



Testing
& Evaluation
& Assessment

The Newsletter of the Testing, Evaluation and Assessment Special Interest Group

TEASIG NEWSLETTER

Issue 63

DECEMBER 2017



- | | | |
|----|--|----------------------------|
| 3 | Editorial | |
| 4 | Letter from the Coordinators | |
| 5 | The TEASIG committee | |
| 6 | Washback of Practical English examinations on the process of teaching and learning | Elżbieta Zawadowska-Kittel |
| 11 | Education or evaluation? | Ilya Denisenko |
| 15 | Maximising collaboration in oral English testing: an example from China | Susanna Wickes |
| 19 | Receptivity to learner-driven feedback | Clare Maas |
| 24 | Giving testing some good press for a change:
Assessment as learning check - a meaningful paradigm shift for trainee teachers | Lynn Williams |
| 28 | "En mi humilde opinión" ...listening to Mexican students' perceptions of an English language proficiency test | David Ewing Ryan |
| 35 | IATEFL TEASIG Webinar(24 October 2017): Developing a test: Where do you start?
When should you stop? | Jo Tomlinson |
| 38 | TEASIG Member Spotlight | Christine Coombe |
| 40 | TEASIG Member Spotlight | Gladys Quevedo Camargo |

"En mi humilde opinión"¹ ...listening to Mexican students' perceptions of an English language proficiency test



David Ewing Ryan

is a full-time professor at the Universidad Veracruzana (UV), Mexico, where he conducts research on language testing and serves as Chief Editor of EXAVER, the UV's tiered-suite of English language certification tests.

dewing@uv.mx

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 and areas that have traditionally been overlooked (Cumming, 2004, p.3). It can be successfully argued that one of these neglected areas is Mexico. Mexico seems to find itself in the paradoxical situation of many Latin American countries which, while witnessing a pronounced increase in demand for high quality English language instruction and assessment, have failed to produce a significant body of research investigating the specific variables that help to define the uniqueness of their contexts.

Without doubt, one of these variables is the students or candidates who actually *take* language tests. Their story is largely untold, and in recent language testing literature it is difficult to find an issue that more scholars seem to agree on than the idea that candidates are among the most important – yet neglected – stakeholders in language testing. Shohamy (2001) perceives that "it is through the voices of test takers who report on their 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" (p.7). Cumming (2004)

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" (p.9).

The purpose of this study, therefore, was to give free rein to the neglected voices of test candidates in one particular context. As McNamara and Roever (2006) insist, "language testing has a real impact on real people's lives" (p.8). This impact starts with the stakeholders who are immediately affected by the test, i.e. test candidates and test developers, and extends outward to society at large. This impact also 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 test validity is by listening to candidates' voices. 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 among a variety of stakeholders (Madaus, in press, as cited in Shohamy, 2001, p.149). Enlarging

the dialogue in this way can help further promote not just the validity of individual tests, but also of the test system, which 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” (Shohamy, p.131).

The article has six parts. Part 1 outlines the goal of the study. Part 2 summarises the EXAVER English language certification tests, which served as the practical context of the study. Part 3 explains the methodology of the study. Parts 4 and 5 offer, respectively, an overview and a discussion of the findings. Finally, Part 6 offers some general conclusions.

1. Goal of the study

The study focused specifically on what Bachman and Palmer (1996) consider as one of the three ways that language tests have a

direct impact on test candidates, namely, the consequences that candidates experience as a result of preparing for and taking these tests (p.31). In order to measure this impact, it was necessary to liberate the voices of the candidates who participated in the study, and this became the study's primary goal. This was accomplished, first, by so-

liciting candidate opinions about the processes of preparing for and taking the test, and then by soliciting their suggestions on how these processes might be improved.

2. Practical context of study: the EXAVER English Language Certification Tests

2.1 General description

EXAVER is the name of the tests used as the basis of the study, and refers to a tiered-suite of English language certification tests developed and administered by the Universidad Veracruzana (UV) in the southeastern Mexican state of Veracruz. The first suite was developed in the year 2000 by a small group of English language teachers at the UV, as well as by representatives from the

EXAVER tests are administered twice a year at 11 language centers throughout Veracruz.

Table 1: Levels of EXAVER tests and their corresponding CEFR Levels (adapted from Abad et al., 2011)

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

2.2 Test Structure

Each EXAVER test contains three separate papers. The structure of each paper is described below.

Table 2: EXAVER test structure (after Dunne, 2007)

Paper 1 Reading and Writing	Paper 2 Listening	Paper 3 Speaking
<ul style="list-style-type: none"> 5 parts Variety of tasks: matching, multiple choice, modified cloze text Indirect measure of writing 	<ul style="list-style-type: none"> 4 parts Range from comprehension of relatively short informal conversations to comprehension of more formal and substantially longer conversations 	<ul style="list-style-type: none"> 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)

British Council, Cambridge Assessment, and Roehampton University's Center for Language Assessment and Research (CLARe). The construct behind the EXAVER tests is to measure three language proficiency levels identified in the Council of Europe's *Common European Framework of Reference for Languages* (CEFR), summarised in Table 1. The

2.3 Test localization

According to O'Sullivan (2011), 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” (O'Sullivan, p.10). In focus-

ing their attention on the local geographic context of the examinations (southeastern Mexico) and the particular needs of the candidates within that context (students, primarily, of the Universidad Veracruzana), EXAVER's test developers have 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..." (O'Sullivan, p.6).

Economic affordability was one of the first local variables that EXAVER's test developers considered. As the majority of EXAVER's candidates could not (and still cannot) afford the cost of more reputable international English language certification tests, EXAVER's test developers decided to create a suite of economically affordable tests, in line with median to lower income brackets based on the Mexican minimum wage.² Table 3 shows the current costs (as of September 2017) of taking an EXAVER test, with approximate equivalents in Euros.³

As of January 2017, the Mexican minimum wage was approximately 80 pesos per day.

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) pages 228-243.

Table 3: Comparative cost of taking an EXAVER test

LEVEL	Cost in MX Pesos	Cost in Euros
EXAVER 1	500	Approx. 23
EXAVER 2	550	Approx. 25
EXAVER 3	600	Approx. 27

3. Methodology

A mixed methods (quan → QUAL) approach for data collection and analysis was used for the study. The quantitative data came from the responses of 245 EXAVER candidates who completed a web-based questionnaire, which was administered in the summer of 2010 following the spring 2010 administration of EXAVER's three levels. The questionnaire included 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, to ascertain candidates' opinions about several topics related to the test. Excel Version 2003 was used to analyze the data. The qualitative data came from the author's research journal from March to October 2010, and from semi-structured interviews conducted in October 2010 with four of the questionnaire's respondents.

4. Findings ⁴

Out of the 964 candidates who took an EXAVER test in May 2010, 245 of them (or 25%) responded to the web-based survey. Of these, 99 (or 40%) ticked the box at the end of the survey, signifying their desire to

participate in a semi-structured interview. This relatively high rate of response was the first indication of the apparent strong desire of EXAVER's candidates to have their voices heard.

Web-based questionnaire survey

As research instruments, questionnaires have their distinct advantages and disadvantages. With regard to the latter, researchers sometimes complain about the lack of depth and richness in multiple-choice responses (Dörnyei, 2007, p.115). For this reason, the researcher included two open-ended questions in the survey along with the 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 (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 took part in the survey), as well as to the diversity of their answers. Summaries of these responses follow.⁵

Question 17: "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 responses, organized into the following categories:

⇒ 23 positive responses, such as:

- “I didn’t hire a tutor or use any books to prepare for the test, as I found the information on the website very useful.”
 - ⇒ 24 negative responses, such as:
 - “The waiting time to get your grade is too long...you really need to find a way to make it go faster.”
 - “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, notably for further documentation or materials to include on the EXAVER website, such as:
 - a video of a sample speaking test
 - a bibliography of literature to consult to help prepare for the test
 - a description of how grades are calculated
- Question 30:** “Do you have any other comments (positive or negative) and/or suggestions that you’d like to add regarding the specific test you took or about the EXAVER Certification Tests in general? If so, please write them below, taking all the space that is necessary.” Question 30 yielded 127 responses.
- ⇒ 38 positive responses, such as:
- “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.”
 - ⇒ 61 negative comments, such as:
 - “While waiting in line to enter the test center, 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.”
 - ⇒ 28 suggestions, such as:
 - “It would be nice to have a more detailed report on how I fared in the test, such as knowing how I performed on each part of the test, maybe in terms of percentages.”

5. Discussion

5.1 Specific concerns

The phrasing of Question 17, with special emphasis on the words “more confident” and “less anxious”, was intentional in order to reflect the researcher’s premise that the less anxious and more confident candidates feel before taking a test, the more likely they are to perform better.⁶ The relatively long list of suggestions (97 in total) that candidates gave in response to this question have proven useful in helping EXAVER’s test developers improve the quality of the website’s preparation materials so that candidates

can indeed feel more confident and less anxious before taking a live test.

Question 30 should seem familiar to qualitative researchers, since it is the classic “Do you have anything else to add?” that usually appears at the end of an interview. It was considered necessary as a ‘safety net’ to ensure that candidates had the opportunity to state anything and everything they wished to about 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 test registrants, relates to the theme of Question 17. It should remind testers of the importance of adopting measures that help to avoid circumstances that can potentially cause unnecessary stress and anxiety for candidates. Test examiners and administrators should develop a list of things that could possibly go wrong on the day of the test, identifying an effective solution for each of them. Such a list should then be printed and given to test invigilators.

By contrast, one of the positive responses to Question 30 illustrates how a seemingly routine task (calmly and clearly reading the initial instructions once candidates are seated) can actually serve to minimize test anxiety and boost candidates’ sense of confidence. Both examples reinforce the importance of ensuring that the test ‘reception’ process (the way that candidates

are physically greeted and treated by examiners and invigilators both prior to and during the test) is as smooth and professional as possible.

5.2 General concerns

Candidate responses to both the questionnaire 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 also provided EXAVER's test developers with important insights into some of the positive and negative consequences for test candidates as a result of preparing for and taking a language test. With regard to the questionnaire, the quantity and variety of responses bring to mind Shohamy's (2001) 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 (p.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 (Shohamy, p. 136-137).

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 using it for.

For this reason, one might correctly refer to the 'final consequences' of consequential validity. The final actions taken by test developers may very well serve to increase the occurrence of positive consequences for future candidates and to de-

"With regard to the questionnaire, 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."

crease the occurrence of negative consequences.

The following actions, based on candidate feedback from this and other studies, have already been taken (or are currently being undertaken) by EXAVER administrators and serve to illustrate how a language test board can convert candidate feedback into positive impact for future candidates:

- 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 (These tests complement the sample reading, writing, and listening tests on the website since EXAVER's inception.)

- Drafting of a list of administrative procedures that may be problematic for examiners and invigilators on the test day, along with corresponding solutions
- Dissemination of candidate feedback questionnaires as a way of continuing to monitor the positive and negative consequences for candidates taking the tests
- Analysis and discussion of appropriate action(s) based on candidate responses

- Follow-through to confirm that appropriate action was taken

6. Conclusion

By now it has perhaps become apparent to the reader that what candidate feedback in language testing actually relates to is a type of assessment that is more inclusive and democratic in nature than the traditional, authoritarian type that was prevalent in so many assessment contexts throughout the world during much of the 20th century and, indeed, prior to that.⁷

When test developers refuse to solicit candidate feedback, or do so without following through on it, the undemocratic nature of the assessment can be seen in terms of the

power and control that testers exert over candidates. Conversely, when test developers solicit candidate feedback and take positive actions based on it, the democratic nature of the assessment is evidenced as a horizontal and 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 assessment.

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 are many things that can go wrong on the day of a test which can interfere in its validity. The air conditioning in a hot and humid room could stop working, forcing candidates to finish the test in uncomfortable physical conditions. An oral examiner could ask a candidate what they did on their last vacation without knowing that someone in the candidate's family died at that time. In both of these not overly extraordinary cases, the candidate's concentration could possibly be affected, thereby modifying his/her performance. This could consequently negatively impact on the candidate's score and provide a false reflection of his or her true ability.

The above examples represent real situations that have taken place during real EXAVER test administrations. As language testers work

with real people in the real world, real (and oftentimes unforeseeable or uncontrollable) problems are likely to continue to occur. However, when it comes to designing and administering tests, there are many things that testers can indeed control, including:

- Being responsible and effective examiners, e.g. giving fair and non-partial treatment to all candidates and following-up with a 'post-exam' report with a list of things that went right and wrong during the test

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.

"When test developers refuse to solicit candidate feedback, or do so without following through on it, the undemocratic nature of the assessment can be seen in terms of the power and control that testers exert over candidates. Conversely, when test developers solicit candidate feedback and take positive actions based on it, the democratic nature of the assessment is evidenced as a horizontal and collaborative process."

- Concern for the test's most important stakeholder: the candidate
- The psychometric quality of the test, i.e. its validity and reliability
- Collective elements of the test system or the operations that are external to the test per se, such as the test registration, orientation, and preparation processes and the way that candidates are treated by examiners and invigilators on the day of the test

References provided on request

Footnotes:

1. Translation: "In my humble opinion". The present article is an abridged version of "Consider the candidate: using test-taker feedback to enhance quality and validity in language testing", published in e-TEALS: An e-journal of Teacher Education and Applied Language Studies 5 (2014): 1-23. ISSN 1647-712X: <http://ler.letras.up.pt/uploads/ficheiros/13086.pdf>

2. As of January 2017, the Mexican minimum wage was approximately 80 pesos per day.
3. 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) pages 228-243.
4. Due to space considerations, findings related to the qualitative phase of the study could not be included but are available from the author at dewing@uv.mx
5. The web-based questionnaire originally appeared in Spanish and was subsequently translated into English. Candidate responses to the questionnaire followed this same pattern.
6. While it could be argued that this premise is based on common sense, it actually mirrored Bachman and Palmer’s similar hypothesis (1996, p.32).
7. The traditional or authoritarian model of education and assessment is still prevalent in many parts of the world today, including in many educational contexts in Mexico.

