

LISTENING TESTS

A fair assessment of language ability?

**GRANTS AND RESEARCH
ADVANCEMENT**

Acing subject-specific examinations is no longer enough to earn a spot in a university. Potential students are also expected to exhibit competency in the language medium the courses are taught in where the ability to listen is as important as the ability to read and write. As a result, language proficiency tests have gained traction in forming part of eligibility criteria for university admission. It is therefore not uncommon to see university hopefuls mugging and preparing for language proficiency tests such as the Test of English as a Foreign Language (TOEFL) or International English Language Testing System (IELTS) to enhance their dossiers. However, recent findings in a study from the National Institute of Education, Nanyang Technological University, Singapore (NIE NTU Singapore) suggests that it is perhaps time to re-evaluate the efficacy of such tests.

Led by Vahid Aryadoust, PhD, Associate Professor of Language Assessment, the study aimed to determine whether the viewing or reading behaviour of listening test takers affected test results. While listening comprehension tests are meant to test the listening skills of test takers, his team hypothesized that doing well in these tests depended on test takers' ability to read and answer questions instead of their ability to listen to the spoken language too.

In this study, 66 university students, aged between 20 to 40, took an on-screen listening test, comprising 11 items and an auditory lecture. The team used eye tracking technology, in the form of Tobii eye trackers, to track the moment-by-moment point of gaze and viewing patterns of the participants.

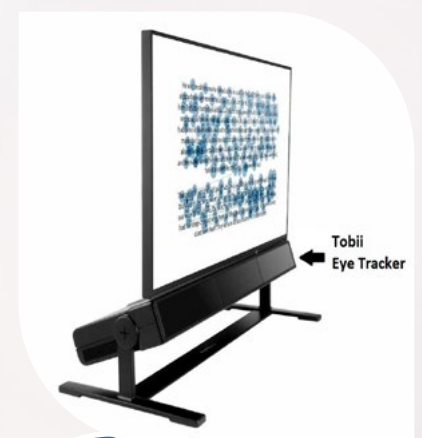
To test their hypothesis, the team used the Tobii eye trackers to measure fixation duration - the gaze behaviour of test takers and the duration of time taken. Fixation duration is used as a proxy to measure attention, capacity of working memory, or reading strategies. If the fixation duration is longer, the test taker is likely to be paying more attention to the word due to reasons such as its significance, its unfamiliarity to the test taker, et cetera.

The findings showed that a higher ability to identify keywords in a question plus a higher fixation duration did predict a high likelihood of the test taker answering the question correctly. In fact, the scale of the findings took the team by surprise, as there was evidence in many test items that participants relied on keyword matching to answer the questions. This points to a possibility of test takers gaming the system by sourcing for keywords to answer questions correctly.

The findings tallied with the team's observations in other high-stakes listening tests that demand a simultaneous listen-read-answer process, and all these served to support the team's hypothesis that listening test results were affected by test takers' ability to read and answer questions rather than listening skills alone.

Overall, the study suggested that the listening comprehension tests that demand too much reliance on keyword matching may not be the best way to assess one's listening skills. The fact that fixation duration predicted a higher possibility of a correct answer when a test-taker could identify keywords and listen out for keywords in test questions indicated that these tests may not, in effect, measure one's listening comprehension skills, contrary to its name and intended purpose. This also contradicts real-life behaviour; daily conversation or academic listening (e.g., listening in lecture theaters) does not involve identifying and listening out for keywords within a given piece of text while talking to someone.

With this, Dr Aryadoust categorically said, "I would suggest the following take-away from the study: Do not trust test scores or exam results, unless test developers provide compelling evidence to show that the tests / exams actually measure what they are intended to measure. In other words, if we 'dissect' the test scores or exam results, what ingredients can we find inside them, language skills or a significant amount of 'residue'? This is what the exam designer should be able to tell you in plain terms before you trust that their tools are trustworthy enough to, for example, determine the future academic life of your child".



An example of an eye tracking device by Tobii