

Language Assessment Matters Issue 6

November 2016

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

In this issue we congratulate ALTAANZ members who have won local and international awards. We also have information about our upcoming AGM, our 2016 conference and the next issue of PLTA. In addition, Naoki Ikeda (University of Melbourne and past Student Representative on the ALTAANZ committee) reports on a lively language testing related workshop held in Melbourne in September. The workshop featured a presentation by Dr Alistair Van Moere of Pearson on automated language assessment.

ALTAANZ AGM 2016

The ALTAANZ Annual General meeting will take place on Friday 18 November from 1.20 to 2.25pm in Room OGGB 5 in the Owen G Glenn Building at the University of Auckland. 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 Presidents and Student Representatives to serve on the committee.

Further information, including our Presidents', Treasurer's and PLTA Editors' reports will be distributed soon. Please contact us through altaanz@gmail.com if you would like to nominate for either position.

Hobbiton,
Auckland.

A day trip
possibility if you
are visiting
Auckland for our
2016 Conference
later this month.

Enjoy!



Congratulations to the winners of the Papers in Language Testing and Assessment (PLTA) Best Article Award

It is our pleasure to announce the winner, runner-up and the finalist of the PLTA Best Paper Award 2013-2015.

Winner: Knoch, U., & Elder, C. (2013). A framework for validating post-entry language assessments (PELAs). *Papers in Language Testing and Assessment*, 2(2), 48-66.

Citation: This paper presents a significant and substantive step in the development of PELAs (Post-entry English Language Assessments) in Australia and New Zealand over the past 20 years. This paper is an extremely useful adaptation of the validity argument conceptualisation into a practical framework for validating PELAs. It provides broader applicability in terms of the discussion of validation/evaluation distinction. The framework will no doubt be influential in time for many institutions developing PELAs.

Runner-up: Hudson, C., & Angelo, D. (2014). Concepts underpinning innovations to second language proficiency scales inclusive of Aboriginal and Torres Strait Islander learners: a dynamic process in progress. *Papers in Language Testing and Assessment*, 3(1), 44-85.

Citation: This paper not only has an invaluable contribution to language assessment in the Australian context, but also practical implications for other similar contexts. It documents the development of an instrument which has arisen out of social and pedagogical need with considerable input from classroom teachers. This paper is an excellent example of how a rating scale can serve a professional development role and how assessment instruments might fit in the nexus of second language acquisition, descriptive linguistics, policy and education.

Other Finalists:

Clark, M. (2014). The use of semi-scripted speech in a listening placement test for university students. *Papers in Language Testing and Assessment*, 3(2), 1-26.

Ruegg, R. (2014). The effect of assessment of peer feedback on the quantity and quality of feedback given. *Papers in Language Testing and Assessment*, 3(1), 24-43.

Kokhan, K. & Lin, C-K. (2014) Test of English as a Foreign Language (TOEFL): Interpretation of multiple score reports for ESL placement. *Papers in Language Testing and Assessment*, 3(1), 1-23.

Selection Committee:

Associate Professor Aek Phakiti (Chair)

Associate Professor Angela Scarino

Dr Rosemary Wette

Dr Susy Macqueen

ALTAANZ 2016 THE UNIVERSITY OF AUCKLAND 17TH TO 19TH OF NOVEMBER



**In the classroom and beyond:
Assessing language ability in different contexts.**

Confirmed plenary speakers are:

Matt Poehner, Pennsylvania State University
Ute Knoch, University of Melbourne
Peter Keegan, University of Auckland
Barry O'Sullivan, British Council

Information on www.altaanzt.org.



Upcoming PLTA Issue.... Available online November 2016

Volume 5, Issue 2 includes:

Articles

- *Using dictation to measure language proficiency: A Rasch analysis.* Paul Leeming, Kindai University & Aeric Wong, Konan University
- *Testing Measurement Invariance of an EAP Listening Placement Test across Undergraduate and Graduate Students.* Soo Jung Youn, Northern Arizona University & Seongah Im, University of Hawai'i at Mānoa
- *Interaction in a paired oral assessment: Revisiting the effect of proficiency.* Young-A Son, Georgetown University
- *The Construct and Predictive Validity of a Self-Assessment Scale.* Jason Jinsong Fan, Fudan University/University of Melbourne

Book reviews

- *Language for Specific Purposes.* S. Gollin-Kies, D. R. Hall & S. H. Moore. Reviewed by Adam Fletcher Koschade, Flinders University
- *Post-admission language assessment of university students.* J. Read (Ed.). Reviewed by Michelle Czajkowski, University of Melbourne

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.

Forward Planner: Upcoming Language Assessment Events

ALTE 2016, Helsinki, Finland November 9-11. www.alte.org

ALTAANZ 2016, Auckland, New Zealand November 17-19 www.altaanz.org

UKALTA 2016, University of Reading, UK November 25 – 27 www.ukalta.org

BAAL TEA SIG 2017 University of Bedfordshire, UK March 31, <http://www.beds.ac.uk/baalteasig>
(cfp closes 30/12/2016, theme is *Innovations in EAP assessment*)

ALTE 2017, Bologna, Italy May 3-5
<http://events.cambridgeenglish.org/alte-2017/>
(cfp closed, theme is *Learning and Assessment: Making the Connections*)

AALA 2017, Taipei, Taiwan June 21 – 23 www.alaaweb.com
(cfp closes 13/11/2016 theme is
Connecting assessment with teaching and learning: innovation and impact)

Symposium on Second Language Writing 2017 Bangkok, Thailand June 30 – July 2
<http://sslw.asu.edu/2017/>

LTRC 2017, Bogota, Colombia July 17 – 20 <http://www.iltaonline.com/LTRC/index.php/ltrc2017>
(cfp closed, theme is *language assessment literacy across stakeholder boundaries*)

**Congratulations Kellie Frost on winning the Lado Award
(Best Graduate Student Paper) at LTRC 2016**



**Kellie Frost (left) and Award Committee Chair Maryam Wagner
June 2016, Palermo**

Congratulations to Kellie Frost and her supervisor, Professor Tim McNamara on this prestigious award.

Citation extract below, paper abstract on following page:

This year's winner emerged based on a presentation that delivered an extremely engaging narrative that left the committee members with a long-lasting impression. It was delivered confidently, clearly and precisely in an authoritative voice that captivated the audience. Of particular significance was the real-world impact of the research that prompted the audience to examine test-taker perspectives and also the dynamic nature of our political environment. The presenter delved into the details of the research investigation, while also providing us with a 'big picture perspective'. The implications of her study were significant and wide reaching. And...all of this was accomplished with the added pressure of opening LTRC with her presentation.

As you may have deduced, the committee is pleased to present this year's Robert Lado Memorial Award to Kellie Frost for her paper titled: "The Dynamics of Test Impact in the Context of Australia's Skilled Migration Policy: Reinterpreting Language Constructs and Scores".

The dynamics of test impact in the context of Australia's skilled migration policy:***Reinterpreting language constructs and scores***

The use of language tests as tools of immigration policy raises critical questions concerning the adequacy of current conceptualisations of test impact and validity in the field of language testing. Within existing theory, these notions rest on the premise that language constructs and associated test score meanings exist as coherent, fixed and stable entities. Test developers and test users are assumed to be in consensus concerning how test scores should be understood and used, and when tests and test scores are used as intended, the premise underlying evaluations of consequences is that positive effects on individuals and societies should outweigh or at least justify any negative effects. Such underlying assumptions, it is argued in this paper, limit our capacity to understand and explain the role of language tests in immigration (and other) policy domains, where, as will be demonstrated here, score meanings, test purposes and test consequences are dynamic, multiple, potentially conflicting and inherently unstable.

This paper explores the dynamics of test impact in the context of Australia's skilled migration policy by examining how four individuals responded to the language test score requirement that exists within the policy as they sought to transition from temporary visas to permanent residency status. Temporary visa holders in Australia are able to become permanent residents via the skilled migration program if they possess certain specified skills and attributes, including English language proficiency as demonstrated on an accredited language test, such as IELTS. In recent years, test scores of IELTS 7 and 8 have been heavily weighted in the overall selection process, and thus represent a vital transition mechanism in the trajectory of those seeking to move from the status of temporary to permanent resident in Australia.

A series of in-depth, open-ended interviews were conducted over an 18-month period with four participants during their migration trajectory from temporary to permanent residency, a transition that for each of them involved repeated language test attempts. A grounded theory approach guided analysis of interview data, which focused on identifying how participants interpreted score meanings throughout their trajectory, how they perceived test purposes, if and how their perceptions changed over time as they interacted with the test, how their thoughts and feelings influenced their behaviour, and the ways in which their actions and decisions produced test consequences.

Findings show that test impact is co-constructed by test takers in dynamic and unpredictable ways as they act and react according to changing perceptions of score meanings and test purposes in their attempt to practice agency under the constraints imposed by policy. Their shifting perceptions produce and are a product of their actions and decisions in response to the testing regime, which in turn lead to shifting test consequences. The study evidences the need for a more nuanced understanding of the experiences of individuals as they engage with language testing practices in the context of their migration journeys, in order to develop theoretical frameworks that can account for the role and impact of language tests in these policy domains.

Kellie Frost, University of Melbourne

Congratulations Sharon Yahalom Student Travel Prize Winner LTRC 2016



Sharon Yahalom and Tony Green

NURSES' PERSPECTIVES ON THE QUALITIES OF REFERRAL LETTERS: TOWARDS PROFESSION-ORIENTED ASSESSMENT CRITERIA

Good-quality written communication among overseas-trained health professionals is critical for accurate diagnosis, safe patient care and effective delivery of multi-disciplinary interventions (Vermeir et al., 2015). In Australia, the written communication of overseas-trained health professionals, including nurses, is assessed through use of the Occupational English Test (OET), a specific-purpose language (LSP) test. Nursing candidates are required to write a letter, usually one of referral, to another health professional (OET, 2014). In the field of LSP testing, test designers and researchers are becoming increasingly aware of the need to develop criteria that more accurately reflect work-related performance (Härmälä, 2010). Assessment criteria are more likely to be professionally relevant if specialists are involved in the decision-making process and their perspectives on what is required of candidates are considered (Pill, 2013). However, the current OET writing test assessment criteria were developed by language professionals without direct input from health

professionals. The aim of this study is to identify what constitutes an effective referral letter from the perspectives of nurses. The findings from this qualitative study will be subsequently used as a basis for reviewing the assessment criteria of the OET writing test.

Data include thirty interviews and six workshops conducted with nurses from two Australian hospitals – one rural and the other metropolitan. For the interviews, participants were asked about their perspectives on the qualities of effective referral letters and the advice they would give to newly registered health professionals to ensure that they write referral letters effectively. Thirty-four nurses took part in six workshops in which they commented on referral letters extracted from real hospital medical records from the same two hospitals. The aim of the workshops was to establish the valued features of referral letters.

The findings show that referral letters are formally written and generally follow a prescribed structure; however, the appropriateness and comprehensiveness of the letters vary. Nurses commented that referral letters must be legible, clear, concise and accurate; although these qualities are not always evident in the letters. The ability to effectively communicate about patients through the inclusion of relevant information in referral letters is considered to be more important than linguistic features such as grammar and spelling. Lack of standardisation of referral letters means that vital information necessary for high-quality patient care is sometimes omitted.

The findings of the study also show that the current assessment criteria for the OET writing test do not fully reflect nurses' criteria for effective referral letters, which has implications for the validity of the test. The findings contribute to the emerging field of research (e.g. Elder et al., 2012) which seeks to include specialists' perspectives about what is considered important in the review and development of assessment criteria for LSP tests.

Congratulations to Sharon Yahalom, and her supervisors, Professors Elizabeth Manias and Tim McNamara, University of Melbourne

Dr Alistair Van Moere in Melbourne

Automated Language Assessment Workshop



Noaki Ikeda attended Dr Van Moere's Automated Language Assessment Workshop at the University of Melbourne in September. Many members of the ALTAANZ committee were disappointed that we could not attend, so we asked Naoki to tell us about it.

The last week of September was an academically exciting week for us. Dr Alistair Van Moere delivered a talk about automated scoring of productive language performance at the University of Melbourne. The day was a school holiday for our university and it was very quiet on campus. But this quietness of the campus highlighted a contrast with our lecture room for the talk, which was full of audience including students, language teachers, and researchers in language testing and in linguistics. The talk started with a physically and academically lively atmosphere in the room, which was maintained through to the end of the lecture.

Dr Van Moere talked about the automated scoring system, which his institution, PEARSON uses for their tests. The talk included how the system was developed, test task formats, what features of language are scored and how they are scored.

The Versant test (<http://www.versanttest.com/>; retrieved on October 4, 2016), one of Pearson's tests, has been widely utilized by a number of institutions in various domains such as government, corporate, and education. In addition, although Dr Van Moere, did not mention it explicitly in his talk, the test can also be used for research purposes, to measure research participants' language abilities. I myself took Versant test (English version) as a research participant before, and I remember that I agreed with the result (my test score) as an indicator of my abilities that the test defined and measured.

Automated scoring is particularly significant in the view of reducing resource intensiveness and maintaining consistency of scoring. Our possible concerns would be “Can we rely on machine scores compared to scores provided by human rating?” With regards to this, Dr Van Moere showed high correlations between versant English test overall score and human grade automated scoring. Other major features that struck me in his talk were that under the automated scoring system used for the PTE Academic (another of PEARSON’s tests; <http://pearsonpte.com/>; retrieved on October 4, 2016), different aspects are rated separately (e.g., *Pronunciation* and *Content* scored as independent criteria respectively). In other words, the automated scoring system is not affected by test-takers’ pronunciation when scoring content. Also, the machine can systematically detect pauses in speech as an assessment feature for fluency. The pause pattern in speech was developed based on patterns of native speakers and the test is based on a native speaker model.

Toward the end of the talk, the audience raised two questions relevant to these two features of the machine scoring. The first question was: Do/Can we as human raters judge fluency by systematically detecting pause in the speech? In this regard, Van Moere and Downey (2016) states “Automated scoring technology does not make computers behave like humans.” (p. 341). I realized that both human scoring and machine scoring have respective strengths and shared strengths in scoring. The talk concluded with the audience’s general consensus of the current limitation of how to define native speakers (related to the second question), which is a challenge for the audience and for the field of language assessment.

There were other questions from the audience as Dr Van Moere allowed us to cut in his talk to give questions to clarify what he explained. Because a number of questions were raised (as this talk attracted interests of many audience) in this one-hour talk, and I was so absorbed in the talk, it felt like an hour passed very quickly. It seemed that in the talk, much time was spent on the details of how speaking performance is scored although at the beginning of the talk, assessment of writing was mentioned in what features were scored. My review here reflects my own understanding of the points relevant to speaking scoring provided in the talk. For the exact details, it is recommended that readers of this review refer to the websites above. It is also recommended to refer to Van Moere (2012) and Van Moere and Downey (2016) as listed below.

After Dr Van Moere’s talk, I recalled the interview that I did with Professor Tim McNamara for the third issue of the ALTAANZ newsletter. In the interview, Professor McNamara mentioned the significance of development of technology for language assessment from the past to the present, and automatic scoring as changing the future of the field. I took a look at the conference program of the first Language Testing Research Colloquium in 1979 to see whether there were any presentation titles that contain key words related to technology such “automated scoring” and so forth but I could not find any. It was absolutely inspiring to listen to Dr Van Moere’s talk about artificial intelligence scoring, which was an engaging, interactive presentation with many questions and insights from the audience and to think about the future of language assessment.

References

- Van Moere, A. (2012). A psycholinguistic approach to oral language assessment. *Language Testing*, 29 (3), 325-344.
- Van Moere, A. & Downey, R. (2016). Technology and artificial intelligence in language assessment. In D. Tsagari & J. Banerjee (Eds.), *Handbook of Second Language Assessment* (pp. 342-357). Boston: De Gruyter Mouton.