Innovative Uses of Eye-Tracking Technology in Language Assessment Research

New Directions Pre-Conference Colloquium
9:00-12:45 on Saturday 7 December at Yokohama National University

In association with:

Opening remarks: Sheryl Cooke (British Council East Asia Assessment Solutions Team)

Discussant: Dr Jamie Dunlea (British Council Assessment Research Group (ARG))

Invited speaker: Dr Vahid Aryadoust (National Institute of Education, Nanyang Technological University, Singapore)

Colloquium Schedule

Time	Activity
09:00-09:15	Opening Remarks by Sheryl Cooke (British Council East Asia Assessment Solutions Team (EAAST))
09:15-10:00	Talk by Dr Vahid Aryadoust (National Institute of Education, Nanyang Technological University, Singapore)
10:00-10:30	Presentation by Dr Nathaniel Owen (Open University, U.K.)
10:30-10:45	Break
10:45-11:15	Paper presentation by Dr Suh Keong Kwon (Korea Institute for Curriculum and Evaluation (KICE))
11:15-11:45	Paper presentation by Dr Aaron Olaf Batty (Keio University)
11:45-12:45	Discussion led by Dr Jamie Dunlea (British Council Assessment Research Group (ARG))

9:15-10:00 Dr Vahid Aryadoust

Vahid Aryadoust, PhD, is Assistant Professor in the National Institute of Education, Nanyang Technological University, Singapore. He is the Associate Director of the Global Listening Center, and a member of international associations such as the International Listening Association, American Association of Applied Linguistics, and the Cognitive Science Society. He has provided consultation on language assessment projects to, for example, Paragon Testing Enterprises (Canada), the DELTA Project of Hong Kong, and Learning Resource Network (London). Vahid has led multiple language assessment projects funded by, for example, Cambridge-Michigan Language Assessment (CaMLA) in 2013 and 2010, and published his research in Language Testing, Language Assessment Quarterly, Assessing Writing, Educational Assessment, Educational Psychology, and Computer Assisted Language Learning, etc. He has also (co)authored multiple book chapters and books published by Routledge, Cambridge University Press, Springer, Cambridge Scholar Publishing, Wiley Blackwell, etc. In addition, Vahid has served as the Principle Guest Editor of a special issue on assessing writing published in Educational Psychology (2017) and the Co-Principle Guest Editor of learners' listening special issue published in The International Journal of Listening (2016). He is a member of the Advisory Board of several international journals including Language Testing (UK, Sage Publisher), Language Assessment Quarterly (USA, Taylor & Francis), Educational Assessment (USA, Taylor & Francis), Educational Psychology (Hong Kong and UK, Taylor & Francis), and The Japan Association for Language Teaching's Testing & Evaluation SIG (Japan). His most current project is a two volume book entitled Quantitative data analysis for language assessment (Vol 1: Fundamental techniques; Vol 2: Advanced methods) to be published by Routledge.

10:00-10:30 Dr Nathaniel Owen

Investigating cognitive validity of reading tests using eye-tracking technology

This paper reflects upon the contribution of eye-tracking to validity studies of tests of reading for academic purposes to date, with specific reference to cognitive validity (Weir, 2005). It briefly outlines cognitive approaches to validation of reading tests before considering the contribution of eye-tracking studies to cognitive validity. It then reports on a study which uses Khalifa and Weir's (2009) model of cognitive processing to examine how different item types influence test takers' behavior using eye-tracking technology. Fourteen L2 English speakers were recruited. And asked to complete an authentic TOEFL iBT reading paper composed of one text and thirteen items. Once complete, each participant took part in a stimulated-recall interview in which they were asked to elaborate their item completion processes. Data revealed that inferencing item types caused greater backtracking (regressions) among participants than basic comprehension items, but there was no difference in the number of forward movements (saccades) or fixations. Examination of one 'reading to learn' item revealed

bespoke item completion procedures by participants. The implications of the findings and recommendations for future eye-tracking studies are discussed.

Key words: Eye-tracking, cognitive validity, non-parametric, stimulated-recall interviews.

Author's bio: Nathaniel Owen is a Research Associate at the Open University. He has a PhD in language testing and has published articles in peer-reviewed journals such as the International Journal of Research and Method in Education and book chapters in volumes such as The Routledge Handbook of English Language Teaching. He has experience of teaching English as a foreign language in the UK and Australia, with expertise in teaching EAP and exam preparation courses. He has previously worked for the examination board Cambridge Assessment and is currently participating in funded research projects with Educational Testing Service and the British Council.

10:45-11:15 Paper presentation: Dr Suh Keong Kwon

The cognitive validity of video-based listening comprehension test: an eye-tracking study

This study seeks to examine the extent to which inclusion of visual cues in a listening comprehension test has an impact on L2 learners' test-taking process and performance. Specifically, this study investigates the cognitive process of L2 test-takers during a video-mediated listening comprehension test by analysing their viewing behaviours. For this, an eye-tracking technology was employed as a main data collection method to investigate the degree to which test-takers view the visual cues and the question and answer choices. In total 57 EFL learners in five different high schools in Korea participated in this study and their listening test scores and eye-movement data are collected and analysed to examine the extent to which individual variances in viewing behaviours have an impact on the listening test performance. Findings indicate that test-takers in general spent a substantial amount of time viewing the visual cues. In addition, reading the question and the key option longer and more frequently is associated with improvements in their overall test score whereas reading the distractors is associated with drops in their overall test score. In terms of viewing the visual cues, significant negative effect of the speaker(s) and significant positive effect of the PPT slide were found. Overall, this study has corroborated meaningful original findings to support the inconclusive debates over the validity of inclusion of visual cues in a listening comprehension test.

Author's bio: Suh Keong Kwon is a researcher at Korea Institute for Curriculum and Evaluation (KICE) since 2012. His main research interests are language testing and assessment, technology assisted teaching and learning, and English education policy.

11:15-11:45 Paper presentation: Dr Aaron Olaf Batty

Eye-tracking attention to visual cues by item type in video-mediated L2 listening tests

Multiple-choice L2 listening items can be broadly categorized into two general item types: explicit and implicit. Explicit items require examinees to comprehend discrete linguistic features in order to answer correctly, whereas implicit items require a degree of inference. Much research has been conducted into these item types on audio-mediated listening tests, but video-mediated tests present a special case. Nonverbal communication research has demonstrated that listeners use the visual channel for gathering emotional or social context to aid interpretation of the verbal channel. As such, examinee attention to these nonverbal cues may differ by item type on listening tests.

No studies have sought to objectively and quantitatively track attention to specific nonverbal cues or to investigate its interaction with item type on video listening tests. The present research addresses this gap via a mixed-methods study employing quantitative eye-tracking measures augmented by stimulated recall interview data.

Twelve Japanese university students sat a six-item video-mediated English listening test while wearing an eye-tracking headset. Three items were explicit, and three implicit. After the test, the researcher conducted interviews of the participants while viewing their scanpath-overlaid videos. Due to the complex nature of the areas of interest, the eye-tracker output was subjected to manual scanpath analysis. The researcher manually coded the oculomotor events in the 72 videos at a resolution of one-tenth of a second according to a list of twelve visual cues of interest. The interview data were coded according to five main reasons for behavior identified in interviews.

Although most visual attention was focused on faces regardless of task type, the facial cues attended to in explicit and implicit items differed. Significant item-type-based differences in dwell times were observed in eight of the visual cues. Most effect sizes were medium to large. Respondents explained these behaviors as attempts to glean more information about the characters' motives and feelings through facial expressions, which they believed would aid in answering the implicit questions.

The findings demonstrate that L2 listening test items relying on inference, especially that of a social nature, are likely to be more affected by the inclusion of video than items relying strictly on linguistic comprehension, with theoretical implications for validity and practical implications for test development.

11:45-12:45 Discussion led by Dr Jamie Dunlea

Jamie works on a range of language test development and validation projects for assessment systems designed and developed by the British Council, as well as collaborating on projects with researchers and organisations internationally. Jamie has advised Ministries of Education and national agencies on large-scale assessment reform projects, overseen research for collaborative, international projects such as linking UK examinations to the China's Standards of English, and is active in the language assessment research community. He joined the British Council in 2013, and was previously Chief Researcher at the Eiken Foundation of Japan, a not-for-profit organization which develops and administers EFL examinations in Japan. He has 25 years of experience working in EFL education, first as a teacher, then in test development and production and assessment research.

Register by Friday 22 November

https://www.britishcouncil.jp/en/new-directions/workshops