Duolingo English Test:Technical Manual

Geoffrey T. LaFlair* and Burr Settles*

Abstract

The Duolingo English Test Technical Manual provides an overview of the design, development, administration, and scoring of the Duolingo English Test. Furthermore, it reports on test taker demographics and the statistical characteristics of the test. This is a living document and will be updated regularly (last update: October 22, 2020).

Contents

1 Duolingo English Test 3
2 Accessibility 3
3 Test Administration and Security 5
   3.1 Test Administration 5
   3.2 Onboarding 5
   3.3 Administration Rules 5
   3.4 Proctoring & Reporting 6
4 Test Taker Demographics 6
5 Item Descriptions 8
   5.1 C-test 8
   5.2 Yes/No Vocabulary 8

* Duolingo, Inc.

Corresponding author:
Geoffrey T. LaFlair, PhD
Duolingo, Inc. 5900 Penn Ave, Pittsburgh, PA 15206, USA
Email: englishtest-research@duolingo.com
1 Duolingo English Test

The Duolingo English Test is a measure of English language proficiency for communication in English-medium settings. It assesses test taker ability to use language skills that are required for literacy, conversation, comprehension, and production. The test has been designed for maximum accessibility; it is delivered via the internet, without a testing center, and is available 24 hours a day, 365 days a year. It has been designed to be efficient. It takes about one hour to complete the entire process of taking the test (i.e., onboarding, test administration, uploading). It is a computer-adaptive test (CAT), and it uses item types that provide maximal information about English language proficiency. It is designed to be user-friendly; the onboarding, user interface, and item formats are easy to interact with.

This document provides an overview of the design of the Duolingo English Test. It contains a discussion of:

- the test’s accessibility, delivery, proctoring and security processes;
- the demographic information of the test taking population;
- the test’s items, how they were created, and how they are are delivered and scored;
- and the statistical characteristics of the test.

The test scores are intended to be interpreted as reflecting test taker English language ability and used in a variety of settings, including for university admissions decisions.

2 Accessibility

Broad accessibility is one of the central motivations for the development of the Duolingo English Test. Tests administered at test centers consume resources which limit accessibility: they require appointments at a physical testing center within certain hours on specific dates (and travel to the test center), and carry considerable registration fees. The AuthaGraph* (Rudis & Kunimune, 2020) maps in Figure 1 show the concentration of test centers in the world (top panel) compared to internet penetration in the world (middle panel), and the concentration of Duolingo English Test test takers (bottom panel; for all tests administered since August 1, 2017). The top two panels of Figure 1 show a stark difference in how much more easily an internet-based test can be accessed than a test center†. While the ratio of population to internet access and to test center access is a somewhat limited metric—not every internet user has access to a device that can run the Duolingo English Test, physical test centers can usually handle dozens of test-takers at

---

* [https://en.wikipedia.org/wiki/AuthaGraph_projection](https://en.wikipedia.org/wiki/AuthaGraph_projection)
† Central Africa is underserved by both models.
once, and not all people need to take an English language proficiency assessment—it is still clear that the potential audience for the Duolingo English Test is orders of magnitude larger than those who could be served currently by more traditional test centers. The map in the bottom panel shows that the Duolingo English Test is beginning to realize this potential with for people taking the Duolingo English test from places with relatively low concentrations of test centers, such as countries in South and Central America (Colombia, French Guiana, and Guatemala); in Central and East Asia (Kazakhstan and China); and Central and East Africa (Central African Republic and Zimbabwe). By delivering assessments on-demand, 24 hours a day, to an estimated 2 billion internet-connected

Figure 1. Heatmaps of test center accessibility (top), internet accessibility (middle), and concentration of Duolingo test takers (bottom)
devices anywhere in the world for US$49, we argue that the Duolingo English Test holds the potential to be the most accessible, valid, and secure language assessment platform in the world.

3 Test Administration and Security

The Duolingo English Test is administered online, via the internet to test takers. The security of Duolingo English Test scores is ensured through a robust and secure onboarding process, rules that test takers must adhere to during the test administration, and a strict proctoring process. All of these procedures are evaluated after the test has been administered and prior to score reporting.

3.1 Test Administration

Test takers are required to take the test alone in a quiet environment. The Duolingo English Test can be taken in the Chrome and Opera browsers worldwide. In China, the test can be taken on the the 360 and QQ browsers as well. An internet connection with at least 2 Mbps download speed and 1 Mbps upload speed is recommended for test sessions.

3.2 Onboarding

Before the test is administered, test takers complete an onboarding process. This process checks that the computer’s microphone and speaker work. It is also at this time that test taker identification information is collected, that test takers are informed of the test’s administration rules, and that test takers agree to follow the rules.

3.3 Administration Rules

The behaviors that are prohibited during an administration of the Duolingo English Test are listed below. In addition to these behavioral rules, there are rules for test taker internet browsers. The browsers are locked down after onboarding, which means that any navigation away from the browser invalidates the test session. Additionally, all browser plugins must be disabled. Test takers are also asked to be visible at all times to their cameras and to keep their camera and microphone enabled through the test administration.

- Leaving the camera preview

\[1\] More information can be found in the Security, Proctoring, and Accommodations whitepaper.
• Looking away from the screen
• Covering ears
• Leaving the web browser
  – Leaving the window with the cursor
  – Exiting full-screen mode
• Speaking unless instructed
• Communicating with another person at any point
• Allowing others in the room
• Using any outside reference material
• Using a phone or other device
• Writing or reading notes

3.4 Proctoring & Reporting

After the test has been completed and uploaded, it undergoes a thorough proctoring review using human proctors with TESOL/applied linguistics expertise, which is supplemented by artificial intelligence to call proctors’ attention to suspicious behavior. This process takes no more than 48 hours after the test has been uploaded. After the process has been completed, score reports are sent electronically to the test taker and any institutions they elect to share their scores with. Test takers can share their scores with an unlimited number of institutions.

4 Test Taker Demographics

In this section, test taker demographics are reported. During the onboarding process of each test administration, test takers are asked to report their first language (L1), date of birth, and their gender identity. Their country of residence is logged when they show their proof of identification during the onboarding process. There were 99,451 people who took certified Duolingo English Tests between July 1, 2019 and June 30, 2020.

Test takers are asked to report their L1s during the onboarding process. The most frequent first languages of Duolingo English Test test takers include Mandarin, Spanish, Arabic, English§, French, and Portuguese (see Table 1). There are 132 unique L1s represented by test takers of the Duolingo English Test and the test has been administered to test takers from 186 countries. The full tables of all test taker L1s and countries of origin can be found in the Appendix (Section 9).

Reporting gender identity during the onboarding process is optional, but reporting birth date is required. Table 2 shows that 50.57% of Duolingo English Test test takers identified

§50% of these test takers come from India and Canada
Table 1. Most Frequent Test Taker L1s

<table>
<thead>
<tr>
<th>First Language</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese - Mandarin</td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td></td>
</tr>
<tr>
<td>Arabic</td>
<td></td>
</tr>
<tr>
<td>English</td>
<td></td>
</tr>
<tr>
<td>French</td>
<td></td>
</tr>
<tr>
<td>Farsi</td>
<td></td>
</tr>
<tr>
<td>Portuguese</td>
<td></td>
</tr>
<tr>
<td>Hindi</td>
<td></td>
</tr>
<tr>
<td>Korean</td>
<td></td>
</tr>
<tr>
<td>Russian</td>
<td></td>
</tr>
</tbody>
</table>

As female, 48.68% of test takers identified as male, and 0.75% chose not to report. Table 3 shows that 82% of the Duolingo English Test test takers are between 16 and 30 years of age.

Table 2. Counts and Percentage of Test Taker Gender

<table>
<thead>
<tr>
<th>Gender</th>
<th>n</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>50,292</td>
<td>50.57%</td>
</tr>
<tr>
<td>Male</td>
<td>48,409</td>
<td>48.68%</td>
</tr>
<tr>
<td>Other</td>
<td>750</td>
<td>0.75%</td>
</tr>
<tr>
<td>Total</td>
<td>99,451</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

Table 3. Counts and Percentage of Test Taker Age

<table>
<thead>
<tr>
<th>Age</th>
<th>n</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 16</td>
<td>5,131</td>
<td>5.16%</td>
</tr>
<tr>
<td>16 - 20</td>
<td>33,932</td>
<td>34.12%</td>
</tr>
<tr>
<td>21 - 25</td>
<td>33,992</td>
<td>34.18%</td>
</tr>
<tr>
<td>26 - 30</td>
<td>13,416</td>
<td>13.49%</td>
</tr>
<tr>
<td>31 - 40</td>
<td>10,368</td>
<td>10.43%</td>
</tr>
<tr>
<td>&gt; 40</td>
<td>2,612</td>
<td>2.63%</td>
</tr>
<tr>
<td>Total</td>
<td>99,451</td>
<td>100.00%</td>
</tr>
</tbody>
</table>
5 Item Descriptions

The test has seven different item types, which collectively measure test taker ability to use language skills that are required for literacy, conversation, comprehension, and production. Because the Duolingo English Test is a CAT, it will adjust in difficulty as the computer updates its real-time estimate of test taker language proficiency as they progress through the test. There are five item types in the computer-adaptive portion of the test. The CAT item types include c-test, audio yes/no vocabulary, visual yes/no vocabulary, dictation, and elicited imitation. During each administration of the Duolingo English Test, a test taker will see at minimum three of each CAT item type and at maximum of seven of each CAT item type. The median rate of occurrence of the CAT item types across all administrations is six times per test administration. Additionally, test takers respond to four writing prompts and four speaking prompts. They are not a part of the computer-adaptive portion of the test. However, the writing and speaking prompts also vary in difficulty, and their selection is based on the CAT’s estimate of test taker ability. These items work together to measure test taker English language proficiency in reading, writing, listening, and speaking.

5.1 C-test

The c-tests provide a measure of test taker reading ability (Khodadady, 2014; Klein-Braley, 1997). In this task, the first and last sentences are fully intact, while words in the intervening sentences are “damaged” by deleting the second half of the word. Test takers respond to the c-test items by completing the damaged words in the paragraph (see Figure 2). Test takers need to rely on context and discourse information to reconstruct the damaged words (which span multiple vocabulary and morpho-syntactic categories). It has been shown that c-tests are significantly correlated with many other major language proficiency tests, and additionally are related to spelling skills (Khodadady, 2014).

5.2 Yes/No Vocabulary

This is a variant of the “yes/no” vocabulary test (Beeckmans, Eyckmans, Janssens, Dufranne, & Van de Velde, 2001). Test takers are presented with a set of English words mixed with pseudo-words that are designed to appear English-like, and must discriminate between them. Such tests have been used to assess vocabulary knowledge at various CEFR levels (Milton, 2010), and have been shown to predict language proficiency skills—the text version (see top panel in Figure 3) predicts listening, reading, and writing.

We use an LSTM recurrent neural network trained on the English dictionary to create realistic pseudo-words, filtering out any real words, acceptable regional spellings, and pseudowords that orthographically or phonetically resemble real English words too closely.

© 2020 Duolingo, Inc
abilities; while the audio version (see bottom panel in Figure 3) predicts listening and speaking abilities in particular (McLean, Stewart, & Batty, 2020; Milton et al., 2010; Staehr, 2008). These tests typically show a large set of stimuli (say, 60 words and 40 pseudo-words) of mixed difficulty at once. The format is made computer-adaptive by successively presenting multiple sets (items/testlets), each containing a few stimuli of the same difficulty (e.g., B1-level words with pseudo-words that should be B1-level if they existed; more on how this is done in Section 6.1).

5.3 Dictation

In this exercise, test takers listen to a spoken sentence or short passage and then transcribe it using the computer keyboard (see Figure 4). Test takers have one minute in total to listen to and transcribe what they heard. They can play the passage up to three times. This assesses test taker ability to recognize individual words and to hold them in memory long enough to accurately reproduce them; both are critical for spoken language understanding (Bradlow & Bent, 2002; Buck, 2001; Smith & Kosslyn, 2007). Dictation tasks have also been found to be associated with language learner intelligibility in speech production (Bradlow & Bent, 2008).

\[\text{Figure 2. Example C-test Item}\]

\[\text{Autocomplete, spell-checking, and other assistive device features or plugins are detected and disabled.}\]
5.4 Elicited Imitation (Read-aloud)

The read-aloud variation of the elicited imitation task—example in Figure 5—is a measure of test taker reading and speaking abilities (Jessop, Suzuki, & Tomita, 2007; Litman, Strik, & Lim, 2018; Vinther, 2002). It requires the test takers to read, understand, and speak a sentence. Test takers respond to this task by using the computer’s microphone to record themselves speaking a written sentence. The goal of this task is to evaluate intelligible speech production, which is affected by segmental/phonemic and suprasegmental properties like intonation, rhythm, and stress (Anderson-Hsieh,
Furthermore, intelligibility is correlated with overall spoken comprehensibility (Derwing & Munro, 1997; Derwing et al., 1998; Munro & Derwing, 1995), meaning that this item format can capture aspects of speaking proficiency. We use state-of-the-art speech technologies to extract features of spoken language, such as acoustic and fluency features that predict these properties (in addition to basic automatic speech recognition), thus evaluating the general clarity of speech.

5.5 Extended Speaking

The extended speaking tasks are measures of test taker English speaking abilities. At the end of the CAT portion of the test, the test takers respond to four speaking prompts: one picture description task and three independent speaking tasks, two with a written prompt and one with an aural prompt (see Figure 6). Each of the task types have items that are calibrated for high, intermediate, and low proficiency levels. The difficulty level of the tasks that test takers receive is conditional on their estimated ability in the CAT portion of the test. All of these task types require test takers to speak for an extended time period and to leverage different aspects of their organizational knowledge (e.g., grammar, vocabulary, text structure) and functional elements of their pragmatic language knowledge (e.g., ideational knowledge) (Bachman & Palmer, 1996)
5.6 Extended Writing

The extended writing tasks are measures of test taker English writing abilities. Test takers respond to four writing prompts that require extended responses: three picture description tasks and one independent task with a written prompt (see Figure 7). Similar to the speaking tasks, these are drawn from different levels of difficulty conditional on the estimated ability level of the test taker. The stimuli in the picture description tasks were selected by people with graduate-level degrees in applied linguistics. They are designed to give test takers the opportunity to display their full range of written language abilities. The independent tasks require test takers to describe, recount, or make an argument; these require the test takers to demonstrate more discursive knowledge of writing in addition to language knowledge (Cushing-Weigle, 2002).

6 Development, Delivery, & Scoring

This section explains how the computer-adaptive items in the test were developed, how the computer-adaptive test works, and how the items are scored. Additionally, it provides information about the automated scoring systems for the speaking and writing tasks and how they were evaluated.

© 2020 Duolingo, Inc
Figure 6. Example Speaking Items

© 2020 Duolingo, Inc
Figure 7. Example Writing Items

1:32

Write one or more sentences that describe the image

Your response

1:32

Respond to the questions in at least 50 words

“I do not live to eat, but eat to live.”

Consider the subtleness of the sea: how its most dreaded creatures glide under water, unapparent for the most part, and treacherously hidden beneath the loveliest tints of azure.

Words: 98