



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”¹ ...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

Editorial

Welcome to Issue 63 of the TEASIG Newsletter! Time has simply sped by since our last publication, the Autumn Newsletter, and it is indeed thanks to the contributions of our industrious TEASIG members that we are able to bring out another Newsletter so soon. I would like to thank them, both TESIG members and non-members, for helping to make this possible, and hope that they will inspire more of you to send in articles (which do not need to be long) or take part in a TEASIG Member Spotlight interview. Further details can be found at <https://tea.iatefl.org/newsletter/>

This issue starts off with three contributions submitted by presenters at the IATEFL Conference in Glasgow last April. The first is based on a talk given at the joint ESPSIG/TEASIG Pre-Conference Event (PCE). In this article Elżbieta Zawadowska-Kittel describes a new exam for BA students together with a study which was used to investigate its washback, recalling the power of tests to influence the teaching and learning process that precedes them. The second two articles are based on presentations also given in Glasgow, both at the TEASIG day. Ilya Denisenko's thought-provoking contribution endorses 'authentic assessment' and encourages teachers to see opportunities and benefits it presents, while Susanna Wickes' article provides an interesting account of the introduction of a group oral test in China, also concerned with authentic, meaningful assessment.

In the remaining contributions, Clare Maas shares her experience of the LDF (Learner-Driven Feedback) approach and students' attitudes towards it. The findings are promising, and Clare kindly provides tips for trying out this useful technique. Lynn Williams takes a positive view of assessment in her work as a teacher trainer, reminding us that as assessment is not necessarily a test and sharing some of the strategies she uses with her trainee teachers. To conclude this section, David Ewing Ryan's article makes intriguing reading as he describes a study in Mexico that investigated the attitudes and opinions of the most important, but often neglected, stake-

holders in a language test, that is, the test-takers themselves.

In the final section of the Newsletter, Jo Tomlinson answers outstanding questions in a follow-up to the informative webinar she gave on test development on 24 October 2017. (Check out the next webinar on the TEASIG website!). We round off this issue by turning the TEASIG spotlight on two of our members from different corners of the globe, Christine Coombe, who is based in Dubai, and Gladys Quevedo Camargo from Brasilia.

Thanks again to everyone who contributed to this Newsletter, and for sharing their ideas with the TEASIG community. Very best wishes to you all for the New Year – hoping to hear from you or see you at a TEASIG event in 2018!

Maggi Lussi Bell

TEASIG Joint Editor



Letter from the Coordinators

Dear TEASIG members

As this may well be my last letter to you in the TEASIG Newsletter as TEASIG Coordinator, Neil has asked me to write and say something about myself. Some of you may have read in the latest issue of the IATEFL eBulletin that I have been elected to represent all the 16 IATEFL SIGs on the Board of Trustees. I feel very honoured to have been entrusted with this position, and hope that I can live up to the expectations that my fellow SIG Coordinators have of me.

I will become SIG Representative from April 2018 at the Annual Conference, which means that I will be stepping down as TEASIG Coordinator. I will, however, remain on the TEASIG Committee and continue to work with Maggi on the Newsletter. I have two reasons for wanting very much to continue with this work, connected with TEASIG itself and with publications.

I became a member of IATEFL in 1980 because I wanted to become part of a community of EFL teachers. Almost the only contact with the association and with other IATEFL members was through printed publications such as *Voices*, which was sent by post. I lurked happily in the background of IATEFL for a very long time, following its activities with great interest. However, for a number of personal and professional reasons, such as small children and freelance work, I wasn't able to attend a conference until 2005, although it had been one of my ambitions for some time. I was working as a test developer and my presentation was selected as part of the TEASIG Day, which turned out to be

the start of a happy and fruitful relationship, and one which I now no longer want to forego.

Not only do I belong to the strange breed of test-lovers, but I am a self-designated manic proof-reader. I discovered the joys of editing and proof-reading through my work as a test developer, and my main professional responsibility is now quality control of assessment procedures, which also involves a great deal of editing. I actually find it hard to read anything without mentally editing it or finding printing mistakes, behaviour sometimes described as obsessive!

My interest in publications is what I feel I have brought to TEASIG. TEASIG produces a regular Newsletter and has produced a number of Conference Proceedings as well as contributing to the IATEFL eBulletin and *Voices*. There are many other ideas for publications in the air, such as a TEASIG peer-reviewed journal, a "Best of TEASIG" publication, a TEASIG blog, and more. However, all of these excellent ideas require more time and energy than one or two committee members can provide.

So I am going to end this message with a call for volunteers to work in the area of TEASIG publications, not only with ideas for development but also with a commitment to do something to make TEASIG's services even better for all its members. Please contact me if you wish to be a part of our dynamic and progressive team.

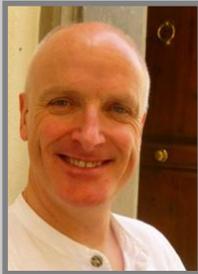
Remember, every little helps!

Wishing you all the very best for 2018,

Judith



The TEASIG committee



Neil Bullock
Joint Coordinator
Webinar Coordinator



Sharon Hartle
Webinar Coordinator



Mehtap Ince
Events Coordinator



Thom Kiddle
Webmaster



Judith Mader
Joint Coordinator
Joint Newsletter Editor



Ceyda Mutlu
Social Media Manager



Dave Allan
Member without Portfolio



Mehvar Ergun Turkan
Membership Officer



Maggi Lussi Bell
Joint Newsletter Editor

To find out more about TEASIG committee members go to
<http://tea.iatefl.org/index.php/about-teasig/teasig-committee/>

People involved in this issue:

Editors: Maggi Lussi Bell, Judith Mader

Photos: TEASIG-CRELLA Conference, Luton 2017

Design and Layout: elc – European Language Competence

Contributions to the TEASIG Newsletter

The IATEFL TEASIG (Testing, Evaluation and Assessment Special Interest Group) Newsletter is published three times a year – firstly, a post-IATEFL Annual Conference issue, between July and September, secondly, at the end of the calendar year, and finally, at the beginning of the calendar year, before the IATEFL Annual Conference. All issues are in digital form.

The next submission dates for articles and advertisements are 31 January 2018 and 31 May 2018.

Contributors do not need to be members of IATEFL or TEASIG, however contributions from TEASIG and IATEFL members, and on matters of interest to TEASIG members, will be given priority.

TEASIG reserves the right to accept or reject articles as it deems necessary. If an article is accepted for publication, it may be subject to editing and alterations may be made in titles, headings, length and other aspects. Minor editorial changes in the text may be made for reasons of space, style, clarity, acceptability and correctness of language. If more than minor changes are considered necessary, the author(s) will be consulted by the Editors of the Newsletter for their approval.

In some cases, articles may not be included in the next published issue of the TEASIG Newsletter, but reserved for a later issue. This will be at the sole discretion of the Newsletter Editors.

Guidelines for contributors

- Articles do not need to be long. They may be as short as 600 words but should generally be no longer than 2000 words.
- Photographs, graphics or diagrams should be sent as separate .jpg files. Please mark clearly in the text where they should be placed and ensure you have permission for reproduction of any such items. If in doubt, the Newsletter Editors may ask a contributor to verify the source and authorisation of any such material.
- Only 5 key references should be given (where required). Other references should be available if requested.
- Pieces should not have been published or be undergoing consideration for publication elsewhere. If this is the case, please include information on where and when the article has previously been or will be published, and ensure you have the right to republish from the Editor of the other publication.

More information on how you can contribute to the TEASIG newsletter can be found on <http://tea.iatefl.org/index.php/newsletter/>

Washback of Practical English examinations on the process of teaching and learning



Elżbieta Zawadowska-Kittel

is currently working at the University of Social Sciences in Warsaw. Her PhD thesis centred on the washback of a school leaving exam in English on the teaching and learning process. Her research interests include testing and assessment, innovations in various fields of methodology and the concept of learning outcomes.

This paper is based on a presentation given at the ESPSIG TEASIG Pre-Conference Event at the IATEFL Conference in Glasgow in April 2017.

1. Exam in Practical English as a reflection of the implementation of the National Qualifications Framework

The introduction of a new exam was connected with the implementation of the National Qualifications Framework (NQF) and new learning outcomes for BA students at the college where the research was conducted. Before the change to the new exam, a traditional exam was used, composed of two parts: written and oral. The written part included reading comprehension, listening comprehension, writing and grammar competence. The oral exam was taken in pairs and consisted of a discussion on a given subject. The format of the examination matched the way the classes at the college were formed, as Practical English classes were divided into conversation combined with reading skills, audio-visual classes, grammar, and writing.

The introduction of the NQF catalysed the changes in the process of teaching and learning. The classes devoted to teaching Practical English (PE) were divided into blocks of integrated skills and, additionally, into classes devoted to productive skills. The new division is intended to facilitate and accelerate the achievement of learning outcomes, especially as far as speaking and writing are concerned.

1.1. Washback of exams on the process of teaching and learning

The research that has been conducted so far (A. Hughes, J. Alderson, Li Cheng, E. Zawadowska-Kittel) shows that it is the manner of testing that changes the process of teaching and learning both as far as priorities and techniques are concerned.

This concept is also reflected in the publication "How to prepare educational programmes in accordance with the NQF" by Andrzej Kraśniewski (2011) who writes: "Proper planning of the methods of verification of learning outcomes may to a large extent decide what outcomes have been, in fact, achieved by the student. This results from the fact that a large part and sometimes the majority of students adapts their way of learning to the content and form of tests. Thus, these are the methods of verifications of learning outcomes which determine what students know to a larger extent than the teaching methods."

The new exam was designed based on this view. Summing up the different opinions on and definitions of washback, washback may be defined as the influence of exams, intentional or unintentional, both positive and negative, on the learning and teaching process, as this is widely understood, and on all its participants as well as on the entire educational system (E. Zawadowska

-Kittel 2013). In the present paper, two major aspects of the washback of the PE exam will be presented. It is worth adding that in order to achieve positive washback, students need to be acquainted with the purpose of the exam and regard its results as credible and fair, and they also need to receive a detailed report on their exam results, rather than only a mark or a percentage score. Additionally, the tasks should be authentic.

2. The concept of the PE exam

The new concept of the exam stresses the role of productive skills in the teaching and learning process and for this reason both speaking skills and writing skills are tested twice. In case of writing skills, the progress students have made during the whole semester is assessed (by means of a portfolio), which supports both autonomy and systematic learning and also helps to increase motivation for learning, as frequent feedback is given to students. Also, critical thinking ability and the ability to express opinions in writing is tested (with reference to reading).

Speaking skills are also assessed twice – first, in connection with listening skills to increase the authenticity of the test, as in real life speaking is most often connected with listening, and then for the second time in connection with a PowerPoint presentation and dis-

cussion of a selected subject (as the ability to speak in public is often expected of the students by prospective employers).

The new format of the examination makes use of modern teaching techniques – film, IT and also (thanks to the portfolio and presentations) reduces to some extent the risk of unreliable testing as, if only one performance is assessed, the quality of this may be affected by, for instance, stress. The new format of the exam has been evaluated both by students and academic teachers in specially constructed questionnaires.

“The new concept of the exam stresses the role of productive skills in the teaching and learning process and for this reason both speaking skills and writing skills are tested twice.”

The results of these questionnaires will be used to modify both the concept of the exam and the way it is graded. The conclusions may serve to change the concept of teaching through the methods of verification of teaching concepts.

3. The study

The aim of the study is to investigate the washback of a new exam

on students and teachers, and find out their opinion of the test format.

3.1. The questionnaire

The survey questionnaire was developed specifically for the study and administered in Polish. Nineteen of the questions were closed questions and one open-ended, the aim of which was to allow the respondents to voice their opinions of the new exam in a more precise way. The surveys of teachers and students were as identical as possible in terms of their content.

3.2. Profile of the respondents

A total of 70 valid responses from the students of philology at a private university and 10 from their teachers were received from those respondents who volunteered to take part in the study. All the students are 1st year BA students; 70% of them study English as their major. More than half of the teachers teach English and almost all are interested in testing issues. Their teaching experience ranges from 2 to 10 years. However, no correlation between their opinions and experience has been found.

4. Discussion of results

A comparison of teachers' and students' most significant answers to the questions in the survey is presented in Table 1.

Table 1 Comparison of opinions

Question/problem	Teachers		Students	
	YES (%)	NO (%)	YES (%)	NO (%)
All tests influence teaching/learning process	90	10	80	20
Class activities perceived as most frequent	Activity	%	Activity	%
	Watching and discussing videos	20	Watching and discussing videos	25
	PowerPoint presentations	20	PowerPoint presentations	20
	Writing tasks	20	Writing tasks	15
	Discussion in groups	15	Discussion in groups	5
	Short oral tasks	15	Short oral tasks	25
	Grammar exercises	10	Grammar exercises	10
Assessment of innovative methods of verification	Method	Points (1-6)	Method	Points (1-6)
	PowerPoint	4	PowerPoint	5
	Portfolio	4	Portfolio	5
	Audiovisual tasks	5	Audiovisual tasks	4
The use of L1	Activity	%	Activity	%
	Explaining class and exam procedures	80	Asking questions about class and exam procedures	100
	Explaining grammar and vocabulary	90	Asking about grammar and vocabulary	80
	Giving homework	50	Asking about homework	80
	Giving instructions	50	Reacting to instructions	70
Progress estimate table	Skill/microskill	Teachers' estimate	Students' estimate	Average final actual grade
	Reading	4.5	3.83	3
	Listening	4.3	3.83	3
	Speaking	3.9	3.5	3
	Writing	3.8	3.3	3.5
	Vocabulary	4.4	3.76	X
	Grammar	3.5	3.3	X
Advantages of the examination	Teachers	Students		
	<ul style="list-style-type: none"> • Methods of verification reflect learning outcomes • More just • Forces students to develop productive skills • Enhances autonomy • Increases motivation • Tasks authentic • Possibility of evaluation (monitoring) 	<ul style="list-style-type: none"> • Shows progress better • Less stressful • Contemporary • Close to life • Focuses on communication • Easier to pass than the previous one • More just 		
Disadvantages of the examination	Teachers	Students		
	<ul style="list-style-type: none"> • No grammar • Checking portfolios time consuming • Accuracy not stressed in enough in the criteria • Requires a lot of explanations 	<ul style="list-style-type: none"> • Lasts long • Requires a lot of work and preparation • A bit complicated 		

lar or identical to the examination tasks: PowerPoint presentations seem to be the activity perceived as even more frequently used by students than by the teachers. It is noteworthy that new methods of testing received high ratings from both teachers and students.

Responses to the question concerning the use of L1 may suggest that the format of the exam is too complicated and leads to overuse of L1 on the part of the teachers, who explain the format of the exam and its rules in L1 and this, in turn, provokes students to use L1 as well. It must be noted, however, that talking about the exam is not the only classroom situation in which L1 is used, so it is rather a matter of teachers' habits and these may gradually change.

The teachers are of the opinion that learners made the most significant progress in receptive skills: listening and reading, as well as vocabulary, which is directly connected with their estimate of their achievements in these skills. Progress in speaking and writing is less spectacular, even though the aim of the changes was the development of productive skills. The least progress was achieved in grammar competence, which most probably results from the fact that grammar is not sepa-

The majority of both teachers and students realize that the manner of testing influences the process of teaching and learning, which is reflected in the activities most frequently used during the classes.

The table does not include activities perceived as most frequent by less than 10% of the respondents (listening to CDs, projects, translations). The table shows that teachers focus on activities that are simi-

rately tested on the exam. The students share the teachers' opinion that they made the most significant progress in receptive skills, listening and reading, as well as in vocabulary acquisition. Similarly to the teachers, they feel they made less progress in speaking, writing and in grammar competence.

It is interesting that teachers consider students' progress to be higher than the students themselves do, whose estimate is closer to the actual results of the exam. It seems likely that teachers observe students' achievements from the perspective of the whole academic year and students see the quality of particular performances or the results of individual tests. The actual grades are lowered by unsatisfactory marks.

Some discrepancies exist in the teachers' and students' views on the pros and cons of the exam. Teachers point out that portfolios are time-consuming and in the students' opinion, make the exam easier to pass. Teachers complain about the fact that grammar is not tested and that accuracy is not sufficiently stressed in the criteria, whereas students approve of the fact that the exam focuses on communication.

Consequently, teachers and students share the view that exam tasks resemble situations encoun-

tered in real life and may be regarded as authentic. They are also of the opinion that the exam is valid and more just.

Conclusions

As it turns out, the concept of washback of important exams on the teaching and learning process is

"It is interesting that teachers consider students' progress to be higher than the students themselves do, whose estimate is closer to the actual results of the exam."

no less true in case of the PE exam that has been discussed here. The most important positive washback is that teachers focus more on productive skills, as in class they most frequently use activities resembling the examination tasks. However, progress in those skills is not automatically higher than in receptive skills, an aspect which needs to be improved. Insufficient progress may result from the overuse of L1 by teachers, something which needs to be changed. Neglecting accuracy seems to be another important problem, which is why including grammar tasks in a new exam is postulated by the teachers. Most probably, the students will not approve of that solution, as they value most tasks such as the portfolio and

presentation, which seem to them far safer than traditional tests. On the whole, the new exam is approved of both by the students and the teachers as more valid, reliable and contemporary. The teachers additionally value the possibility of evaluation of the exam. For this reason a new study is planned after the changes in the exam are introduced.

References

- Alderson, J.C./ Wall, D (1993), *Does washback exist?* (w:) "Applied Linguistics" 14, 115-129.
- Cheng L./ Watanabe, Y./ Curtis A (2004). *Washback in language testing. Research context and methods.* Mahwa.
- Hughes, A (1994). *Testing for language teachers.* Cambridge.
- Kraśniewski, A (2011). *Jak przygotować programy kształcenia zgodnie z wymaganiami Krajowych Ram Kwalifikacji dla Szkolnictwa Wyższego?* Warszawa.
- Zawadowska-Kittel, E. (2013). *Nowa matura z języków obcych: szanse i zagrożenia. Efekt zwrotny egzaminu z języka angielskiego.* Piaseczno.



IATEFL Testing, Evaluation and Assessment
Special Interest Group (TEASIG)

Pre-Conference Event

9 April 2018

IATEFL Annual Conference, Brighton

Assessing Listening

- Why do we want to test listening?
- What listening skills do we want to test and how?
- How can we better identify and assess these skills?
- Where can we source appropriate material? Should it be scripted or taken from the real world?
- What is the importance of statistics in the assessment of listening skills?

In this 2-part PCE we will be attempting to take the lid off this mysterious world with the help of experts and practitioners in this domain.

The PCE is aimed at all teachers, whether they are new to this aspect of testing, evaluation and assessment or wish to refresh their knowledge and put it to practical use. There will also be room for questions to experts and discussion of best practice.

Morning – Plenary Speakers

John Field
The University of Bedfordshire, UK

Cognitive processes in listening

Sheila Thorn
The Listening Business

A new matrix for the testing of listening

Rita Green
Test Development Training & Analysis Ltd

Statistical evidence to support assessment of listening

Afternoon – Hands-on workshops

led by NILE and LT123 with opportunities to match the theory and ideas to your world

For more information and registration

<https://tea.iatefl.org/upcoming-teasig-events/>

https://secure.iatefl.org/registration/conf_reg_login.php

This event is being generously sponsored by Cambridge English Language Assessment.



Education or evaluation?



Ilya Denisenko

Ilya Denisenko is Head of the Languages Department at the State Academic University for Humanities (SAUH Moscow, Russia) and a founder of Authentic ELT Association. This year he has completed a study on assessment and given a talk at the IATEFL Conference.

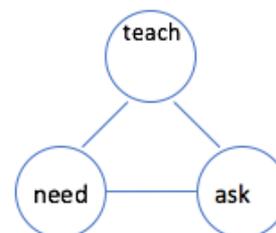
This paper is based on a presentation given at the TEASIG Day at the IATEFL Conference in Glasgow in April 2017.

Imagine that a graduate English exam is obligatory for all your students and it is taking place tomorrow. Will they pass? Imagine your brightest students, will they have any problems? Think of the student that worries you most. Do you think they will be lucky enough to get through the test with mostly correct answers? Is it the speaking and writing parts that concern you? Do you agree that what we teach, what students need, and what is being tested are the same thing? When a student gets top marks in an exam, does it mean they are well prepared for life and their career?

Well, we still have many questions about evaluation. Most teachers usually agree that students should be taught relevant skills but others may claim that they themselves lack the tools for developing such skills. Normally, an EL teacher has to upgrade students' grammar and vocabulary on different topics as well as develop students' skills. Furthermore, subsequent examinations will add some more uncertainty, so the task formats used in exams should be explained and thoroughly instilled. Maybe all this critical thinking, IT and communicative competence should be left to somebody else because we don't have time for it?

Can you put an equals sign between

what you teach, what your students need, and what the exam will test?



If you consider these *competences* important, you no doubt spend considerable time and effort organising practice of them. You might be struggling with the syllabus, trying to find enough time and searching for sufficient materials. You do it simply because there is a real world outside the classroom and correct grammar alone won't suffice. We, as teachers, just need to suggest something more applicable to office or freelance work in business. School is not just a place where knowledge and skills are learned. Every subject should contribute to cross-subject competence building – 'should', but doesn't 'have to'. However, there are still things that are a must: the syllabus, educational programme, year plan, and even the teacher's book guide.

So, will an EL teacher risk the limited hours available on building up language skills to develop something that neither a test nor an exam can really assess? What if they do find the extra time?

English as a subject makes it possible to speak about anything from ancient history to space flight; it doesn't have content restrictions (that depends on the cultural traditions of your country, of course). This means that an EL lesson can be built on various contexts, related to interests, experience and the professional or future educational needs of the students. This objective makes an English lesson unique in that it prepares students for real-life problems much better than anything else in the school curriculum.

An English lesson, moreover, is an ideal time and place to build up the crucial skills needed today. A teacher doesn't have to reduce or cancel traditional grammar or vocabulary practice (if they can't do without it) to spare some time for another task. There is another solution.

There is no need to go too deeply into the calculations: the largest amount of time spent in every classroom around the world is used for evaluation. The time EL teachers spend on testing can be used more efficiently if it's used for teaching purposes as well.

Assessment is a valuable part of the educational process that is often not used enough. Many scholars have agreed that assessment has a dou-

ble-duty and works as a scaling and a teaching tool. However, many practicing teachers find it takes too much effort to spend a lot of time evaluating their students, at least formatively. In addition, most teachers don't try to develop testing materials themselves. A well-organised assessment system could solve such problems.

“Another vital thing is to make performance meaningful for students, which can be done through the context of a task. For this reason, tasks should be correlated with the students' potential knowledge and provide them with relevant feedback.”

Imagine such a system exists. What are its characteristics? First of all, it should be built on students' *performance*. This means that students need to be evaluated on how they deal with a real practical task. In this case, observation of the students' work shows real acquisition of competences.

Another vital thing is to make performance *meaningful* for students, which can be done through the context of a task. For this reason, tasks should be correlated with the stu-

dents' potential knowledge and provide them with relevant feedback. Meaningful means interesting, connected with their experience, as well as thought and creativity provoking.

Such meaningful contexts bolster motivation. When a student can apply their own personal experience, it provides scope for creativity and readiness to work in the proposed setting.

I'd like to illustrate this with one example. A school boy was considered to be a poorly performing pupil by all his teachers. Even those who didn't actually teach him were aware of his low potential. He obviously behaved in a way that was consistent with how he was seen, and his teachers usually didn't expect much of him. Yet, on an English project about certain tourist attractions he wanted to see in his area, he really did his best. It turned out that he was into extreme sports, for which there are many English words describing tricks, moves and equipment. So, English lessons can reach everyone. Maths or native language teachers spend a lot of time with students but they don't normally discuss personal things with them and, if they do, they are unable to apply this in the lessons. English lessons, however, are more suited to this.

A task can be *multi or cross-subject* because, in real life, we rarely come across isolated tasks – think of conversations, sending e-mails, working on projects, travelling and so on. We base our work on all our skills and experience, so teachers should try to set the same conditions for their students. Knowing what topics are being discussed in history or maths at the same time can help in developing such tasks.

Although assessment is often stressful for students, stress doesn't often have negative results. A stressful situation also makes it more real, that is, closer to true foreign language communication.

By combining the above ideas, we can turn assessment into *authentic* assessment, that is, evaluation in a situation where foreign language competence is applied when it is essential and natural.

There are several ways to make assessment practical. One that can be controlled at all stages is a project. A project calls for students to perform a variety of tasks and check different goals. There can be several stages, each with different aims and objectives.

The first stage works as an introduction to the topic. Written and audio tasks are given. If there is a need, reading and listening skills can also

be tested here. The materials set the context for all subsequent work.

At the second stage, students receive the project task, which usually involves some research and productive parts. A study plan is either given to the students or they develop one themselves, with some points as questions or statements. Then they create a presentation. At the next stage, they give a presentation.

Depending on the objectives, the teacher can add different tasks to the project: group research, collective oral presentation, written task, poster, or letter. The main aim of this stage is to see the language competence the students have truly acquired, and analyse their performance. The teacher plays the role of interlocutor and stimulates discussion around the presentation, asking additional questions and bringing up points on some details that can be upgraded and improved. This provides considerable speaking practice and produces relevant information about skills levels.

The feedback stage is crucial for authentic assessment. The teacher can identify common mistakes more easily

when viewing students working together.

To sum up, I'd like to emphasise that, once the teacher decides to develop and assess competence acquisition holistically, authentic assessment features can be very helpful. Authentic assessment can be beneficial for both teachers and students. The former obtain an opportunity to see to what extent their students have really acquired various competences, and to draw conclusions about the efficacy of their teaching. The latter can try out their skills under natural conditions, and understand their weaknesses and abilities before leaving the safety of the classroom.

Evaluation is a treasure chest when it comes to teaching languages, and teachers should be familiar with it and use it. Multiple choice tests should be left for general testing and other less fortunate subjects. The EL teacher should try to make the most of evaluation, ensuring every assessment session is a practical one.



Join us as an
**Early Career Teacher
Member**

- Are you a newly-qualified English language teacher?
- Have you been in the profession for less than 5 years?
- Not previously a member of IATEFL?

If the answer is yes, you can benefit from **2 year membership for the price of 1!**

Find out more and join IATEFL at: members.iatefl.org

What you can do for TEASIG

IATEFL TEASIG members all have one thing in common: their interest in issues connected with testing, evaluation, and assessment. Some members are experts in the field and contribute to the SIG by sharing their expertise with other members in various ways: holding webinars, speaking at conferences and events, and writing for the Newsletter. As IATEFL is a teachers' association, many members are teachers, some just starting on their teaching career, and others a long way into it. A high number of teachers have a great deal of experience of testing, while not considering themselves experts, whereas others are keen to get into what they feel is an important and dynamic area of teaching EFL. Many teachers hope to learn something from experts and from other teachers, and to have the opportunity to ask questions and share their own ideas. Whichever group you belong to, there are many ways in which you can contribute to TEASIG and our efforts to provide effective services to all our members. Here are a few ideas.

Newsletter

Contribute

- by answering member spotlight questions and giving all members an idea of who else is in the SIG
- by sending us a short answer to the Mosaic question for the next issue
- by reporting on any (however small) research into assessment or testing you have done
- by reporting on an assessment or testing event or conference you have attended

Help the Editors by

- approaching potential contributors
- approaching potential advertisers
- proof-reading
- giving feedback on the Newsletter (what you like, what you would like to see, and what you don't like – but please provide an alternative where possible!)

Other publications

Let us know what other publications we could offer.

Website

Send

- information on external events, seminars, workshops, and conferences (strictly commercial activities, including fee-bearing courses, will not be advertised free of charge). A decision on whether to include any event is at the discretion of the SIG Co-ordinator(s).
- feedback on the website (what you like, would like to see, and don't like)

Webinars

Make suggestions for speakers and formats.

Events

- Help to organize an event in your area.
- Provide technical support with live-streaming and recording of sessions.
- Act as a helper at a TEASIG event.

Social media

- Encourage colleagues to join TEASIG social media groups.
- Start and contribute to discussions.

Maximising collaboration in oral English testing: an example from China

Susanna Wickes

began teaching in her home town of Edinburgh in 2009. Since then she has taught in Australia, India, Ireland and - most extensively - China. During her years in this country she has developed research interests in assessment, learner autonomy, and collaborative learning, particularly in the university context. She currently teaches EAP at the University of Nottingham Ningbo China.

susannawickes@gmail.com

This paper is based on a presentation given at the TEASIG Day at the IATEFL Conference in Glasgow in April 2017.

Introduction

The teaching and assessment of spoken English has arguably been neglected by China's exam-focused education culture. For Chinese students, becoming confident users of English offers myriad opportunities, but since 'success' so often merely equates to passing exams, many of these young people have little conception of lifelong learning. I strived to change this attitude with my students by designing a speaking test which aimed to be both useful and meaningful: a spontaneous group conversation activity. This article begins with a few background details and then explains the test's design and marking procedures.

The learners and the course

My class, at a public university in Inner Mongolia, China, consisted of fifteen Mandarin L1 learners who all struggled with oral English to some degree. As English majors, their compulsory modules included grammar, listening, reading, writing and translation, mostly taught by native Chinese teachers in preparation for national exams like TEM (Test for English Majors) and CET (College English Test). Since oral fluency is not integral to these exams, speaking classes are normally left to foreign teachers, whose interactive teaching methods often contradict the traditional teacher-centred curriculum. Because of these factors, most of my learners were anxious about speaking spontaneously in class and often memorised and recited texts during presentations or

other activities, making it challenging to observe or obtain samples of 'natural' speech to monitor, and ultimately assess, their progress.

Developing the confidence, knowledge and skills required for engaging in informal discussion was the central learning objective of our 18-week course. Our weekly lessons focused on various themes – often suggested by the learners – and included related vocabulary and grammatical structures with regular pair and group discussion activities to maximise speaking opportunities as well as build knowledge and skills through active practice. To foster L2 confidence the focus was shifted from accuracy to fluency and I encouraged the learners to speak as much as possible without worrying about mistakes. They had good receptive knowledge of grammar, so I knew with practice they could gradually begin to notice and correct their errors.

Test design

Conducting an end-of-term achievement exam was a course requirement, but I was given the freedom to design, administer and mark the test. A crucial consideration for non-traditional assessment is that tests should be "integrated with the goals of the curriculum and ... have a constructive relationship with teaching and learning" (McNamara, 2000, p.7). A 'conversation' test activity clearly reflected the course objective and the communicative, fluency-based classroom content, thus firmly linking the test to the learning

goal. The test results then enabled the learners to gauge their progress, diagnose their needs, and set new goals.

The individual competences required for conversation and informal discussion were adapted from the CEFR (Council of Europe, 2001) and are shown below with a description of the ability level I expected my learners to reach in each category. These competences were taught and practised throughout the term and provided the basis for the marking rubrics.

- Linguistic competences
 - Phonological
 - Lexical
 - Grammatical
 - Semantic

→ *The learner can use appropriate and intelligible words, expressions and structures, with clear, intelligible pronunciation, to express his/her ideas.*
- Pragmatic competences
 - Discourse competence
 - Functional competence

→ *The learner can use appropriate techniques to initiate, engage in and maintain conversation; can ask questions, agree, disagree, etc.; can give descriptions, explanations, etc. where appropriate, and can perform all of these in a confident manner without excessive pausing or hesitation.*

Test method and content

Although conversation-style tests can take place between two people – a test taker and an assessor/ interlocutor or two test takers – I organised my learners into groups of three as talking in groups can reduce anxiety (He & Dai, 2006). Peer interaction also promotes more realistic communication than assessor-examinee conversation as learners are in an “equal power position” (Luoma, 2004, p.187) and can demonstrate more interactive functions than simply answering an interlocutor’s questions. This enables the assessor to observe, rather than

“Although conversation-style tests can take place between two people I organised my learners into groups of three as talking in groups can reduce anxiety.”

participate in the activity and therefore pay more attention. Additionally, group assessment is practical and efficient; the fifteen test takers worked in five groups, allowing all the tests to be completed during one two-hour lesson.

Open-ended conversation content was prompted using topic cards. Before the test week, six broad themes covered in the course were chosen by student vote and written on the backs of six cards. During the test, one card was randomly selected to inspire a short conversation. The topics – *personality, technology,*

relationships, environment, travel and education – were all familiar to the learners and were considered equally ‘difficult’ as they could all be discussed at any level from basic and concrete to more abstract and complex. Knowing the six topics in advance allowed learners to practise and revise relevant vocabulary and grammar, thus reducing nerves and maximising fairness, but not revealing the specific topic until the test ensured that learners could not ‘cheat’ by scripting and memorising their conversations.

Test administration

After explaining and practising the assessment format, the test was administered during class time in the second-last week of term, leaving the final week free for a peer/self-evaluation session. I initiated a short, unassessed ‘warm up’ to help the test takers relax, then a group member was asked to randomly select one of the six cards, which provided the topic for the conversation. I started the recorder and set a timer, which beeped after seven minutes to let the group know they had reached the minimum limit for completing the task. This method allowed the learners to end their discussion when and how they wished, thus producing a more representative, uncontrived conversation sample. The timer was set to beep again after ten minutes – the upper limit – to tell the group to round things off if they were still talking. Notes were not allowed before or during the conversation and test takers were

not given time to prepare. This was a deliberate decision to maximise authenticity, as ‘real’ conversations are spontaneous. I only intervened if the conversation had completely halted or if the learners directly asked me a question, which they could choose to do.

Teacher marking

In authentic tests it is vital to involve learners in the assessment process as it enables them to develop awareness and take responsibility for their own learning (O’Malley & Pierce, 1996), however peer and self-assessment can only realistically supplement the expert evaluation of the teacher (Luoma, 2004). For these reasons, I made marking a collaborative endeavour, with an analytic rating scale for teacher assessment and a separate holistic scale for peer and self-assessment, the latter devised by the learners. The teacher’s scale was written to rate and score individual test takers’ performances in the conversations, and was adapted from the communicative language competences mentioned earlier:

1. Linguistic comprehensibility
2. Phonological comprehensibility
3. Fluency
4. Conversational interaction
5. Evidence of learning

‘Linguistic comprehensibility’ was included to discourage test takers from overly focusing on accuracy and thereby threatening fluency. It was defined as being able to use

appropriate and intelligible words and construct intelligible phrases and sentences. Learners were marked down if their errors impeded understanding, but mistakes and slips were disregarded. Pronunciation was also defined in terms of intelligibility, though a high score in this category also involved good control of intonation and syllable stress. Fluency was defined according to the learning goal of speaking

“To maximise learner involvement in this process, the class was responsible for creating the scale during one of their lessons.”

spontaneously and confidently without unnatural pauses and excessive hesitation. Interaction included turn-taking, co-operating and participation. Finally, ‘evidence of learning’ was included to assess learners’ use and understanding of taught content, rated according to its frequency, accuracy and appropriate usage.

I adapted five descriptors from the CEFR and wrote them as five-point scales in as positive wording as possible.

During the test, I simplified this into a basic 5x5 grid so I could make quick notes after each conversation and then use these with the full rubric when listening to the recordings.

Peer and self-assessment

Another rubric was created for peer and self-evaluation, and to give a group score. This five-point holistic scale made for an efficient, straightforward marking process that promoted discussion and negotiation amongst learners. To maximise learner involvement in this process, the class was responsible for creating the scale during one of their lessons. In small groups, they were asked write a short description of what they thought constituted an ‘excellent’ and a ‘not so good’ performance, and eventually, after some reporting and comparing, we wrote a five-point scale that

	1	2	3	4	5
1. Linguistic comprehensibility	Limited control of vocabulary and structures, which impedes understanding	Can use vocabulary and simple structures mostly correctly but still needs to rephrase and repeat to be understood	Can select words and structures to communicate with reasonable accuracy; can be generally understood	Has good control of vocabulary and structures; can be clearly understood most of the time despite making some mistakes	Shows high degree of lexical and grammatical control and has no problems being understood; any slips are minor
2. Phonological comprehensibility	Words and utterances can be understood with some effort; repetition may be required	Pronunciation is generally comprehensible despite a noticeable foreign accent	Pronunciation is clearly intelligible, though intonation is slightly unnatural	Pronunciation is clear and easy to understand; intonation is natural	Clear, natural and fully comprehensible pronunciation; intonation is varied appropriately, including syllable stress
3. Fluency	Phrases are short and isolated; there are a lot of long pauses and some repairing is required	Phrases are generally short but well-constructed; there are regular, noticeable pauses	Can speak with relative ease and without help; pauses and hesitation short and less noticeable	Can speak spontaneously and confidently with a natural tempo despite some hesitation	Can speak spontaneously and confidently at length; uses a smooth flow of language; any hesitation is natural and well-controlled
4. Conversational interaction	Can make short and basic contributions when asked directly	Can handle short exchanges and provide some input despite lacking confidence	Can participate with some confidence although has difficulty maintaining the conversation	Can converse confidently in a clearly participatory way, maintaining the conversation well	Can converse with ease, participating naturally, confidently and appropriately; maintains the conversation effortlessly
5. Evidence of learning	No use of items taught in class	Uses a few items, though does not seem familiar with their uses	Uses a few items accurately and appropriately	Uses a range of items though not entirely accurately or appropriately	Uses a range of items, demonstrating clear understanding

Teacher’s marking rubric (adapted from CEFR)

everyone agreed on.

A week after the test, the learners returned to their groups and I played the recordings. Each group agreed on a number from one to five to score each recording, including their own, and wrote some constructive comments on an anonymous evaluation sheet. If a number could not be agreed on, group members could award a half mark such as 3.5. I recorded the scores

1	2	3	4	5
The speakers sound uncomfortable and the conversation isn't very natural. There are some problems understanding each other and keeping the conversation going. There are a lot of long pauses.	The speakers sound a bit nervous but they are more or less able to maintain a basic conversation with some long pauses. Some group members talk much more than others so participation is a bit unequal.	The speakers sound generally comfortable and the conversation is quite natural, although there is still some pausing and hesitation. The share of participation is generally equal most of the time.	The speakers sound comfortable and have no problems maintaining the conversation. There's some hesitation but it doesn't affect the flow of speech. There is a mostly equal share of participation.	The speakers sound very confident and the conversation sounds very natural. All the speakers participate actively and interact well with each other. The conversation flows effortlessly.

Peer and self-assessment rubric

and calculated an average for each group, adding it to all the members of that group on top of their individual scores. At the end of class the evaluation sheets were distributed to the learners to be read and discussed.

Group marks encourage learners to work together and take joint responsibility for their performance while individual marks make learners accountable for their own contributions (O'Malley & Pierce, 1996). Thus, allocating both maximises the benefits of collaborative learning yet remains fair for everyone.

A note on validity

Success in the test provided me and

learners with evidence that the learning goal had been achieved; the students were able to have spontaneous English conversations. However, the nature of open-ended group conversation tests poses some validity issues. Assessing learners in groups introduces numerous variables, making it difficult to isolate individual performances. While this is problematic, it must be remembered that conversation is

inherently collaborative and context bound, and so isolating speech and test items

would invalidate the whole test.

Another issue is that learners inevitably act differently when being assessed (Luoma, 2004), most likely due to test anxiety and the desire to demonstrate their best performance (He & Dai, 2006). Of course neither of these would be present in 'real world' conversations. To minimise such threats I did everything possible to ensure my learners felt at ease, including giving them time to prepare and practice and allowing them to choose their own groups.

Conclusion

This article has outlined a simple yet innovative speaking test, designed to help a class of Chinese university

students develop the skills, knowledge and confidence to engage in English language conversations. The classroom content that provided the basis of the test facilitated a meaningful and valuable learning experience, and the collaborative evaluation process promoted inter-peer dialogue and encouraged learners to take control of their learning. Though group tests introduce inevitable threats to validity I took action to minimise these, creating a test that I contend is authentic, fair and useful.

References

- Council of Europe. 2001. *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.
- He, Lianzhen & Dai, Ying. 2006. A corpus-based investigation into the validity of the CET-SET group discussion. *Language Testing*, 23(3): 370-401.
- Luoma, Sari. 2004. *Assessing speaking*. Cambridge: Cambridge University Press.
- McNamara, Tim. 2000. *Language testing*. Oxford: Oxford University Press.
- O'Malley, Michael & Pierce, Lorraine V. 1996. *Authentic assessment for English language learners: practical approaches for teachers*. New York: Longman.

Receptivity to Learner-Driven Feedback



Clare Maas

holds post-graduate qualifications from the University of Wales and Trinity College London. Before moving into tertiary education, she taught English at German grammar schools, and English for Specific Purposes at several language academies in the UK and Germany. Her professional interests include EAP, materials development, and CPD for teachers. Her most recent research has been into feedback on L2 writing, especially in EAP and higher education. She also blogs at ClaresELTCompendium.wordpress.com and is a member of the team behind ELTResearchBites.com.
cmmaas@uni-trier.de

In a recent journal article, I presented my initial research into students' openness to a new approach to giving feedback on written work – Learner-Driven Feedback (LDF). Here is the reference for the full article (available openly as Editor's Choice on the ELTJ website): Maas, C., 'Receptivity to Learner-Driven Feedback in EAP', *ELT Journal*, 71/2 (2017), pp. 127-140.

The LDF approach understands that learners often feel happier receiving feedback from a teacher than through peer review or self-editing. In LDF, learners 'drive' feedback by asking questions about their work, and then re-draft based on this feedback. The teacher gives the feedback, but the learners decide how and on what they receive comments: they can choose between various formats (e.g. hand-written, email, audio recording), and are required to pose questions to which the teacher responds (e.g. about grammar, vocabulary/register, referencing, text structure).

Some Background

I devised this approach by combining insights from other fairly recent studies on giving feedback on L2 writing, particularly in EAP contexts. I looked into two key areas:

- Making feedback more dialogic

- Using technology to deliver feedback

With regard to involving students in a feedback dialogue, Bloxham and Campbell (2010) trialled 'Interactive Coversheets', which students use to pose questions about their essays when submitting drafts. Bloxham and Campbell found that most students in their study were grateful to receive feedback that was more individualised than usual, which then also motivated them to evaluate their own work more thoroughly. Moreover, the teachers involved reported being able to provide feedback on drafts more quickly, by only commenting on areas highlighted in students' questions.

Campbell and Schumm-Fauster (2013) also required their EAP students to pose feedback questions on their work when submitting essay drafts. Here, students noted their questions in footnotes or in the text's margins. The students were surveyed and reported finding the feedback personal and motivating, and effective in meeting their individual needs regarding academic writing.

Studies into using technology to deliver feedback have explored, for example, using audio recordings or emails. Johanson (1999) lists some advantages of recording audio feed-

back for ESL students, which include its possibly saving teachers time, making comments clearer through intonation, and feeling more personal to students. These ideas are supported by studies outside ELT, such as Brearley and Cullen (2012), who found three minutes of audio recording could include ~500 words of feedback and did not take more time than marking essays by hand.

Also investigating digital feedback, Farshi and Safa (2015) compared grades of EFL students who received hand-written, emailed, or no corrective feedback on written tasks for one term, and found that emailed feedback led to significantly greater improvement on post-test grades than hand-written feedback.

My Research

To put this summary in context, readers need to know that I used LDF on three drafts of an essay that my undergraduate EAP students (B2 level on CEFR) wrote during one semester at a German university. The students could:

- choose between various (not mutually exclusive!) modes of feedback: in-text corrections, correction symbols, handwritten feedback, email, audio recording, or face-to-face consultation.

- ask questions about any aspect of their writing, e.g. text structure, referencing, vocabulary, grammar, etc.
- pose both specific and general questions.
- include their questions as footnotes or in margins, or at the end of the text.

At the end of semester, I surveyed the 40 students to discover their attitudes towards the LDF approach.

“Students’ comments to open-ended survey questions highlighted some reasons why audio recordings and email were so popular, for example, mentioning that feedback was more detailed and timely, and felt more individualized than the feedback they were used to.”

The questionnaire included questions on students’ perception of the usability of various delivery modes, the perceived effectiveness of LDF for improving their language accuracy and academic skills related to essays, and any problems they experienced with LDF.

Some Findings

The delivery modes most students requested were audio recording (by 67% of students) and email (60%). The figures for handwritten comments (13%) and in-text corrections (20%) were much lower, which seems to show that students are happy to move away from ‘traditional’ hand-written feedback.

Students’ comments to open-ended survey questions highlighted some reasons why audio recordings and email were so popular, for example, mentioning that feedback was more detailed and timely, and felt more individualized than the feedback they were used to, and that, in audio recordings, intonation helped them to identify positive and negative comments.

Questions also asked which aspects of general language accuracy students felt had significantly or somewhat improved by working with LDF. High numbers of students felt they had improved their use of transition signals (42% significantly, 33% somewhat), general grammar (42% significantly, 40% somewhat) and sentence structure (60% significantly, 33% somewhat). Naturalness of expression, text structure and general vocabulary were also perceived as having improved at least somewhat by most students. The

reasons for these perceptions, as reported in open-ended survey questions, centred around the feedback being more specific and placed throughout the text, in contrast to rather more summative feedback comments usually written at the ends of texts.

The academic writing skills of critical thinking and self-editing were seen as having significantly improved by over half of the students surveyed, and as having somewhat improved by around 40%. Large majorities of students also reported perceiving improvement in logical argumentation skills (87%) and their ability to engage in academic discourse (80%). Overall, over three quarters of the students surveyed felt that LDF had been more helpful in improving their academic skills related to essay writing than the hand-written feedback they traditionally receive.

Comparisons of students' grades on the essays for which they received LDF and their grades on other essay assignments supported the self-reported perceptions, and confirmed that there was some real improvement. On a grading system out of 15 points, students' essay scores on the LDF assignment were on average 2.22 points higher than those received one year earlier and

on average 1.7 points higher than those received on an assignment with traditional feedback in the same term. The students' essay scores on an assignment with traditional feedback in same term as the LDF assignment were on average just 0.64 points higher than those received one year earlier – smaller improvements here would seem to indicate that LDF had a role to play.

Discussion

These results show that learners responded positively to the LDF ap-

“Students' willing use of digital delivery modes for feedback can be practical for teachers, who can often provide more, and more detailed, feedback more quickly than by writing comments by hand.”

proach and that it was beneficial for their written work. The findings thus highlight LDF as a viable feedback practice in this context, with advanced-level EAP students. Students' comments in response to open-ended survey questions hint at concepts which may underpin the efficacy of the approach: LDF addresses issues of intelligibility, 'authority' over the texts, learner

autonomy, and personalisation. These issues are not unique to EAP, and thus the LDF approach may warrant piloting in other contexts, too.

Students' willing use of digital delivery modes for feedback can be practical for teachers, who can often provide more, and more detailed, feedback more quickly than by writing comments by hand. It can also be efficient for tutors to give feedback on selected aspects of the writing during the drafting stage, rather than attempting to address all of the issues at once. Involving students in the decisions about the delivery of their feedback and content may also help remove some urgency from teachers to agree on one 'correct' feedback procedure, and can moreover enhance differentiation by allowing for students to request feedback that suits their individual strengths and weaknesses.

There are, of course, a few caveats to these promising findings. For example, students may not know which aspects of their work to ask about, or may not be able to identify their own weaknesses, and lower-level learners may not have sufficient meta-language to pose effective feedback questions. I address some of these drawbacks below.

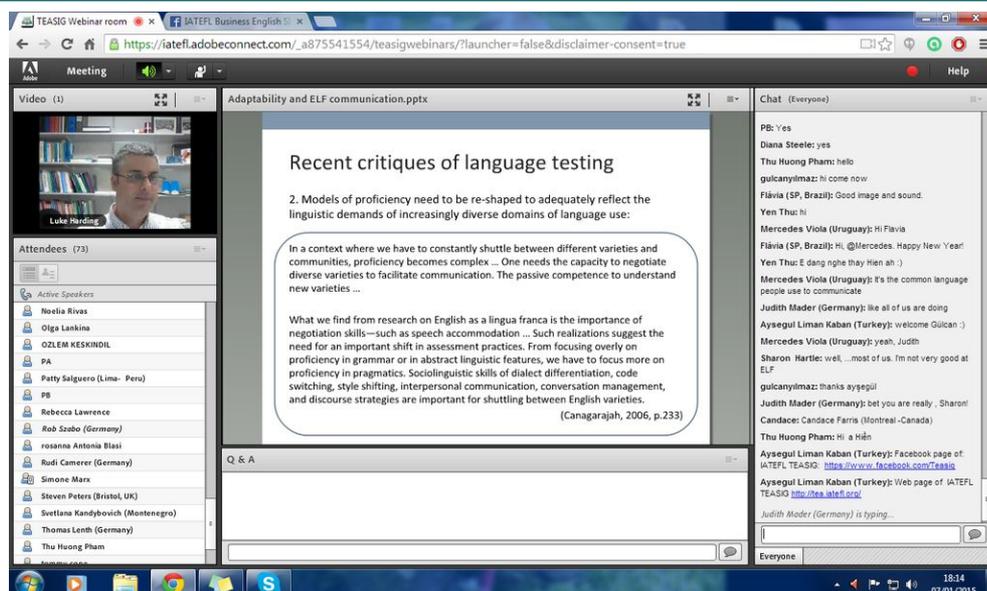
Tips for implementing LDF

- LDF is easiest (I find) if students submit their work electronically.
- Decide which delivery modes you wish to offer, and how students should pose the questions (e.g. footnotes, comments in margins, or cover-sheet).
- Allow any/only feedback delivery modes that are workable in your context. (i.e. perhaps not face-to-face consultation if you have 300 students!)
- Talk through the LDF process and provide a handout to support students the first time(s) they work with it.
- Combine LDF with writing workshops, peer review, and other ways of getting students thinking more carefully about their work, to scaffold the LDF process.
- Aim to provide feedback in a timely manner and allow enough time between draft submissions for students to thoroughly engage with the feedback.
- Give students guidance and/or training on what makes for a good response to the writing task at hand so that they can ask about these points in their own writing.
- If possible, provide students with a grading matrix so they know what aspects of their writing they should pay attention to.
- Encourage students to pose both specific questions (e.g. Which word is most appropriate here?) and general questions (e.g. Is the vocabulary I'm using formal enough?).
- If necessary and feasible, allow feedback questions in students' L1.
- For audio recordings or emailed feedback, insert line numbers to be able to refer to specific parts of the text.
- For audio recordings, you can email students an .mp3 file, or use free recording websites such as www.vocaroo.com and send students the links to their feedback.
- Remind students to have their written text at hand whilst listening to / reading the feedback.
- Note down common problems that students do not request feedback on, and use these to plan future teaching or workshops.
- Sit back and enjoy a cup of tea, as LDF will probably save you time. (You might even want to throw out your red pens!)



Foto: Sharon Hartle

TEASIG webinars 2017



The webinar can be accessed on the day at: <http://iatefl.adobeconnect.com/teasigwebinars> and is open to everybody. It will be recorded, and the recording will be accessible to everybody for a week on the TEASIG website, after which it will only be available to IATEFL members. TEASIG webinars will be continuing in 2017 and details will be announced on the TEASIG website.

Date	Presenter	Title
19.12.17	Felicity O'Dell and Russell Whitehead	<i>Getting it right, getting it wrong: a tour of the LT123 testing principles.</i>
	<p>This webinar with Felicity O'Dell and Russell Whitehead will discuss LT123's list of ten principles for 'good' testing, sharing plenty of examples of successful and unsuccessful items and tasks. This promises to be a useful round-up for newcomers and experienced testers alike.</p>	
26.02.18	Gudrun Erickson	<i>The Dual Function of Assessment – Enhancing Learning and Equity</i>
	<p>This webinar will focus on the dual function of assessment, whether in classrooms or large scale contexts, namely to enhance learning as well as fairness and equity. Gudrun will discuss the balancing act between the two, emphasizing common principles that need to guide planning, practices and materials at individual as well as structural levels.</p>	



Giving testing some good press for a change Assessment as learning check - a meaningful paradigm shift for trainee teachers



Lynn Williams

grew up and went to school near Manchester, England. She studied for a language degree at university and later added a teaching qualification. She now teaches English language and literature at a Swiss grammar school and also works as a teacher trainer at the PH FHNW in Basel. She is particularly interested in assessment practices, the teaching of literature, and opportunities for individualisation.

lynn.williams@fhnw.ch

After several years in teacher education I continue to be amazed by the trainee teachers in my seminar groups. They consistently bring with them a huge amount of knowledge, whether linguistic, literary or cultural. They also bring a wealth of skills in areas ranging from tutoring to administration. Finally, and in my view crucially, for the purposes of training new teachers who are alert and aware, they have of course clocked up thousands of hours as learners in school contexts themselves. These vital experiences (good lessons and bad, inspiring topics and less inspiring ones) can very usefully be put to work in a range of ways in our seminars – from examining learner (and teacher!) motivation, for example, to reflecting on classroom methods which learners appreciate. This provides an ideal situation for productive seminar work which focuses on classroom practice(s).

This focus on classroom practice necessitates a pragmatic approach to learning how to teach, and we therefore work very practically in our seminar sessions. It goes without saying that, as a university teacher education department, we consistently draw on research findings and accepted best practice in our work, but we nevertheless always have in mind the ultimate transfer to the classroom and possible implementation with secondary

school learners. To this end, trainee teachers are regularly asked to design a task or sequence of tasks for one or more lessons in which they put their theoretical readings into practice.

Against this backdrop we engage with many key issues of upper secondary (language) learning, looking into principles of lesson-planning and curricula, the four skills, and task- and competence-oriented learning. Within this demanding menu of important angles, one major area it's easy to feel overwhelmed by as a beginning teacher is surely that of assessment and testing. Testing gets a lot of bad

"The link between teaching and assessment is not always clear at first sight and sometimes needs to be made more explicit."

press: pupils have on occasion been known to complain about the volume of material to be learned, and teachers about all the marking they

have to do. For this reason, as well as many others, I feel it's important to put a positive spin on assessment and to show trainee teachers that assessment can be a manageable, rewarding and ultimately interesting aspect of teaching and learning.

We typically start by looking at classic test design and formats, identifying the markers of a successful test, and at the same time contextualising our intuitive reactions against a background of the theoretical un-

derpinnings of test design. After considering theoretical aspects such as validity and reliability, we move on to look at seemingly banal but ultimately essential practical aspects like concise and precise rubric formulation or even numbering lines in a reading text for ease of orientation when discussing a test in class at a later date.

unsurprising, that I've noticed over the years is that trainee teachers frequently associate assessment with tests. In response to this observation, I have typically started a seminar session on assessment and testing with a fundamental discussion on the nature of assessment with a task like this:

A test is an assessment but an assessment is not necessarily a test!

Discuss what you understand by this and be prepared to feedback in the seminar group.

As we know, the range of options which go beyond the classic 'test' scenario is huge, ranging from quiz questions at the end of a lesson to self-evaluation

according to prescribed learning objectives. Trainee teachers are therefore encouraged to trawl through their (language) learning biographies to identify possible assessment scenarios, particularly those that explode the classic canon of controlled grammar practice exercises. In a further step, the trainees devise an appropriate assessment strategy for a teaching unit they are developing and reflect on the rationale behind their choice.

However, in my experience, the link between teaching and assessment is not always clear at first sight and sometimes needs to be made more

explicit. For example, trainee teachers frequently (and justifiably) take a topic or particular article or film clip as the starting point for a potential teaching unit without necessarily having any special learning objective in mind – they simply trust their instincts, knowing that it is likely to motivate upper secondary learners or that it serves to exemplify a particular language point within the context of an authentic text – bingo! And good for them!

In my seminars, therefore, as well as in individual coaching settings and in written feedback on draft units, I seek to guide trainee teachers towards examining the underlying rationale for the unit, the reasons why they have chosen the particular topic, text or task. Having established the competences or knowledge they want their upper secondary learners to acquire, it becomes easier for them to think about their (implicit) objectives for the unit and, consequently, also about how they can then check that these objectives have been achieved. This is the point at which we turn our attention to assessment procedures. Trainees realise – sometimes with a jolt, sometimes with a degree of disappointment, but more often than not with relief – that a classic 'test' might not be necessary to check their learners' progress. Or it might not be feasible, plausible or manageable within certain constraints. Or, quite simply, it might not be the best way of checking learning.

"So what is the best way?" my train-

Practical tips for test design

Finalising your test and putting it to good use

- Try your test out on a colleague. Are the instructions clear enough? Do they come up with any unexpected answers? Make a note of them.
- Proof-read & proof-read & proof-read again 😊
- Once you are satisfied with the test, use it to inform students / shape your teaching / student preparation without disclosing key aspects.

Furthermore, I encourage trainee teachers to engage with principles of backwash, my message being essentially:

Once you are satisfied with the test, use it to inform students / shape your teaching / student preparation without disclosing key aspects.

Something interesting, and perhaps

"Assessment" or "test"?

Central idea for this session:

- A test is an assessment but
- an assessment is not necessarily a test!

➤ *Discuss what you understand by this and be prepared to feedback in the seminar group.*

ees then start to wonder, and I'm always nothing short of thrilled at this point. It means we are breaking away from the mindset that sees teachers replicate the kind of test they themselves sat at school or those that they might find through external searches for material, often designed to fit all. At this point, let it be said that a good test is a good test, no quibbles there (and I've taken, set and corrected enough of them by now!), but we are now at a juncture setting out into new territory – namely the realisation that a good assessment doesn't need to be a test at all...

In my opinion, the 'best' way to assess a unit of work is at the same time an objective and subjective decision. Objective because much of the literature on testing rightly argues that an assessment should mirror the learning as far as this makes sense (i.e. it would be odd to follow up a series of discussion lessons to develop fluency with a discrete-item grammar test, and by the same token, writing a formal letter may not necessarily be the best way to check learners' understanding of a range of new grammar points). Once trainee teachers have identified what they consider to be the best way to check learning of the unit they have put together – i.e. that which best aligns with the learning they are striving for and the overall competence-oriented outcome they are aiming at – they are encour-

aged to engage with a variety of questions to help them refine their assessment strategy. These might productively include the following:

- Does the assessment method reflect the style of work undertaken? (Does it assess the learning which has gone before? Are learners well-prepared to respond to the task? In what way will it demonstrate learners' grasp of the topic in question?)
- Is the assessment task manageable? (Timing, level, resources, teacher feedback)
- Is the assessment task motivating? (Does it offer learners choices? Does it appeal to different interests? Does it take up a relevant topic?)

Once trainee teachers have 'vetted' their assessment procedure to check its viability in these areas, they are in a position to make final adjustments and refine their assessment task. In a final step, we typi-

cally share teaching units and the related assessments at the end of the semester. There is no small degree of pride during this phase and, having had the privilege of guiding trainee teachers in the realisation of their proposed assessment tasks, I understand their pride! I also hope that a few key tenets will stay with them after our joint work on assessment strategies:

- See that testing is about more than just attaining scores. Make tests worthwhile to all involved: motivating, focused, realistic, informative.
- Link testing to classroom work as far as possible – scores. Make tests worthwhile to all classroom work and testing.

These are principles which guide my own classroom work, leading to learning which I hope is all the richer, sustainable and more enjoyable for it.

Takeaways on testing

- See that testing is about more than just attaining scores. Make tests worthwhile to all involved: motivating, focused, realistic, informative.
- Link testing to classroom work as far as possible – aim for interaction between regular classroom work and testing.

Assessment in the upper secondary ELT classroom
(introductory handout for trainee teachers)

Forms of testing and assessment

It is important to note that an assessment does not necessarily need to be a test. It is equally possible to check whether students have achieved the relevant learning objectives for a unit in a different assessment set-up than a formal written test (or the equivalent speaking assessment). Some ideas include:

- concept questions (written or oral) individually, in groups or in the plenary (always with feedback and teacher clarification/expansion where necessary and appropriate)
- self-evaluation (based on teacher prompts)
- peer feedback (followed by discussion in plenary/consultation with the teacher)
- written or oral feedback comments from the teacher
- an application of the competences acquired in a new setting (transfer)
- a product (poster, mind map, reader's guide, handout, flyer, theatre programme...)

Principles of assessment

All forms of assessment must demonstrate genuine validity and stand up to scrutiny. Therefore,

- adhere to the standard principles of testing (reliability, validity, fairness) and offer opportunities for positive backwash
- communicate test requirements, format and criteria early on
- be prepared to answer any questions (and re-think as and when necessary)

Practicability

Tests should be straightforward for teachers to administer and students to take, and should not present unnecessary challenges in marking.

Written tests should...

- test what has been learned
- reflect previous teacher information on the test
- offer clear instructions and indicate marks per question
- ideally, present a variety of tasks
- give all students a fair chance, and strong students the chance to excel

Feedback/corrections are essential and should...

- clearly show students where they went wrong, help address these points, and offer students paths for development
- report on performance in content terms, insofar as this is being tested

Suggested further reading

- Alderson, J.C., Clapham, C. & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge: CUP.
- Douglas, D. (2010). *Understanding language testing*. London: Hodder Education.
- Fulcher, G (2010). *Practical language testing*. London: Hodder Education.
- Hughes, A. (2003). *Testing for Language Testing* (2nd ed.) Cambridge: Cambridge University Press.
- McNamara, T. (2000). *Language Testing*. Oxford: Oxford University Press.

It is also worth consulting the guidelines for good practice on language testing and assessment issued by the European Association for Language Testing and Assessment (EALTA); see www.ealta.ue.org.

“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. Findings⁴

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:

“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

- 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.

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.



IATEFL TEASIG Webinar

Developing a test: Where do you start? When should you stop?

24 October 2017

Jo Tomlinson

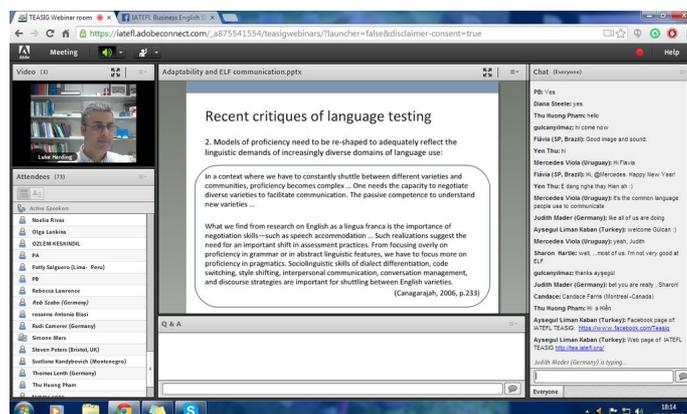
As we didn't get time to discuss the scenarios at the end of the webinar, Jo has provided us with a small collection of thoughts and considerations. She is sure there are many more.

You are responsible for developing an intermediate level test for your school. There are ten intermediate classes with ten different teachers. They all use different course books. If the students do well enough, they can move to the next level. Teachers don't have a lot of time for marking. Tests will be run monthly and there will be 150 students taking the test.

Possible approaches/considerations:

- Collect course books used and analyse language, functions, skills etc. used in each book.
- Identify common crossovers.
- Evaluate whether the test should include any skills element, taking timing and resources into account.
- Consider that you may need multiple versions.
- Pilot tests with the teachers (ask them to complete within the time etc).
- For practicality, you could assess the functioning after the first tests (at a most basic level, ask teachers to name the students in class that are the strongest and weakest, and compare against results).
- Changes can be made after testing to further improve test functioning, possibly based on placement of students and how well this is working.

You are designing a business English test for telesales workers. You will be responsible for giving the test to new workers (roughly 5 workers a month), and with the results they will be given specific modules



to improve their working capability in English. All new employees already have a B2 level of English.

Possible approaches/considerations:

- Analyse the language, functions and skills that are used in the role by studying existing workers. Are they using a script? Do they go off script?
- Consider the impact carefully; could employees lose their job due to this test? Do they know or not know this? Do you know this?
- Practicality is less of an issue compared to scenario 1, and you may be able to activate the test language in authentic tasks that they will carry out in their work.
- Ensure test design has clear links to the modules that you can recommend.
- Consider carefully when the tests will be done; there could be some interactivity here between their training for the job and their test performance.
- Pilot the test on existing employees; they could reap the benefits of free further training ideas!

The unanswered questions asked during the course of the webinar have been answered by Jo in the following:

Is a test always a compromise governed by rules?

I would say so. From my experience, test development and the resulting test is always a series of compromises. Test development often requires two (possibly) distinct approaches. The first is understanding and using good practice standards. The second is pragmatism. Put simply, it's the compromise between what should be done and what can be done. I'm sure all of us testers have an ideal way of testing in a specific situation, but available resources often limit that. We just have to do the best we can with what we've got!

Isn't test development more cyclical than linear? Or like a jigsaw puzzle where the pieces never manage to fit together?

I think it definitely can be. Especially if the test is to be repeated (perhaps with subsequent versions) over time. If there is a single test being created, I personally like the stair metaphor as it gives an idea of moving onwards (and upwards!), yet we must acknowledge and embrace the fact that we're bound to move up and down quite a few stairs before getting to the top.

I also like the jigsaw puzzle idea. Perhaps our aim should be to fit enough pieces together to be able to 'see the picture'!

How important is it that you speak to all stakeholders in their own language – learners, school head, HR, bosses, parents?

When creating a test, I see no reason why we shouldn't do this if we can. The more we know from everybody involved, the more we can understand the context. Almost nobody likes anything that is forced upon them. The more we val-

ue the opinions of those involved, and express this in a dialogue, the more likely they are to value the end product.

How much value do you place on ancillary parts of test development such as item writing, rater training, statistical analysis, and pilot feedback?

These areas are fundamental to test development, but the part they play depends on the test design and context. However, if your test design includes these elements and they aren't carried out well enough, then you can't actualise that particular test design. Take, for example, criteria that raters don't particularly understand or are self-conflicting; it doesn't matter if you have the best plans in the world, the results from that test won't be able to tell you much.

What type of test do you personally like constructing best?

This is a difficult question because I enjoy constructing all tests (honestly!). I would say I particularly enjoy analysing the construct. I work with a colleague who comes from a linguistics background in contrast to my testing background, so you can imagine the conversations. I particularly enjoy work that gives us the creative means to try something new. Integrated tests are interesting to work on, simply because they can be quite challenging from a development angle.

Can you tell us something more about note-taking as a task? Is note-taking a fair way of testing these days, as we write so little? Or could we do it on a tablet?

Let's look at this from an EAP perspective. Note-taking still exists, of course, and might be called the most authentic (yet messy) way of testing listening. There is a lot that can go wrong with note-taking as a testing tool, especially in the

rating, yet there are ways in which these problems can be limited.

However, as we discussed in the webinar, note-taking in general is moving on. More and more students are recording their lectures and taking photos of the slides. Now, I suppose an essential question for us as test developers is 'What are those students doing with that information after the lecture?'

Good point regarding the way different examiners interpret particular assessment criteria (e.g. spelling). How should we deal with this and make sure that different assessors interpret the assessment criteria they're supposed to be using in the same way? Regular monitoring and annual (re-)training sessions don't seem to help a lot as the grades are still different when we do random monitoring.

Yes, I think it's always a somewhat constant struggle. If we take spelling into consideration, then from my experience allowing degrees of flexibility in spelling can be extremely problematic as what is an acceptable error for one rater will be unacceptable for another. There are also numerous problems with having a list of 'acceptable spelling errors'. In this instance, I would be tough and say no deviation from the correct spelling is allowed. I also explain why and ask the raters whether they think it is fair that one student should 'pass' while another should 'fail' due to a generous rater giving a few more marks and a tougher rater giving a few less. Most agree that it's not acceptable.

In terms of wider interpretation of the marking criteria, I think it is essential to discuss the fundamental terms for which we perhaps take for granted that we all have a shared meaning (for example, 'task achievement' or 'cohesion'), as well as analyse and discuss degrees of achieve-

ment and what they mean (e.g. 'convincing', 'minimal' etc.) and show examples of these. Timing is also important; ensure the training/standardisation is fresh in raters' heads wherever possible, and perhaps double mark and feedback (depending on whether this is feasible).

How much space do you think there is for creativity in testing?

I hope there is a lot of space for creativity. The more we can facilitate testing that is learner and context specific, the more meaningful it can be for the test-takers and also in terms of results within the particular context. I think that one of our aims as test developers (and perhaps this is sometimes understated) is to assess in a way that candidates find a test a joy to do, rather than a hurdle to overcome.



TEASIG Member Spotlight: Dr Christine Coombe

Where are you based?

Dubai, UAE

How long have you been there? / Which other countries have you worked in?

I've been in the Gulf for 25 years and have worked in the US, France, Oman and the UAE.

How long have you been a member of IATEFL?

As far back as I can remember.

Why did you choose TEASIG as your SIG?

My primary area of specialization is testing/assessment.

Are you a member of any other SIGs?

No not with IATEFL but I chair two SIGs with in TESOL Arabia: the Testing, Assessment and Evaluation SIG as well as the Leadership and Management SIG.

What do you get out of TEASIG?

TEASIG is a community within IATEFL where everyone has the same or similar interests; I enjoy this dynamic and especially meeting testers from around the world and learning about what they are doing not only in the field of ELT but in language testing and assessment.

What other services or activities would you welcome?

I would welcome more peer-edited and reviewed publications from our community.

What sort of work do you do?

I am a full time faculty member at Dubai Men's College where I teach General Studies courses.

How much of your work time is spent on testing, evaluation and assessment? In my job about 15% of my time is spent on assessment or assessment related activities; in my time outside the workplace I spend a great deal of time training teachers how to write tests and writing books on assessment.

Do you see yourself as more of a teacher or as more of a tester?

I'd like to say 'both' but in actual fact I am now spending more of my time as a teacher.

Do you work in a team or alone?

I believe that teamwork is crucially important so I work in teams much more so than alone.

Do you set tests?

Yes, I do but more of my 'testing' time is spent editing tests that others have written and advising on testing and assessment matters.

What sort of tests are they? How high are the stakes for the test-takers?

They are usually midterm or final assessments of both the formative and the summative kind; the latter are pretty high stakes as they usually count for 30% of a student's final grade.

Are the tests you set published in any form?

Not any more although I worked for many years as the Chief Examiner for the English B component of the International Baccalaureate and many of those tests were published.

Do you enjoy setting tests?

Yup, after travelling, watching Game of Thrones and scuba diving, it's my favorite thing to do!

Can you use the same set of skills and principles in **several different contexts?**

I believe so. I feel the way we package ourselves and this includes our skills and principles is very important.



Do you carry out research on testing and related fields?

Yes.

Have you published on testing issues?

Yes, I have a few research-based articles in top tier journals like Language Testing, TESOL Quarterly and Language Teaching. However, my favorite kind of publishing is on helping teachers increase their assessment literacy.

Have you spoken at conferences on testing issues?

Yes, lots.

Do you discuss testing issues with colleagues?

My colleagues often consult me on testing issues.

Do you feel qualified to set tests?

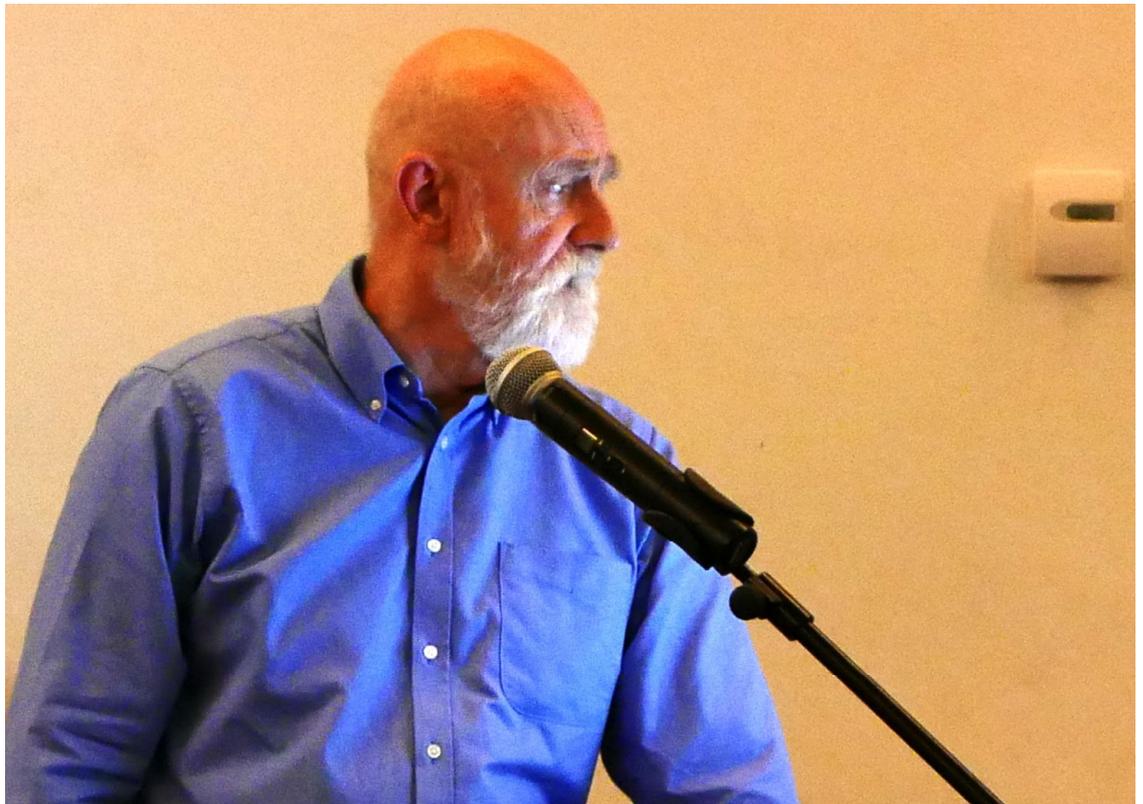
Yes very much so.

If you do not set tests, what testing-related work do you do (advisory, marking, curriculum design ...)?

I pretty much do it all, test setting, advising on tests, marking, examining etc.

Christine, thank you very much!

Email: ccoombe@hct.ac.ae



TEASIG Member Spotlight: Gladys Quevedo Camargo

Where are you based?

I'm based in Brasilia, the capital of Brazil, right in the centre of the country.

How long have you been there?

For almost five years.

Which other countries have you worked in?

I've worked only in Brazil.

How long have you been a member of IATEFL?

For about 4 years I think.

Why did you choose TEASIG as your SIG?

Because I think testing, evaluation and assessment are crucial issues for any language teacher to know about.

Are you a member of any other SIGs?

No, I'm not.

What do you get out of TEASIG?

The most important thing, in my opinion, is that TEASIG allows me to be in contact with people that share the same interests as mine – this allows me to learn from other people, exchange ideas with people from different parts of the world and find solutions to common problems we might have in our daily work.

What other services or activities would you welcome?

As I don't live in Europe, I would very much welcome having more information on what was presented and discussed in the events promoted by the TEASIG, particularly the one-day ones!

What sort of work do you do?

I am an English language teacher educator and I teach both undergraduate (future English teachers) and MA students (most of them teachers working in public schools).

How much of your work time is spent on testing,

evaluation and assessment?

A lot, not only as a teacher but also as a researcher and as an MA tutor.

Do you see yourself as more of a teacher or as more of a tester?

As a teacher, for sure.

Do you work in a team or alone?

Mostly alone, unfortunately.

Do you set tests?

Yes, I do.

What sort of tests are they?

Most of the time they are classroom tests, but I am also part of the committee that sets entrance exams for the post-graduation programme in Applied Linguistics of my department.

How high are the stakes for the test-takers?

They are high stakes because they are related to the student's progress at university – in the case of the classroom tests, and the success or not in starting an MA.

Are the tests you set published in any form?

No, they're for our own use.

Do you enjoy setting tests?

Yes, a lot!

Can you use the same set of skills and principles in **several different contexts?**

Yes, sure, as long as you know the basic principles of assessment.

Do you carry out research on testing and related fields?



Yes. At the moment I'm carrying out some research on assessment literacy in Brazil for my post-doc studies.

Have you published on testing issues?

Yes, I've got some articles published in Brazil and one abroad.

Have you spoken at conferences on testing issues?

Yes, I have.

Do you discuss testing issues with colleagues?

Yes, a lot and whenever I have the opportunity!

Do you feel qualified to set tests?

Yes, but assessment is a very broad and complex area, so there's always something to learn.

Gladys, thank you very much!

Email: gladysquevedocamargo@gmail.com





The aims of IATEFL are to:

- Benefit English language teachers all over the world providing opportunities for personal and professional development.
- Enable the international network of ELT professionals to grow, by encouraging and fostering the regional and local groupings, so that members can learn from each other.
- Encourage grassroots professionalism where all categories of members at whatever stage of their career can make significant contributions and continue to learn.

Linking, developing & supporting ELT professionals worldwide

IATEFL Patron: Professor David Crystal

Who is IATEFL for?

IATEFL is truly international with the majority of our members living in countries where English is not the first language. Members contribute greatly to the life of the Association in a wide variety of ways: through participation in our 16 Special Interest Groups (SIGs); contributions to any of our magazines: *IATEFL Voices* and SIG publications; volunteering on IATEFL committees; communicating developmental ideas to Head Office and giving papers at, or simply attending our many conferences, workshops and seminars. The Association is managed and administered from an office in Faversham, UK. The nine remunerated staff are supported by volunteer committees, including the IATEFL Trustees and the Special Interest Group committees consisting of ELT professionals from a range of countries.

Copyright Notice

IATEFL retains the right to reproduce part or all of this publication in other publications, including retail editions.

Contributions to this publication remain the intellectual property of the authors. Any requests to reproduce a particular article should be sent to the relevant contributor and not IATEFL.

Articles which have first appeared in IATEFL publications must acknowledge the IATEFL publication as the original source of the article if reprinted elsewhere.

+++++

Disclaimer

Views expressed in this IATEFL—Newsletter Issue 63 (December 2017) are not necessarily those of the editor(s), of IATEFL TEASIG, of IATEFL or its staff or trustees.