PARLIAMENT OF AUSTRALIA

Answer to question:

PARLIAMENTARY JOINT STANDING COMMITTEE ON MIGRATION

MIGRATION, PATHWAY TO NATION BUILDING INQUIRY

QUESTION: Dr WEBSTER: I am interested in the adaptive tests. Do you worry about the risk of false positives? If it is adaptive, somebody may think that they are doing very well if the questions are becoming dumbed down, shall we say, and then they actually get a score that is not going to be acceptable for whatever level of English they need to have. How is that managed?

Mr Holden: That is a really good one. It is essentially managed by doing enormous amounts of piloting of those different items to make sure that they are calibrated appropriately. Another benefit of digital testing is that you can see all tests being taken in real time and, therefore, you can identify problematic items or

questions much faster than a traditional analysis of pen-and-paper tests.

In terms of that particular one—and I can only give examples of the way the Duolingo test works, which is one of dozens of digital tests—is that a test taker will get five calibration questions and that will determine roughly where they are at. Then it will go straight into what we say is the adaptive section. The calibrated section is pretty accurate and then the adaptive section will commence.

In terms of specific examples of false positives, I would probably like to take one on notice to give you a really concrete example. It is something that we look at and we have our measures for, but I would just say that it is due diligence, in making sure that the questions that go in there are producing the outcomes as they go.

JOINT STANDING COMMITTEE ON MIGRATION
PO Box 6021, Parliament House, Canberra ACT 2600 | Phone: (02) 6277 4560
Email: migration@aph.gov.au | www.aph.gov.au/mig

OFFICIAL

ANSWER:

The chance of getting 'false positive' results on the Duolingo English Test (DET) is extremely small, thanks to the computer-adaptive nature of the test. The initial impression about a computer-adaptive test like the DET may be that a test-taker with low English proficiency is misled by receiving easy items into believing that they will get a higher score when in reality, they receive a low score that reflects their proficiency level. However, this is not the case. In order to understand how the DET avoids the 'false positives', it is important to get acquainted with its structure.

The DET is administered in four phases: a calibration phase (randomly selected items of increasing difficulty), a picture description phase, a computer-adaptive test (CAT) phase, and a language performance phase. During the CAT phase the difficulties of the items administered to a given test taker depend on their performance on **each previous item**. In the "calibration" phase, test takers are presented with items which deliberately span a wide range of difficulty. A test taker's responses to these items are used to compute a provisional ability estimate, which informs the difficulty level of the item selected to **begin** the CAT phase (Cardwell et al., 2022). The provisional ability estimate is then **iteratively updated throughout the CAT phase** as the test taker responds to more items, so that the match of difficulty to individual proficiency improves as the CAT phase progresses.

Thus, because of the adaptive nature of the DET, the test constantly readjusts the initial, provisional language proficiency estimate that is established during the calibration stage. This means that if, for instance, a test-taker responds to an easier question correctly, they will then receive a more difficult item. If this more difficult item is also answered correctly, the next item will be of an increased difficulty, as well. When a test-taker with low English proficiency answers a more difficult item incorrectly (i.e., reaches their proficiency threshold), their proficiency estimate is re-adjusted and the next item they see is of lower difficulty (and vice-versa). In other words, the automated scoring engine monitors test-takers' every correct and incorrect response and constantly adjusts the difficulty level accordingly. This means that it is highly unlikely that a test-taker will answer all of the questions correctly and still receive a low score. Therefore, it is also unlikely that a low proficiency test-taker will have a false expectation of performing really well on the DET and receiving a high score.

It is also important to know that although the adaptive algorithm ensures that high-proficiency test takers are given systematically more difficult items than

JOINT STANDING COMMITTEE ON MIGRATION
PO Box 6021, Parliament House, Canberra ACT 2600 | Phone: (02) 6277 4560
Email: migration@aph.gov.au | www.aph.gov.au/mig

2

OFFICIAL

lower-proficiency test takers, the IRT models used to compute scores account for these systematic differences. Thus, even though no two DET test takers have ever received the exact same set of items, test scores are comparable.

The Duolingo English Test is able to monitor for test bias or any inconsistencies in scoring algorithms (such as false positives) in real time through its AI-powered Analytics for Quality Assurance in Assessment (AQuAA) dashboard. The Duolingo English Test uses AQuAA to monitor, in real-time, the performance of every test taken. This includes not only the performance of test sessions across the general test taker population, but also by various personal characteristics of the test taker, including gender, country, first language and intent in taking the test. In this way we could rapidly and effectively identify the emergence of events such as 'false positives' if they were to occur.

Finally, another guard against 'false positives' on the test is that in addition to the computer-adaptive part of the test, the DET also includes two sections where the items given to test takers are selected randomly from an extremely large pool of items available. These items include open-ended production tasks that are scored after the test has been taken by automated scoring models (e.g., Speak/Write About the Photo; Read, then Speak/Write; Listen, then Speak; Speaking/Writing Sample; Interactive Listening Summarization). These tasks do not have a clear-cut correct/incorrect answer and are evaluated on a range of language categories including content, discourse coherence, vocabulary, grammar, fluency (for speaking only), and pronunciation (for speaking only). Simply choosing a correct answer in these tasks is not an option; therefore, the complexity of the task makes it harder for test-takers to 'deceive' themselves about their language ability and receive a 'false-positive' impression about their performance.

References

Cardwell, R., Naismith, B., LaFlair, G.T., Nydick, S. (2022). Duolingo English Test: Technical Manual. *Duolingo Research Reports*. *https://duolingo-papers.s3.amazonaws.com/other/technical_manual.pdf*

OFFICIAL Answer to question:

PARLIAMENTARY JOINT STANDING COMMITTEE ON MIGRATION

MIGRATION, PATHWAY TO NATION BUILDING INQUIRY

QUESTION:

Mr Holden: That's a great question for stimulating thinking about the policy space. There was a recent EOI process that was put forward by Home Affairs last year in select English tests—not that digital tests were allowed in that process—and they convened an expert panel of three members from the English language testing community and academics in Australia. In that same vein you could try and set up a very independent panel. The context of independent panels is always challenging when an area is quite narrow, and so people are often on boards of different competitor organisations. You'd obviously expect a level of integrity there anyway, but that is a challenge in a small industry where there is a lot at stake about how you guarantee the independence of that and the size. But I think you've highlighted a good question that I probably don't have a great answer for. I know we've had a jobs and skills organisation set up that this could sit within as we look into English language testing and policy, but we certainly need to look at the best way to do it. **What I can probably do on notice is actually provide you with what other countries are doing because we're working**—
CHAIR: Yes, I think that would be very useful.
....
CHAIR: When you do that could you **please have a look at what is known as the 'everyday language requirements'** that would be suitable for people who will go out into—I'm thinking in the deputy chair's electorate. I think the people in Robinvale have a better grasp of what kind of English capacity their workers will require— also, many people do learn on the job and while they're living here—as opposed to, perhaps, Professor Smith from Melbourne University, my alma mater, who might have a different idea—but certainly more of an idea—of what English standard students and PhD students should
have. So, that's what I'm thinking. Could you **come back to us with some sort information about who measures the functionality—the score for functional English on a daily basis as opposed to one that's required to write a PhD or something.** If you can give us some ideas on that, it would be very useful.

JOINT STANDING COMMITTEE ON MIGRATION
PO Box 6021, Parliament House, Canberra ACT 2600 | Phone: (02) 6277 4560
Email: migration@aph.gov.au | www.aph.gov.au/mig

OFFICIAL

ANSWER:

1. Competitor Analysis

Below is a table of the responsible parties in determining language proficiency requirements for immigration in other countries. As you can see from the information in the table, language policy currently resides within the respective immigration departments. The global trend towards digitalisation of language proficiency testing in the immigration process is becoming increasingly evident. For example leading immigration nations have recognised the potential of digital testing solutions and have incorporated them into their higher education immigration policies through two specific ways:

In Canada, the United Kingdom, the United States and New Zealand, the respective immigration authorities do not require an English test for higher education student visas, provided that the applicants have an offer from an education institution (and thus have met their requirements). They simplify and expedite the process by delegating responsibility for English language to higher education institutions, who rely on comprehensive academic boards and both English langue and admissions professionals to determine appropriate English requirements.

In Ireland, the immigration authority does require applicants to submit evidence of English testing, however they have adopted digital tests as a way to demonstrate this evidence.

In contrast, Australia's current policy, which in addition to university English requirements, also mandates applicants provide proof of English proficiency through only traditional, non-digital testing methods, is limiting our ability to compete for global talent. This policy stands as a unique hurdle amongst our progressive international counterparts, deterring highly capable international students who are faced with a more accommodating process elsewhere.

Country	Who is responsible for language requirements for immigration?	English test required for higher education student visa?	Students can meet English requirements with a digital test?
Canada	Immigration, Refugees and Citizenship Canada (IRCC)	Not required, as long as applicants have an offer from an education institution	Yes
United Kingdom	UK Visas and Immigration (UKVI)	Not required, as long as applicants have an offer from an education institution	Yes
New Zealand	Immigration New Zealand (INZ)	Not required, as long as applicants have an offer from an education institution	Yes
USA	Department of Homeland Security	Not required, as long as applicants have an offer from an education institution	Yes
Ireland	Irish Naturalisation and Immigration Service (INIS)	Requires applicants to provide English test in addition to an offer from education institution. Digital tests accepted.	Yes
Australia	Department of Home Affairs (DHA)	Requires applicants to provide English test in addition to an offer from education institution. Digital tests not accepted.	No

2.   Measuring Functional English

Apart from standardised tests, there are other ways that immigration departments may measure functional language proficiency, although these methods are generally less common. Here are some alternatives that some countries may use:

University Degrees or High School Certificate: Some countries accept evidence of completion of education taught in English as proof of functional English proficiency. The assumption is that if an applicant was able to successfully complete an educational program in English, they should have a sufficient command of the language. This is often the case for individuals who have studied in English-medium universities in countries where English is the official or a dominant language.

Work Experience: In some cases, relevant work experience in a profession that requires English language proficiency can be used to demonstrate language skills. This often applies to skilled workers who have worked in English-speaking environments.

Language Courses: Some countries may offer language courses to immigrants, either before they arrive or once they are in the country. These courses often include an assessment component, which can be used to gauge the individual's language proficiency.

Interviews: In some cases, an interview with an immigration officer can be used to assess language proficiency. This is typically used in conjunction with other methods, and the interview may be conducted in English to assess the individual's ability to communicate effectively.

Exemptions for Long-Term Residents or Older Adults: In some countries, long-term residents or older adults may be exempt from certain language requirements. For example, in the United States, applicants for naturalization who are 50 years old and have lived in the U.S for at least 20 years are exempt from the citizenship English language requirement.

However, standardized language tests remain the most common and internationally recognized method of assessing language proficiency for immigration purposes. They offer a systematic, scalable and uniform way of evaluating language skills across different contexts and countries.

3.   Pilot Proposal

As a path forward, Australia could consider a pilot program that allows digital English testing for higher education student visas. This pilot would not only align Australia with global best practices, but it would also provide Australian universities with the tools to compete globally for talent. Furthermore, it would leverage existing policies that theoretically permit the use of digital English tests, thereby necessitating minimal policy changes. The pilot could focus on a specific cohort of university student visas, ensuring it targets low-risk, high-priority migrants and provides a rich dataset for the Department of Home Affairs. The data obtained from this pilot could inform the digitisation of other visa categories in regards to, for example, functional English requirements.

4.   Enhance the Migrant Experience

The immigration departments mentioned in the table above are responsible for determining english language requirements for both academic and functional english. Those agencies in other countries have a mandate to enhance the migrant experience, and Australia should likewise identify ways to do so. For example, DHA or Jobs and Skills Australia, if given a mandate to enhance the migrant experience, may be able to take learnings from a pilot of digital testing for student visas and apply it towards other visas as well that require functional English. By making the migrant experience more

equitable and less burdensome, Australia is able to attract a larger pool of potential migrants to choose from.

Standardised tests remain the gold standard for assessing language proficiency for immigration purposes. By embracing digital testing solutions, Australia could attract more global talent and reinforce its reputation as a forward-thinking destination who cares about the migrant experience.