occur. The selected-response items were easier for the participants and thus the facility index for the selected-response items was higher than the constructed-response one, resulting in a higher total score for the multiple-choice test items.

	Mean	Paired Difference	Correlation	Sig.
P1-P2 (MC-CGF)	79.51-56.98	22.528 (.000)	0.569	0.000
P3-P4 (MC-CGF)	77.53-67.25	10.273 (.000)	0.679	0.000
P5-P6 (MC-SGF)	79.05-67.93	11.127 (.000)	0.410	0.002
P7-P8 (MC-CGF)	81.88-57.63	24.250 (.000)	0.546	0.000
P9-P10 (MC-CGF)	60.80-55.00	5.015 (.056)	0.430	0.000
P11-P12 (MC-CGF)	74.38-47.50	26.875 (.000)	0.678	0.000

Table 4. *Question Format * Performance Scores*

Similar correlations were reached in Danili and Read, Johnson and Ambusaidi, and Currie and Chiramanee for selected and constructed response items. Additionally, MC questions were found to be easier and thus resulting in a higher score than the Constructed-response questions supporting evidence of previous studies mentioned in the Literature Review.

5.2. Test Question format * Language Test Confidence

Confidence scores were also analysed in detail, but the results were not similar to the performance scores. It is evident from the descriptive statistics that in all the cases multiple-choice degree of confidence was higher than in gap-filling. However, the differences were not substantial.

CONFIDENCE		Mean	N	Std. Deviation
C1	MC	78.10	49	20.997
C2	CGF	72.90	49	16.989
C3	MC	75.26	47	19.332
C4	CGF	73.43	47	26.225
C5	MC	73.09	46	19.723
C6	SGF	67.67	46	20.978
C7	MC	79.18	39	21.541
C8	CGF	62.64	39	24.508

Table 5. *Language Test Confidence Scores – Descriptive Statistics*

As shown in the following table there is a strong correlation between all pairs of questions showing high consistency between the two formats of questions. In addition, there are differences between the confidence scores obtained by the selected and the constructed response items, showing that the question format could not be considered as an important factor responsible for variations on confidence scores. These findings lead us to assume that factors other than the question format could be responsible for variations in confidence scores. The facility index, the performance/accuracy scores and the knowledge of the language items/content could be one of them but also some inherent characteristic of the test-taker. Therefore it was attempted to calculate the correlation between different pairs. A number of correlations were found to be very strong showing that there might be some personal trait related to self-confidence.

	Mean	Paired Differences	Correlation	Sig.
C1-C2 (MC-CGF)	78.10-72.90	5.204 (0.053)	0.552	0.000
C3-C4 (MC-CGF)	75.26-73.43	1.830 (0.573)	0.566	0.000
C5-C6 (MC-SGF)	73.09-67.67	5.413 (0.118)	0.361	0.014
C9-C10 (MC-CGF)	79.18-62.64	16.538 (0.000)	0.769	0.000

Table 6. *Question Format * Confidence Scores*

Similar findings were recorded in Pallier et al.: MC test items gathered higher confidence scores although in our research the differences were not statistically significant in all the cases.

6. Conclusions

Our initial hypotheses have only been partially supported by the evidence. In particular, our findings revealed that task-type/question format could be a source of variation of scores while it could not be considered as a determining factor responsible for variations in confidence scores. It should be emphasized that there was a strong correlation between the scores obtained by the selected and constructed-response question formats, showing that there is no significant impact of question formats on the ranking of students. As indicated by Farhady,

> in a norm-referenced situation, the increase at the level of the scores would not influence the ranking of the students, i.e. all testees will score higher on the multiple-choice form than they will on the open-ended form. In a criterion-referenced situation however, where there exists a predetermined criterion for the students to meet, low scores would hurt those at the borderline. (222)

In norm-referenced situations the increase at the level of scores, which could be attributed to question format, may not have a significant impact while in criterion-referenced situations the choice of the question formats would need to be carefully examined and implemented. Alternatively, any possible variations would need to be supported by equivalent measurements.

The non-significant amount of confidence variation which could be attributed to question formats implies that participants feel equally comfortable with the question formats under investigation, expressing some kind of extra confidence when dealing with multiple-choice questions, something which could be attributed to the higher facility index.

It should be mentioned that the present study may be subject to some limitations as there were different groups of participants responding to different test-items and not one single group or different groups responding to the same set of items. This was a result of the actual setting and a conscious decision not to jeopardize authenticity of the conditions under which the research was conducted. In support of this decision, Cormick suggested that "it is still unclear whether laboratory-based findings from metacognitive studies can be generalized to actual classroom tasks" (qtd. in Nietfeld, Cao, and Osborne 11). Most certainly this research design could be considered as an implementation of a small-scale replication study which is offered to add to the discussion on the topic.

Notes

[1] The issue of the bar has been investigated in a number of papers (Kambaki-Vougioukli and Vougiouklis, Kambaki-Vougioukli et al. and Vougiouklis).

[2] It should be mentioned that the variables EAP language course and level of proficiency were not proved to affect their overall language test performance.

[3] As mentioned also by Bachman (*Statistical Analyses* 39) for the criterion-referenced language test scores. A similar procedure was also followed for the confidence scores by Stankov (128) and Schraw et al. (434).

[4] MC is an abbreviation for Multiple-Choice, CGF for Cued Gap-Filling, and SCF for Selective Gap Filling.

Works Cited

Alderson, Charles J., Caroline Clapham, and Dianne Wall. *Language Test Construction and Evaluation.* Cambridge: Cambridge UP, 1995.

Bachman, Lyle F. *Fundamental Considerations in Language Testing*. Oxford: Oxford UP, 1990.

- - - . *Statistical Analyses for Language Assessment*. Cambridge: Cambridge UP, 2004.

Bachman, Lyle F., and Adrian S. Palmer. *Language Testing in Practice*. Oxford: Oxford UP, 1996.

Burton, Steven J., Richard R. Sudweeks, Paul F. Merrill, and Bud Wood. *How to Prepare Better Multiple-choice Test Items: Guidelines for University Faculty*. Brigham Young University Testing Services and Department of Instructional Science, 1991. Web. 12 March 2014 <http://testing.byu.edu/info/ handbooks/betteritems.pdf>

Cheng, Hsiao-fang. "Comparison of Multiple-Choice and Open-Ended Response Formats for the Assessment of Listening Proficiency in English." *Foreign Language Annals* 37.4 (2004): 544-53.

Currie, Michael, and Thanyapa Chiramanee. "The Effect of the Multiple-choice Item Format on the Measurement of Knowledge of Language Structure." *Language Testing* 27.4 (2010): 471-91.

Danili, Eleni, and Reid Norman. "Assessment Formats: Do They Make a Difference?" *Chemistry Education Research and Practice* 7.2 (2005): 64-83.

Douglas, Dan. *Understanding Language Testing*. New York: Routledge, 2014.

Farhady, Hoossein. "Varieties of Cloze Procedure in EFL Education." *Roshd Foreign Language Teaching Journal* 12 (1996): 217-29.

 In'nami, Yo, and Rie Koizumi. "A Meta-analysis of Test Format Effects on Reading and Listening Test Performance: Focus on Multiple-choice and Open-ended Formats." *Language Testing* 26 (2009): 219-44.

Johnstone, Alex. H., and Abdullah Ambusaidi. "Fixed Response: What Are We Testing?" *Chemistry Education: Research and Practice in Europe* 1.3 (2000): 323-28.

Kambaki-Vougioukli, Penelope, and Thomas Vougiouklis. "Bar Instead of Scale." *Ratio Sociologica* 3 (2008): 49-56.

Kambaki-Vougioukli, Penelope, Alexander Karakos, Nikolaos Lygeros, and Thomas Vougiouklis. "Fuzzy Instead of Discrete." *Annals of Fuzzy Mathematics and Informatics* 2.1 (2011): 81-89.

Koehler, Derek. "Hypothesis Generation and Confidence in Judgment." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20 (1994): 461-9.

Liu, Feng. "The Effect of Three Test Methods on Reading Comprehension: An Experiment." *Asian Social Science* 5.6 (2009): 147-53.

Nietfeld, John L., Li Cao, and Jason W. Osborne. "Metacognitive Monitoring Accuracy and Student Performance in the Classroom." *Journal of Experimental Education* 74.1 (2005): 7-28.

Pallier, Gerry, Rebecca Wilkinson, Vanessa Danthiir, Sabina Kleitman, Goran Knezevic, Lazar Stankov, and Richard Roberts. "Individual Differences in the Realism of Confidence Judgments." *Journal of General Psychology* 129 (2002): 257-300.

Purpura, James E. *Assessing Grammar*. Cambridge: Cambridge UP, 2004.

Rodriguez, Michael C. "Construct Equivalence of Multiple-choice and Constructed-response Items: A Random Effects Synthesis of Correlations." *Journal of Educational Measurement* 40.2 (2003): 163-84.

Salehi, Mohammad, and S. Bagheri Mohammad. "The Impact of Response Format on Learners' Test Performance of Grammaticality Judgment Tests." *Journal of Basic and Applied Scientific Research* 3.3 (2013): 1335-345.

Schraw, Gregory. "The Effect of Metacognitive Knowledge on Local and Global Monitoring." *Contemporary Educational Psychology* 19.2 (1994): 143-54.

- - -. "The Effect of Generalized Metacognitive Knowledge on Test Performance and Confidence Judgments." *Journal of Experimental Education* 65.2 (1997): 135-46.

Schraw, Gregory, Michael E. Dunkle, Lisa D. Bendixen, and Teresa DeBacker Roedel. "Does a General Monitoring Skill Exists?" *Journal of Educational Psychology* 87 (1995): 433-44.

Shohamy, Elana. "Does the Testing Method Make a Difference? The Case of Reading Comprehension". *Language Testing* 1.2 (1984): 147-70.

Stankov, Lazar. "Complexity, Metacognition, and Fluid Intelligence." *Intelligence* 28 (2000): 121-43.

Stankov, Lazar, and Lee Jihyun. "Confidence and Cognitive Test Performance." *Journal of Educational Psychology* 100.4 (2008): 961-76.

Tsagari, Dina. "Method Effect on Testing Reading Comprehension: How Far Can We Go?" Master Thesis. University of Lancaster, 1984.

Vougiouklis, Thomas. "Enlarged Fundamentally Very Thin Hv-Structures." *Journal of Algebraic Structures and Their Applications* 1.1 (2014): 11-20.

Vougiouklis, Thomas, and Penelope Kambaki-Vougioukli. "On the Use of the Bar". *China-USA Business Review* 10. 6 (2011): 484-89.

Weir, Cyril J., Ivana Vidakovic, and Evelina Galaczi. *Measured Constructs: A History of the Constructs Underlying Cambridge English Language (ESOL) Examinations 1913-2012*. Cambridge: Cambridge UP, 2013.

Wolf, Darlene. "A Comparison of Assessment Tasks Used to Measure FL Reading Comprehension." *Modern Language Journal* 77 (1993): 473-89.

Yule, George, Jerry Yanz, and Atsuko Tsuda. "Investigating Aspects of The Language Learner's Confidence: An Application of the Theory of Signal Detection." *Language Learning* 35. 3 (1985): 473-88.

Zheng, Ying, Liying Cheng, and Don A. Klinger. "Do Test Formats in Reading Comprehension Affect Second-language Students' Test Performance Differently?" *TESL Canada Journal* 25.1 (2007): 65-80.