

Language Assessment Matters

Issue 8, October 2017

The Newsletter of the Association for Language Testing and Assessment of Australia and New Zealand

Welcome to the eighth issue of Language Assessment Matters, the newsletter of the Association for Language Testing and Assessment of Australia and New Zealand. This is by far our biggest issue to date.

A very big thank you is due to the ALTAANZ committee members who have written items and recruited contributors for this issue. A particularly big thank you to past Student Representative Sharon Yahalom and to current Student Representative Megan Yucel for their interviews with Professor Jin Yan of Shanghai Jiao Tong University and IELTS expert Pauline Cullen respectively. Both interviews offer insights into the world of large scale high stakes testing. Our conference reviews reflect the degree of engagement of our membership with the international community of language testers. We have reports from Colombia, Taipei, and Leuven, from Morena Magalhaes, Leo Xiaohua Liu, and Michelle Czajkowski. Finally, our co-presidents have updated us on parts of their work in language assessment this year. Rosemary Erlam reports on Assessment Literacy work with school teachers in Auckland and Noriko Iwashita updates us on language assessment standardisation work at The University of Queensland.

My comment during committee meetings has been that this is a bumper issue which will need the attention span of four or five coffee breaks - I hope you will forgive us the distraction and enjoy the content.

Johanna Motteram, Communication Officer

Look out for the next issue of Papers in Language Testing and Assessment

PLTA Volume 6.2 will be available this November <http://www.altaan.org/>

Language Assessment Matters is the newsletter of ALTAANZ, the Association for Language Testing and Assessment of Australia and New Zealand.

Do you have a Language Testing or Assessment related item you would like to have included in the newsletter? If so, please send your submission to altaanz@gmail.com with "newsletter" in the subject line. Possible contributions could include a report on work in progress, a PhD or Master's thesis report, a conference review or report, discussion of a current issue in language testing or assessment, a description of an assessment task which has worked well for your learners or an assessment-related anecdote that you think our membership would find interesting.

INAUGURAL ALTAANZ WEBINAR SERIES

This October and November we are trialling a new initiative. The ALTAANZ Webinar series.

The Webinar series is intended to provide a platform for knowledge sharing and community building within the association. We surveyed our student membership for content requests. We hope the sessions will be interesting for many of our members.

Please keep an eye out for full abstracts for the sessions and registration instructions on our Facebook page, on our website, and via email.

The schedule for the webinar series can be found on page 20 of this newsletter.

Johanna Motteram, Megan Yucel, and Leo Xiaohua Liu

AGM

The ALTAANZ Annual General meeting will take place during lunch break on Monday 27 November. We hope that many ALTAANZ members will be able to attend to hear about our activities in the last year and to vote in new Vice President, Treasurer, Secretary, Communication Officer, Teacher Representative, and Student Representatives to serve on the committee.

Further information, including the venue and exact time of the AGM and our Presidents', Treasurer's and PLTA Editors' reports will be distributed soon.

Call for nominations**Election of office-bearers, ALTAANZ Committee**

We would like to call for nominations to the positions of Vice President, Treasurer, Secretary, Communication Officer, Teacher Representative, and Student Representatives for the ALTAANZ Committee. The Student Representative Position is for one year and the President position for two years commencing January 2018. If you would like to nominate for any of these positions or to nominate another ALTAANZ member contact us altaanz@gmail.com no later than 5pm AEST on Monday 13 November, 2017. Elections for office-bearers will be held at the forthcoming ALTAANZ conference (the exact day/time of the AGM will be confirmed soon).

Please contact the current ALTAANZ Secretary, Kate Quigley [<Katherine.Quigley@vuw.ac.nz>](mailto:Katherine.Quigley@vuw.ac.nz) if you would like further information about any of these positions.

Is it ok to use Kahoot for assessment in the language classroom?

Rosemary Erlam
The University of Auckland



Background

As outlined in Erlam (2016), the author is the Academic Director of TPLT (Transforming Practice in Language Teaching, www.tplt.ac.nz), a year-long In-service curriculum support programme for teachers of ten languages other than English (French, German, Japanese, Mandarin, Samoan etc) in New Zealand schools. The TPLT programme caters for up to 70 teachers per year teaching at all levels of the school system – primary, intermediate, secondary - and in a variety of contexts. Some are qualified and experienced language teachers, others have no language teaching qualifications and experience. The programme includes 3 components:

1. Language study: classes to support language study and the opportunity to sit internationally recognised exams;
2. In-school support: teachers are visited 4 times for lesson observations. Following each observation they take part in an evidence-based learning conversation where they are encouraged to self-evaluate their practice.
3. Pedagogy: 8 days delivered regionally in blocks. The focus is on Second Language Acquisition theory and classroom practice. These days include language-specific work and a university course.

During the TPLT programme teachers learn about tasks and task-supported language teaching in the stage 3 University of Auckland pedagogy course. As part of the assessment for this course they are required to design, teach and evaluate a task, the evaluation is written up and handed in as a report.

Teacher assessment literacy

The TPLT team members had noticed that teachers were, in their evaluations of the tasks they taught in their classrooms and wrote up as part of their assessment for the pedagogy course, reluctant to investigate student learning. Instead they tended to restrict their task evaluation to an investigation of students' affective response to the task, having them complete a survey or questionnaire. Furthermore, it was apparent in the part of the pedagogy course that dealt with assessment that teachers had difficulty relating assessment practice to lesson aims. For example, they might propose that a lesson involving aural language input only could be assessed by requiring students to demonstrate comprehension of written input. The conclusion was that teachers were lacking assessment literacy. It was therefore decided that the 2017 TPLT programme would include a focus on assessment (this was unfortunately, because of time restrictions, limited to two half-day workshops only). The In-School Support Facilitators also reported, from their observations that a number of teachers were using Kahoot to assess student learning. It was therefore suggested that

the appropriateness of using Kahoot for assessment could be explored in the assessment workshop (hence the title).

'Focus on assessment' workshops

This section presents an overview of material covered in the workshops.

1. Challenge

The teachers were first challenged to consider that they will need, at times, evidence of what their students have learnt. They were referred to the article 'Lots of games and little challenge – a snapshot of modern foreign language teaching in English secondary schools' (Wingate, 2016) and encouraged to think about the importance of establishing evidence for student achievement.

2. Different types of assessment

In order to capitalise on teacher's existing experience and practices (Fulcher, 2012) discussion ensued about the differences between formative, summative and self-assessment. In considering each type of assessment practice, participants were encouraged to ask the question - 'who wants to know what about whom and for what purpose?'

3. Assessing tasks

Participants were given 4 different language tasks that had been designed by past teachers of TPLT. They were told that they had to suggest appropriate formative, summative and self-assessment practices for these tasks.

Participants were given two 'tools' to help them:

(a) A lesson plan template with attention drawn to the 'learning intention' section as the place to start.

(b) reference to Nation's 4 strands (meaning focused input, meaning focused output, language focused learning, fluency). As part of identifying learning intentions participants were encouraged to think of which of these strands the task catered to.

For activity 3. (above) answers were elicited and analysed in a whole-class discussion.

So what about Kahoot?

The teachers then took part in a task inspired by Shintani (2016). It was input-based and designed for ab initio learners of French. Each participant had a picture of a zoo and an envelope of zoo animals. They had to listen and place animals in the correct cage or enclosure according to instructions in French. Many of the vocabulary items were cognates with English (e.g. tigre, lion etc) but some were not (e.g. singe, phoque, ours). After each task participants were shown a picture of what the correct answer should have been.

They then completed an output task where students in teams competed to be the first to name an animal the teacher pointed to on a power point display. In a subsequent 'survey' task they had to ask every class member which of a number of pictured options their favourite animal was, the aim being to find out which animal the class liked most.

The participants then played 2 classic kahoot games to assess their vocabulary learning.

1. Stimulus was 'le singe' – possible answers were: 'zebra'/'monkey'/'bear'/'seal' Etc.
2. Stimulus was 'il aime les bananes' – possible answers were: 'zebre'/'singe'/'ours'/'phoque' Etc.

The limitations of Kahoot for assessment purposes were then discussed in relation to the tasks taught. These were summarised as:

- no assessment of aural input possible – only written (in tasks they completed teachers had not been exposed to written language)
- no assessment of productive language possible – (teachers had engaged in production during output task)
- Kahoot can mean an over-reliance on translation (but use of target language as in 2. is possible)
- length of stimuli that Kahoot can accommodate is limited (so comprehension of longer text cannot be assessed)

The teachers also highlighted another limitation:

- credit is given for speed which is not necessarily a construct relevant to language learning

However, the motivating aspect of Kahoot was also discussed, and one teacher described how he had had students in his class, in groups, design a Kahoot game to challenge each other in their learning.

Conclusion

It was rewarding to work with these teachers and to challenge them in their thinking about assessment. However, it was also perhaps frustrating as so little can be achieved in such a limited time frame. More information is needed about the nature of their current assessment practices in order to know what their ongoing needs are (Tsagari, 2017). In the meantime, these teachers were encouraged to consider attending the Teacher Assessment workshop at the forthcoming ALANZ/ALAA/ALTAANZ 2017 conference.

References:

- Erlam, R (2016). Using evaluation to promote change in language teacher practice. *Papers in Language Testing and Assessment*, 5(1), 41-65.
- Shintani, N. (2016). *Input-based Tasks in Foreign Language Instruction for Young Learners*. Amsterdam: John Benjamins.
- Tsagari, D. (2017). Assessment literacy of Foreign Language Teachers around Europe: Research, Challenges and Future Prospects. *Papers in Language Testing and Assessment*, 6, 1, 41-63.
- Wingate, U. (2016). Lots of games and little challenge – a snapshot of modern foreign language teaching in English secondary schools. *The Language Learning Journal*. DOI: 10.1080/09571736.2016.1161061



Forward Planner: Upcoming Language Assessment Events

IATEFL TEASIG Seminar

The University of Bedfordshire, UK

28 and 29 October 2017

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

The 19th Annual MwALT Conference

Wright State University, Dayton, Ohio, Saturday, October 28, 2017

"Language Assessment and its Sociopolitical Context"

The Applied Linguistics Conference

ALANZ / ALAA / ALTAANZ Auckland, New Zealand 27-29 November 2017

(cfp open, dedicated testing and assessment stream, ALTAANZ AGM 2017)

<http://www.alanz2017.org/>

The 3rd International Conference on Language Testing and Assessment and the 5th British Council New Directions in Language Assessment Conference

Shanghai, December 2-3

<https://www.britishcouncil.cn/en/exams/conference>

15th EALTA Conference

Technology-Based Language Assessment: Benefits and Challenges

25th-27th May 2018, Bochum, Germany

<https://ealta2018.testdaf.de/>

LTRC 2018, University of Auckland, New Zealand July 2-6 2018

Please note, there will be no stand-alone ALTAANZ conference in 2018

Professor Jin Yan speaks with Sharon Yahalom: an ALTAANZ interview.



Jin Yan is a professor of applied linguistics at the School of Foreign Languages, Shanghai Jiao Tong University, China. Her research interest focuses on the development and validation of large-scale and high-stakes language assessments. She is Chair of the National College English Testing Committee of China. She is also 1st Vice President of the Asian Association for Language Assessment and co-editor-in-chief of the Springer open-access journal *Language Testing in Asia*. She is on the editorial boards of journals such as *Language Testing*, *Classroom Discourse*, *International Journal of Computer-Assisted Language Learning and Teaching*, *Asia TEFL*, *Contemporary Foreign Languages Studies*, *Foreign Language Research*, *Foreign Language Testing and Teaching*, *Foreign Languages in China*, *Foreign Language World*, *Foreign Language Education in China*.

Professor Jin Yan spoke to Sharon Yahalom, a PhD student at The University of Melbourne, about her career in language testing.

SY: Thank you very much for your time today, Professor Jin Yan. If you don't mind, let's start by discussing your career path as a language tester. How did you first become interested in language testing?

JY: I became interested in language testing quite early in the late 1980s. This was mainly because there was a testing program at my

university, Shanghai Jiao Tong University. I completed both my undergraduate and my graduate studies at Shanghai Jiao Tong and we had a testing program, I don't know whether you have heard of it or not, this is the CET test – College English Test.

SY: Yes, I have. But I'm interested to hear more.

JY: This test was developed by a group of professors in the mid-1980s. So the professors in charge of this program – or who initiated this program – were based at Shanghai Jiao Tong University. They were actually my master and PhD supervisors, Professor Liu Hongzhang and Professor Yang Huizhong. The professors and several key members from other universities set up the 'Working Group of the CET' and started this test in 1987. This was the test of the lower band, CET Band 4. We also have a test of the higher band, CET Band 6. The Band 6 was started in 1989 and in the same period of time as when I started my Master program. I started my Master program in 1988 –about 30 years ago. So I chose testing mainly because I was interested in this program and hoped that I could work on it after my graduation. And actually later I became a member of the CET team and I have been working on this test for almost 28 years now. So that's basically how I started my career.

SY: Very interesting! For what reason was the test first developed?

JY: The test was developed for college students. In the mid-1980s, a group of professors developed the national curriculum for what we call 'college English education' or 'college English teaching'. So there was a national curriculum which sets the requirements of English language teaching in higher education in China. To promote the implementation of this national curriculum, the professors decided to implement a national test. So the CET was

actually testing for the purposes of college English teaching, or what we call an educational testing program.

SY: As you mentioned, you've been working in the field of language testing for quite a while now. What do you find most challenging about language testing either on a practical or theoretical level or both?

JY: I consider myself mainly a practitioner of language testing and assessment. From my perspective, I find the technical aspects of language testing very challenging. By these technical aspects, I mainly refer to the design and development of language tests. For me, the most intriguing part of test development is: what are we going to test? That is the construct definition. So what is the ability to be measured? And also, as a language test developer, the operationalisation of this construct – how do we operationalise this theoretical construct in our assessment practice-- is also a big challenge. For the purposes of our teaching and learning, we often decompose this construct into separate skills like listening, reading, speaking, writing, and also in the Chinese context, we sometimes have a fifth component, that is translation. This seems relatively easy, but the skills, as we know, are often used in an integrated way. We now prefer to use so-called integrated tasks and ideally we would also prefer to use task-based performance assessment. But the problem is: how do we generalise from performances on these tasks? The generalisability of the scores of this kind of integrated task or performance-based assessment is very challenging. Even if we have a relatively satisfactory test design, quality control is an issue: the control of the quality of the items, the consistency of rating, and rater training. We may also have technical issues like test form equating. So I think we have put a lot of efforts into the technical issues of language testing. But as the chair of a testing organisation responsible for a very large-scale, high-stakes testing program, I began to feel that the

social dimension – McNamara and Roever called 'the social dimension of language testing' – was actually much more challenging.

SY: Could you tell me more about the social dimension of language testing from your perspective?

JY: Yes, of course. By this I mean, particularly in our situation, the impact of testing on teaching and learning - test washback or impact, and also fairness concerns: fairness in test design and delivery, accessibility of the test to test takers, and the problem of cheating, especially in recent years, high-tech cheating. There seem to be no easy answer to these questions. Can I give you some examples of the so-called social dimension? I find it extremely important to a language testing program like the CET.

SY: Yes, please do.

JY: I've talked about the issue of test design, and a very high-stakes test surely has a very strong impact on teaching and learning. To bring about positive impact or what we call 'impact by design', we would hope to use performance-based assessment tasks as much as possible in our test, tasks like note-taking for listening, essay writing or short answer questions for reading and so on. However, given the large scale of our test and also constrained by other practicality issues like scoring of constructed-response items, we may encounter a lot of challenges in implementing these very good ideas in our testing program. As a result, we often rely to some extent on what we call 'objective items'. So in our testing program, although we have a writing component, a translation component and also a separate speaking test, 70% of our items are objective items. Teachers sometimes use these objective items in class to train their students. Normal classroom teaching is sometimes replaced by what we call 'testing practice', practising multiple-choice items or other selected response items. To us, language tester developers, we have our concerns, our priorities, but to teachers and learners, the

compromises we make could mean negative washback. This is an example of what we call the social dimension. The other aspect is the moral issues or 'ethical issues' that endanger test fairness. For example, how do we combat cheating? This is a big issue not only for the CET, but also for other testing programs.

SY: Definitely.

JY: In recent years, with the rise of the stakes of the test, high-tech cheating has become a major concern of CET test developers. I would give you some idea of this high-tech cheating: test takers use communication devices such as mobile phones, needle cameras, invisible watches, mini ear-pieces and also bluetooth transmitters for sending and receiving messages in a test venue.

SY: Okay, so how do you deal with these issues then?

JY: As you probably know, in a computer-based test it's relatively easier to solve this problem because we can use an item bank or deliver the test in a more secure environment. But in a paper-based test, this is not possible. What we've been trying in the past several years is an innovative measure called 'Multiple Versions, Multiple Forms'. To obtain multiple versions, a practice we've been doing for decades, we reorder the options of multiple choice questions and we reorder the texts, and assemble them in different versions. Students are working on the same test questions, the same content materials, which are presented in different orders. This has proved effective for many years but later it was cracked by those people who wanted to cheat. Instead of sending messages like 'answers a, b, c, d' or 'answers 1, 2, 3, 4', they sent key words to test takers using cheating devices. We then started to develop multiple forms for each test implementation. In each test room, test takers are working on different texts and different test questions arranged in different orders. This measure has been adopted in

our paper-based test for four years now and feedback from national public security bureaus has shown that it is a very effective counter-cheating strategy.

SY: Is that difficult to administer in a practical sense?

JY: Yes, very right. This is very labour intensive and also very costly, a very expensive measure. We need a larger number of items and we need to pilot all these items, review them, analyse them, so that the quality of the items is ensured. And the other technical issue is to identify which form test takers take. We give each test taker a bar code and identify the version and the form for each test taker using this bar code. But you know, in the first administration, over 5000 test takers forgot to stick their bar code to their answer sheets.

SY: Oh no!

JY: What a mess!

SY: And so how did you then change that for the next test sitting?

JY: With the help of test centres, we resolve this issue manually. Later on we tried various means to inform students about the measure and now students all understand how to do it, so that's no longer an issue. But the technical side is equally challenging. For the marking of multiple forms, we have to use different sets of benchmark scripts to train markers. And there is also an issue with equating. We have to equate these multiple forms in terms of item difficulty.

SY: It sounds very complex.

JY: It is indeed very complicated. For each CET test, we have more than ten forms. Equating is performed to make sure that students are not advantaged or disadvantaged for taking an easy form or a difficult form. These are typical examples of the so-called social issues which test

developers are trying to solve through using our expertise in language testing.

SY: In terms of the CET, you've been chair of the National College English Testing Committee since 2004, is that correct?

JY: Yeah, it's been a long time, I know!

SY: What have you found most rewarding or interesting about the role?

JY: It is an interesting job, I have to say. I like this job, I like this career. The CET Committee has been trying its best to be responsible for test takers, for teaching and learning, and for society. So that is why I think the program can survive 30 years - this year is the 30th anniversary of the CET test.

SY: Congratulations!

JY: Thank you. The fact that this test has survived 30 years means that it has, to some extent, met the social needs. It has met the needs of teaching and learning. This is the rewarding part of my job, to witness the continuous development of our testing program, the reform of the test, the application of new technology, and so on. The social recognition of the value of this test and also teachers' and students' participation in and support to the test, I think, are the most rewarding aspect of my role as the chair of the CET Committee. And professionally I think it is very rewarding because this testing program gives me the opportunity to explore the field of language testing from both technical and social perspectives.

SY: It sounds rewarding and interesting.

JY: Yes but it is also a very challenging job.

SY: What do you find most challenging?

JY: As the developer of the CET, we are constantly under criticism for bringing a negative impact to teaching and learning. We are even criticised for compelling students to

cheat in exams, because if they cannot pass the test, they may lose their job opportunities.

SY: And also the test caters for so many test takers a year, doesn't it? So in that respect, it's a very large-scale test.

JY: It is a huge-scale test. We have two tests each year: one in June and the other in December. And for each test we have nine million test takers, so each year we have eighteen million test takers. It's a huge responsibility and a lot of work to operate the test. The testing committee is not responsible for the entire operation. The operational structure is like this: The committee works for the National Education Examinations Authority (NEEA), a government institution under the Minister of Education. NEEA is in charge of the test operation or the management of the test. The CET committee members are appointed by NEEA and we are responsible for the technical aspects of this test. We work on test design, item writing, item reviewing, piloting, rating scale development, quality control of rating, test form equating, and score reporting.

SY: Sure, and then someone different regulates the test centres, do they?

JY: That's right. They are the National Education Examinations Authority and local educational examinations authorities.

SY: So then who does CET-related research? Is that part of your role?

JY: Yes, it is part of the CET Committee's responsibility. NEEA also calls for research proposals and sponsor research projects; and CET is part of their research focus as well. The committee is also doing some very practically oriented research projects. For example, we started online marking in the early 2000s. The CET is the first large-scale test which used on-screen or online marking of the constructed response items. We also developed the internet-based CET tests,

including the item bank and the platform for test delivery. The internet-based CET was started in 2007 and assesses listening, reading, speaking, and writing. In recent years we have been collaborating with IT companies to develop automated scoring systems for essay writing, translation, and also speaking tasks. It is part of our job to conduct these very practically-oriented research projects.

SY: It sounds like you're doing a lot of interesting and important work.

JY: Yes, it's very interesting.

SY: You're currently in Melbourne and you're a visiting professor at the Language Testing Research Centre at The University of Melbourne. Could you talk about your collaboration with the LTRC?

JY: Yes. The LTRC is one of the earliest and most influential centres dedicated to language testing research and practice, so I have long been interested in getting to know more about this centre. When Prof. Tim McNamara and Dr. Ute Knoch invited me to apply for an Asian Scholar Program, which is funded by The University of Melbourne, I happily accepted and got this funding. The funding is for three years, 2016 to 2018. What I like most about this centre is that the research projects here are very much theoretically grounded and practically oriented. Studies here are always closely related to various operational testing programs like the Occupational English Test (OET), IELTS or TOEFL. So the research helps to solve practical issues, practical problems. When I'm here at the LTRC, I join their discussions on research projects, participate in doctoral students' discussions, and go to dry run sessions for conferences. I have also given talks on language assessment in China. Also, we've been collaborating on interesting projects. Last year we worked on an assessment literacy project, which is very interesting and new to me and I learned a lot.

SY: Lastly, as a PhD student myself, I would like to know what advice you have for those just beginning their career in language testing.

JY: From my experience, I think it is important for a language testing researcher to get yourself involved in testing practices, which will help you understand what you are doing, why you are doing this. I mean, theories about testing programs are important but the area of language testing is actually very empirically and practically oriented. If you are involved in language testing projects, it will motivate you to conduct more meaningful research. This will benefit the whole field and will give you continuous motivation to do research. The second thing is that you need to update your knowledge by extensive reading and also by going to conferences. The field of language testing is a quite small circle, although it is growing rapidly in recent decades. There are a couple of international and regional conferences. I strongly recommend that as a doctoral graduate, you need to participate in these conferences to get to know people in the field, to get to know the most important, urgent needs in the field. The third thing I would like to recommend is to expand your scope of interest gradually to enable yourself to better understand the construct and also conduct interdisciplinary research. I mean you need to join research groups or teams and broaden your scope of interest for interdisciplinary inquires, and this, I think, is the future of language testing. The application of new knowledge, the application of technology in language testing, I think, is key to the future development of our field.

SY: Thank you very much for the excellent advice.

JY: I hope it will be useful.

SY: Yes, definitely! And I'm sure that the other PhD students who read this interview will greatly appreciate your insights as well.

The 4th International Conference of the Asian Association for Language Assessment (AALA)

The AALA is a young but fast-growing association of language assessment from Asia. The fourth annual conference of the association took place at the National Taiwan University from June 22 to 23, 2017, and attracted more than 50 paper presentations and over 20 poster presentations around the globe. Despite the large amount of presentations, the two-day conference was exceptionally well-organised by the conference host, the Language Training & Testing Center (LTTC): individual papers were properly arranged into four strands, with each strand happening in a well-equipped lecture room.

The conference was opened by the AALA president, Prof. David Qian (Hong Kong Polytechnic University), who welcomed all conference delegates. This was followed by an invited speech (“Purposing Writing Assessment: Focusing Complex Constructs in Variable Contexts”) delivered by Prof. Alister Cumming (University of Toronto). The subsequent paper and poster presentations covered a variety of interesting topics, which resonated with the conference theme (*Connecting Assessment with Teaching and Learning: Innovation and Impact*) in general. Among them, the four thought-provoking plenary speeches are worth particular mention. In the first plenary speech, Dr. Nick Saville (Cambridge English Language Assessment) expounded and illustrated the concept of Learning Oriented Assessment. In the second one, Prof. Sebastian Hsien-hao Liao (National Taiwan University) discussed and argued for the localisation of language assessments from a cultural perspective. In his talk, Dr. Ping-cheng Yeh (National Taiwan University) illustrated his interesting idea of incorporating computer games into assessments to enhance learner motivation. In the last plenary speech, Prof. Yuko Goto Butler (University of Pennsylvania) discussed the usefulness of self-assessment for young learners based on a comprehensive review of past research.

Other highlights of the conference include the three pre-conference workshops given by Prof. Antony John Kunnan (University of Macau), Prof. Yasuyo Sawaki (Waseda University) and Prof. Alister Cumming (University of Toronto) on assessment literacy, Generalizability theory and writing assessments respectively, the two student awards (Best Paper and Best Poster), the student lunch meeting co-ordinated by the AALA Student Committee, and the reception dinner at the National Taiwan University History Gallery. The conference was closed by the AALA vice-president, Prof. Yan Jin (Shanghai Jiao Tong University), who also announced that the next conference would take place at Shanghai Jiao Tong University in October 2018. All in all, the fourth AALA conference offered all conference delegates a great platform for exchanging new ideas in the field and socialising.

Leo Xiaohua Liu, Doctoral Candidate, The University of Auckland, ALTAANZ Student Rep.



Delegates at the 4th ALAA Conference

ALTE 50th Meeting and Conference, September 2017

ALTE's 50th Meeting and Conference 2017 in Leuven, Belgium included, in addition to ALTE member meetings, an open conference day for a broader audience in language testing themed around the impact of language tests on education, migration and society. I was lucky enough to be in Europe at the time and took the opportunity to hear how the speakers and the other attendees approached these issues.

The plenary speakers focused on themes and issues that nudged the audience of language testers to consider a somewhat broader context than we may be used to examining. Professor Lourdes Ortega (Georgetown University) spoke first on the need for language testing to create stronger links to other disciplines, keeping the role of language testing situated in the wider world of the many uses, and users, of language tests. In particular, there needs to be more attention paid to those who are marginalized by the industry's focus on the relatively more educated and wealth 'elite multilingual'. Professor Jan Hulstijn (University of Amsterdam) spoke on the same issue from the opposite angle. His work on Dutch native speakers has focused on the core features that all native speakers have, regardless of opportunity, education, experience and other variables that can affect language use, range and mastery. How can the CEFR, which has built into it aspects of literacy, education and experience reconcile the idea of a 'core language proficiency'. The CEFR, Hulstijn reminded us, is built a large part upon the perception of teachers, and is weakened by this limited range of input, given the breadth of the use of the scales. The language we use in the curriculum is not the language we use in society; an easy fact to overlook, and yet one that qualifies a great deal of what language testing does.

This thread led us to Professor Constant Leung's (King's College London) discussion on language and identity and how this maps onto what is taught in the classroom. If one's whole linguistic identity as an English speaker is based solely on the transactional and socially predictable language of schooling, as is the case of many who grow up speaking another language outside of school, what is missing? Curricularised language excludes the language of intimacy and power that is learned in other social contexts, and if we teach, and then assess only curricularised language, we are placing boundaries around, and projecting values about what is and isn't 'important' language for success in society.

All of these issues flicker around broader questions of what is the right thing for language testers to do in order to do right by those being assessed. Dr Bart Deygers (KU Leuven) tackles this weighty problem by drawing from philosophical discussions on justice. We are often concerned with providing tests that are fair – free from any biases that might confound test results. We are less comfortable operating with the idea of justice, which requires assessment of the different circumstances by which test takers may come to the test. In a world where testing is a primary gatekeeper to a number of opportunities for people, we must consider whether a test, regardless of how empirically fair, is being used justly. Dr Deygers outline some basic principles by which test takers can begin to do so.

Professor James Simpson (University of Leeds) situated this broader issue of justice in testing to the context of language assessment for adult migrants, many of whom have little or no formal education. Language tests such as those that are required for naturalization and citizenship in the UK and beyond may place unjust demands on these marginalized groups, requiring skills and abilities that lie outside what can be arguably considered necessary for social integration, and that these migrants may not possess in their L1s.

Finally, Professor Kris van den Branden (KU Leuven), bringing the topic back to perhaps our most familiar ground – classroom assessment, spoke on how assessment is used in compulsory education. How can teachers use testing to positively affect a student's learning process, both cognitively and socially? Perhaps more importantly, in what ways does testing have a negative influence on students? The last session of the day, parallel workshops, gave the attendees the chance to explore issues such as these in a more hands-on way, learning from each other. The selection of workshops mirrored the main themes of the day – low-educated learners, core language abilities, multilingualism in and out of the classroom, and language use itself in and outside of the classroom. I attended the session held by Professor Leung, which examined the way in which young school children in the UK were assessed in their writing. The scales applied to both native speaker children and those in the process of learning English as well. How useful would feedback be to the latter group? What values are we extolling by focusing on certain linguistic features over another? Can assessment scales be used with both native speakers and language learners fairly or justly?

In all, the key take-away from the day was the need to ask such questions. In a world in which migration and multilingualism is becoming a norm and testing is becoming an important barrier to access, it is important that we consider these difficult questions to keep language testing firmly grounded in the current realities of education, migration, and government policy.

Michelle Czajkowski

Research Fellow - Language Testing Research Centre at The University of Melbourne

2017 Language Testing Research Colloquium

This year the Language Testing Research Colloquium (LTRC), the conference of the International Language Testing Association (ILTA), took place at the Universidad de Los Andes in Bogotá, Colombia, from July 17-21. LTRC 2017 was the first time ILTA's annual conference was held in South America. The theme of the conference was 'Language Assessment Literacy Across Stakeholder Boundaries', with a special invited plenary on 'Education, Language and Assessment in the Colombian Context' opening the second day of the Colloquium. The invited plenary speakers were Ana María Velásquez and María Lucia Casas, both from Colombian institutions. A number of presentations (Paper Sessions, Work-in-Progress and Posters) were also delivered by South-American researchers, providing delegates with interesting insights from contexts which are somewhat new to the language testing community. LTRC 2017 concluded with Ofra Inbar-Lourie's Davies Lecture in which she argued that the focus of the conference on language assessment literacy should encourage the reflection on identity issues within the language testing community.

ALTAANZ was represented by attendees from both Australia and New Zealand and some ALTAANZ members presented in the Colloquium. Ute Knoch, former co-president of our Association, reported on a research project carried out by her team at the Language Testing Research Centre in Melbourne looking at developing a rating scale for assessing writing skills at the workplace. Cathie Elder and Tim McNamara, both also former co-presidents of ALTAANZ, participated in the opening symposium, which brought together different perspectives on language assessment literacy. John Read facilitated a symposium on assessing the literacy skills of university students through post-admission assessments, with the participation of colleagues from various contexts: the U.S.A., Denmark, Colombia and South Africa.

LRTC 2017 was highly motivating and we hope the 2018 edition of the event will be just as stimulating. LTRC 2018 will be hosted by the University of Auckland, bringing language testing scholars and practitioners from around the globe down under. ALTAANZ members are strongly encouraged to attend the Colloquium, especially as there will be no independent ALTAANZ conference next year.

Morena Dias Botelho de Magalhaes

Doctoral Candidate, DELNA Administrator, The University of Auckland

Cathie Elder's report on the conference can be found at

<http://arts.unimelb.edu.au/ltrc/welcome/2017-language-testing-research-colloquium>

Pauline Cullen Interview

Author and test writer Pauline Cullen speaks with ALTAANZ Postgraduate Student Representative Megan Yucel.



Pauline Cullen is a freelance test writer and author of 5 IELTS preparation books including the *Official Cambridge Guide to IELTS* and two *Cambridge Vocabulary for IELTS* books. In 2016 Pauline launched her own IELTS Vocabulary teaching apps.

MY: Thank you for your time today, Pauline. Test item writing is a very specialised field. How did you first become interested in it?

PC: I was originally a French and Spanish teacher in a UK high school and so had to prepare classes for O levels and A levels, and I was also put in charge of teaching ESL within the school. When I moved to Australia I began teaching ESL fulltime, and I was naturally drawn to the exam preparation classes. This was in 1988, so pre-IELTS, and the main tests at the time were FCE and CPE, or TOEFL and TOEIC. The language centre where I worked steadily grew in size and I was very involved in testing and placement right from the start. I really liked that side of my role, so I was writing in-house tests for a long time before I was writing them professionally. As a teacher, I like the structure that a test gives to a class, as well as the fact that a formal test gives a course some sense of an 'end'. We would often have students studying with us for 6 months or more, and so

this idea of having completed a stage in the learning process was something I found really useful.

In the early 1990s, I became an oral examiner for Cambridge Main Suite exams, and I can remember being really interested in the idea of IELTS right from the beginning, because we had a lot of EAP classes, and we knew that IELTS would bring to those courses what FCE and CAE had brought to our GE courses.

I trained as an IELTS examiner in the very early days of the test. I'd been an examiner for about 2 years when I was approached to join the Australian IELTS test writing team in 1995. For my first commission, I had to write an Academic reading section. That first passage is now in *Cambridge IELTS Test Book 3*. I've been involved in test writing since then.

MY: There seems to be some controversy in testing circles about whether item writing can be seen as an art or science (see the discussion in Green and Hawkey, 2012, for example). On the one hand, item writing requires creativity and flexibility, while still keeping within the guidelines. On the other hand, item writers need to be fairly prolific, and able to quickly produce items that are standardised and reliable. What's your perspective on this?

PC: I think that this is a really interesting question and discussion to have. When it comes to perspective, in the last few years I've become more and more aware of just how important it is to understand the perspective of the person who is speaking or writing. Even when we're writing an academic textbook, the ideas and materials will always in some way reflect the writer's personal perspective - their background and experiences. So, I do think the different perspectives of people working in the language assessment field are interesting, and I think you will get different answers to that question depending on whether you are talking to a test writer or, for example, a statistician who has never had to write test

materials. I can only give my test writer perspective, which is that writing reliable test materials requires both science and art. If there wasn't a creative aspect, I certainly wouldn't have been able to stick with it for over 20 years.

The science comes from the technical aspects and, as you say, the need for standardisation. It takes years to learn and hone the skills to be able to produce test materials that are fair, valid, and reliable. The art (and perhaps the issue that is least understood) comes from the fact that the specifications and guidelines which item writers are given, are not a set of instructions or some sort of recipe that can easily be replicated with just a few variations. They are closer to a theory of how testing works that then needs to be applied in a skilful way to create something new each time. There is an art to that. We are repurposing texts, but to be able to do that, you need to be able to see the possibility within the original as well as the 'shape' of what you can create from it.

In listening, of course, the creative aspect is even more apparent, as we have to write and produce credible scenarios and scripts.

MY: What's your process when you sit down to write a test, for example, when you want to write a reading test? Can you share that with us?

PC: The test writing process cannot begin when you are asked to produce a test. The process has to begin weeks, often months, before. When a writer estimates that they would spend half a day, or a day, finding appropriate materials, they rarely mean that those hours were spent in one sitting. The time it takes is spread over many weeks, when we search habitually. We are never not looking for texts, and possible source materials. Because of that, any guesstimate about how long it takes to find a source text is wildly inaccurate, I think. When I first started, I used to constantly cut out and collect hundreds of articles, but then when it came time sit down to write a test, I would sift

through them only to find that they were not as test-friendly as they had appeared at first glance. As a new writer, it is very easy to be initially impressed by a text only to find that it actually contains only one or two testable ideas. So, I soon learned not to print or even save anything unless I could see 12 – 13 testable ideas. The most important work is done at this text selection stage, before the writing even begins. You learn from painful experience that, if you try to rush this phase, then you have a high chance of having your material rejected. What is worse, those are the texts that you need to work on for far too long to be able to produce items that you hope will work, only to have the material ultimately rejected. The main problem with a lot of the test materials we see online nowadays is that the writer has begun with a source text that is not suitable for the task. The writer persists in producing questions, but these simply don't work, because there aren't enough salient points within the material to test on. It's actually much more difficult than people imagine to find useable test materials.

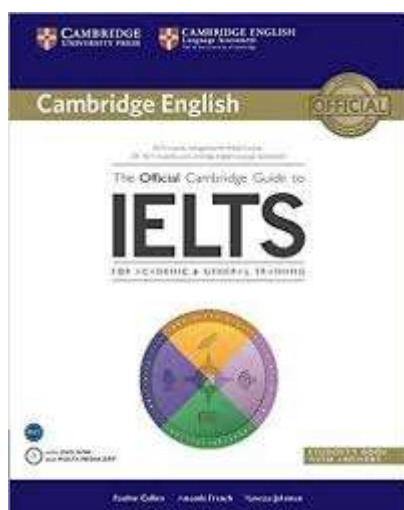
MY: Do you have any advice for teachers who might need to write their own tests?

PC: My first advice would be to use good models. Your aim should always be to produce materials that are fair, valid, and reliable. So, make sure you make a study of test materials that reflect this. Once you have found test materials that you trust, look carefully at the different question types and how they work.

It is easy to get the form, or the look, right, but if the questions don't perform their required function then they aren't useful – you will just end up with a list of questions rather than a testing tool. So, try to identify, or be aware of, the skills you have to use to get to the correct answer. Then, when trying to replicate that type of question, make sure you focus on testing those same skills, not simply on writing a question. You also need to be very aware of the difference between writing productive questions (where candidates need to write a word or words)

and objective test questions (where candidates choose a letter from a list). There are very different skills involved in writing these different question types.

New test writers also often produce questions that are what we call ‘tricksy’ or unfair. They often feel that this is necessary in order to create a test that is ‘difficult’. This is often because the source material is not suitable, and so to create an illusion of difficulty, the writer has to rely on trick questions. My final advice would be to practise, practise and practise, but make sure to get meaningful feedback. In order to produce test materials that are fair and reliable, you must get feedback from both students and from other language professionals. The difficulty comes in not seeing this feedback as criticism. A test will only be fair if others can see the same idea within the passage and interpret it in the same way as you. It is often a surprise to learn how others interpret (or misinterpret) your text or your questions. And it is important not to dig in and fight against their interpretations but to look again to see if you can make the idea clearer. Testing is very much about writing in a clear and precise way.



MY: As a published author of IELTS preparation materials, you're an expert on the IELTS test. I believe you also have an online platform which you use to communicate with students. Do you have any key pieces of advice that you give to students who are preparing to take the IELTS test?

PC: I started to be active on social media in 2012. That was because I was disappointed in the sales of several of my books and I wanted to find a way to help publicise them. My contact with candidates, students, and teachers from all over the world has taught me a lot about the user experience and goes a long way to inform the talks that I give at conferences or in webinars. I quickly realised that there is a lot of confusion and misinformation about IELTS. It will sound naïve, but I really thought that once people could contact an expert, that would solve the problem, and everyone could then focus on teaching and studying in the right way to prepare for the test. I was actually really shocked to find that people were more than ready to tell me I was wrong and point to other ‘experts’ who have attracted a far greater audience, and whose views and theories completely contradict what I tell them. The root of the problem, in my view, is the examples that don’t represent IELTS – what I refer to as inauthentic test materials. These materials often have guessable questions and represent the test as a confusing and tricky test. People have looked at these materials, which abound online, and drawn conclusions about the real test which are not accurate. They then give advice based on those conclusions, and produce even more examples to support their theories. So, a vicious circle is created, where the bad examples, lead to bad advice, which is supported through the production of more bad examples.

Nowadays, we have to be careful to filter what we read or see in terms of news stories, and the same applies to IELTS. So, my main advice to students is always to use only authentic test practice questions. The issue of test validity is an important one when we are writing test materials. It is the validity that produces the desired washback effect to the classroom or the test preparation of the individual student. So, it shouldn’t be surprising that, in my experience, when people use inauthentic materials, where this idea of test validity has not been a factor, their preparation does not go far enough and

so their test results do not improve. For example, in my experience, inauthentic reading test materials focus generally on matching vocabulary rather than testing reading skills, so students using these to practise do not develop the reading skills needed for the real test. Similarly, many of the writing tasks I have seen force students to write in a repetitive way and do not require them to take a position, so they don't develop the writing skills that are essential for the writing test. So, my number one tip for both teachers and students is to use authentic test materials as much as you can.

MY: Great advice. Out of all your publications, do you have any particular favourites?

PC: That's a tricky question! I am probably most proud of my two *Vocabulary for IELTS* books, because they really reflect my own teaching style and the way that I would teach an IELTS class myself. The two levels focus on building language, and I would love the chance to develop them into an IELTS course book together with the *Official Guide to IELTS*, which focuses on developing skills. But I am also very proud of my two IELTS vocabulary teaching apps, and the free e-book I am writing at the moment. Over the last 5 years I realised that one of the main problems people have is knowing how to study language at a high level. Many of the people I deal with are doctors, who have not studied language in a formal way for a very long time, yet have to achieve a minimum of band 7.5 in all skills, including writing, which they really struggle with. My free book helps show them how to become a language learner again, and draws on my own experiences when I struggled at B2 myself when I first began studying for my Advanced level exams in French and Spanish. I had to learn how to write essays on serious topics like crime, and the environment, and so on, just as they have to in writing task 2.

MY: Assessment literacy is a topic which seems to be attracting a lot of interest nowadays. In your experience, how much do you think teachers and students know about language assessment?

PC: Again, it's a very interesting point. When it comes to IELTS, I would say that the focus of the teachers and students I meet, either through my talks or through social media, is almost solely on scores and scoring – this is what I see in the questions I am asked. I constantly have to try to bring their focus back to language and the skills being assessed. This was the aim of the *Official Cambridge Guide to IELTS*, to show teachers and students how to prepare for each part of the test and to put the emphasis back on skills and language learning so that genuine progress could be made. The focus for teachers and for students is generally on looking for patterns within questions. This is the wrong approach if you want to make genuine progress.

I also think that perhaps there is a tendency to see the theory of language testing as somehow separate to the practice of language testing. So, while teachers may be very familiar with the guiding principles of language assessment from an academic point of view, they don't always equate this or relate this to their preparation and teaching. It's one thing to be aware of language assessment theory, but being able to apply this in the classroom in a meaningful way is a completely different matter. Again, that is what the *Official Guide to IELTS* aims to help teachers do.

It's really important not to see ideas like test validity, washback, and reliability as mere jargon. They really are at the forefront at every stage of IELTS test production. Understanding this means that you can begin to see that there is real benefit in teaching specific skills.

Again, a lot of the problems here may have come about from the use of inauthentic test materials, which don't follow, or reflect, these principles. In my view, those types of materials have had a very damaging effect on trust in the test, and perhaps on testing in general. If people are exposed to confusing and unreliable test materials, it is not surprising that their preparation becomes

ineffective and that they stop thinking in terms of the assessment of language and skills.

MY: Pauline, what trends do you expect to see regarding the future of language testing? What about test preparation, in terms of the modes in which instructional materials are delivered? As an author and test writer, how do these changes affect you?

In terms of test delivery, the future will surely lie in more computer-based testing. I welcome that, and I think it adds an extra dimension to language teaching and test preparation. Like a lot of writers, I now have to self-publish, and I'm already working on some materials that will be interactive and only available as e-books rather than a standard print version. That side of writing has always interested me, so it is something I'm enjoying working on.

I suppose that the future trends in language testing are following a pattern similar to other industries, and people will no doubt be hoping to create an algorithm that will take the place of the test writer. While it might be argued that this would produce a more standardised test, I would offer a counter argument that when we try to find some 'hidden formula' or a 'recipe' for producing any sort of content, the result will be predictable rather than 'standardised'. Once you have predictable testing materials, then you have an unreliable test, because the same 'formula' that is used to create such a test can be used to master it. What gives IELTS its high credibility now is the fact that there is no formula to follow – you just have to study, and master, the language and skills needed for academic or professional success.

MY: Thank you, Pauline, for an enlightening and thought-provoking discussion.

INAUGURAL ALTAANZ WEBINAR SERIES

Friday 20 th October	2pm NZ 12pm Sydney 11am Brisbane	Denise Angelo , Australian National University and Cath Hudson , Australian National University	<i>English language proficiency assessment in the school domain.</i>
Friday 27 th October	2pm NZ 12pm Sydney 11am Brisbane	Sally O'Hagan , University of Melbourne and Lynette May , Queensland University of Technology, PLTA Editors, Aek Phakiti , University of Sydney, Expert Abstract writer.	<i>Ancillary Writing for Language Assessment HDR Candidates</i> <i>Getting published in PLTA and getting your conference abstracts accepted.</i>
Friday 3 rd November* Please note different times.	5pm NZ 3pm Sydney 2pm Brisbane	Johanna Motteram , British Council, Singapore Nick Zhiwei Bi , University of Shanghai for Science and Technology	<i>Career Planning (or Career imagining) for Language Assessment HDR Candidates</i>
Friday 10 th November	2pm NZ 12 Sydney 11 Brisbane	John Read , The University of Auckland	<i>Assessing Vocabulary</i>

Revitalising Language Assessment:

Sustainable change through international benchmarking and knowledge transfer

The School of Languages and Cultures at The University of Queensland has won a Teaching Innovation Grant from UQ's Institute of Teaching and Learning Innovation to run a two-year project revitalising assessment practices across its eight language programs.

The project will provide additional descriptions of what students are capable of on course completion using "can do" statements from an internationally recognised set of standards called the Common European Framework of Reference for languages (CEFR). A significant advantage of the CEFR is that it is non-language specific and therefore applicable to all languages not just European ones.

Benchmarking the School's courses against the CEFR will deliver sought-after benefits to language students, including the important ability to describe language proficiency in terms that are widely understood by exchange organisations and future employers.

The benchmarking process will also be an opportunity for the School's academic staff to revisit the design of assessment tasks and marking criteria to ensure both alignment with the CEFR and students' ability to demonstrate their capacity to achieve real-world tasks.

The extensive experience of the School's academic staff in language assessment will be harnessed in workshop forums and then documented in a repository of sample assessment tasks. Cutting edge assessment tools will deliver proof students have reached a certain level in interaction as well as a consistent whole-of-School approach to language assessment, putting the UQ School of Language and Cultures at the forefront of standard setting in Australian higher education.

Researchers involved in the Revitalising Language Assessment Project at The University of Queensland

A/Prof. Marisa Cordella
Associate Professor in Spanish
Linguistics

Dr Juliana De Nooy
Senior Lecturer in French Studies

Dr Adriana Diaz
Lecturer in Spanish Studies

Dr Deanne Gannaway
Lecturer in Higher Education
Institute for Teaching & Learning Innovation

Dr Barbara Hanna
Senior Lecturer in French Studies

Dr Kayoko Hashimoto
Lecturer in Japanese Studies

Dr Noriko Iwashita
Senior Lecturer in Applied Linguistics

Dr Paul Moore
Lecturer in Applied Linguistics

Ms Alicia Toohey
Project Manager

Ms Monica Waung
Research Assistant