

## **Evaluation of the Duolingo English Test: Implications for K-12 English Language Learners (ELL)**

### **Cara A North**

Center on Education and Training for Employment (CETE)  
The Ohio State University  
1900 Kenny Road  
Columbus, OH 43120  
[north.129@osu.edu](mailto:north.129@osu.edu)

### **Anna R Leach**

Master of Learning Technology  
The Ohio State University  
11007 E. Midnight Moon Lane  
Tucson, AZ 85747  
[leach.155@osu.edu](mailto:leach.155@osu.edu)

### **Natalie R Gintert**

Master of Learning Technology  
The Ohio State University  
Ramseyer Hall  
29 W. Woodruff Ave.  
Columbus, OH 43210  
[gintert.5@osu.edu](mailto:gintert.5@osu.edu)

### **Tim Nunn**

Master of Learning Technology  
The Ohio State University  
Ramseyer Hall  
29 W. Woodruff Ave.  
Columbus, OH 43210  
[nunn.15@osu.edu](mailto:nunn.15@osu.edu)

### **Ana-Paula Correia, PhD**

The Ohio State University  
Educational Studies Department  
310M Ramseyer Hall  
29 W. Woodruff Ave.  
Columbus, OH 43210  
[correia.12@osu.edu](mailto:correia.12@osu.edu)

## **Abstract**

The Duolingo English Test (DET) is an English proficiency exam that “generates a CEFR-aligned score” and compares to “other major assessments such as the TOEFL and the IELTS.” (Duolingo English Test, 2016). The test is designed to provide an English proficiency score and “...make standardized tests more accessible to Duolingo users, particularly those outside the U.S., who want to apply for jobs and schools that require English certificates.” (Rodriguez, 2014, p. 1). The test provides “a video that showcases the applicant’s practical communication skills” and “a rigorous, quantified score that signals the applicant’s English level.” (Duolingo English Test, 2016). Paired with the Duolingo English app, data the DET provides can help teachers and administrators benchmark English proficiency across their English as a Second Language population. For this evaluation, the evaluation team negotiated with Duolingo for temporary, free access to the test. Based on the evaluation of the product, we created a learning technology evaluation framework CPR: Context, Pilot, and Report.

## **Introduction**

Duolingo is a technology company based out of Pittsburgh, PA that offers customers different products around language learning. Since its creation in 2012, the company has expanded to include three main products, Duolingo, Duolingo English Test (DET), and TinyCards. Duolingo, original product created by von Ahn and Hacker in 2012, is a free language learning platform offering customers a way to learn a variety of languages on the web or through a mobile application. DET, created in 2014, is an online proficiency testing that allows customers to showcase his/her level of knowledge with the English language. The DET English test, unlike the language learning tool, is not free to use. The funding model for this product is \$49 a person, marketed heavily towards universities and businesses using it to accept non-native English speaking students or employees.

## **About the DET**

A common issue for schools that work with English Language Learners (ELL) students is categorizing the learner’s English language proficiency. Currently, this is often evaluated by the Test of English as a Foreign Language (TOEFL) scores. However, the TOEFL is expensive. Prices vary by country, but usually fall between \$160 and \$250 USD. A primary outcome of the DET is to certify the test taker in the English language. A potential reason Duolingo decided to go in this direction is due to the cost of the TOEFL test as well as the time it takes to receive the results. Duolingo’s version allows users to take the test in the comfort of their own home via the web, cost significantly less, and allows test takers to post results to social media and institution applications.

One concern would be whether DET would be an acceptable comparison the TOEFL scores. Ye (2014), found that amongst a sample of non-native English speakers, “scores from the Duolingo English Test were found to be substantially correlated with the TOEFL iBT total scores, and moderately correlated with the individual TOEFL iBT section scores, which present strong criterion-related evidence for validity” (p. 11). Despite it being proctored online, there are multiple safeguards in place such as having a live proctor watch a recording of the assessment. The test is said to provide universities with “a video that showcases the applicant’s practical communication skills” and “a rigorous, quantified score that signals the applicant’s English level.” (Duolingo English Test, 2016). These benefits make the DET a great option for English language testing. DET administration opens a world of possibilities for future applications.

## **Evaluation Framework**

To analyze and evaluate the DET, we created an evaluation framework called CPR. CPR stands for Context, Pilot and Report. Often in learning technology, proper evaluation is costly and time consuming. By using CPR, administrators and stakeholders can focus on elements of an evaluation that can lead them to data informed decision making.

## **Context**

For the first step, Context, we took a hard look at the fundamentals of DET. Using our log ins, each member of the team stepped through logging in and taking a practice test. We each took notes about our own experiences and meet to discuss the pros and cons. We then created several evaluative questions around our initial needs for the DET. From here, using our initial thoughts and guiding questions, we created a practical approach to

reviewing the product. This allowed us to critically explore the technological requirements, cost, and feasibility of mobile access.

## **Pilot**

For the second step, Pilot, we identified a population to test out DET. Using the evaluative questions as guides and our prior experiences, we surveyed the population after their review of the product to gather additional evidence for the evaluation of DET. The pilot step is often the one that is over looked, but is critical to gather perspective from the users.

## **Report**

Lastly, for the Report step, we gathered the data and began to create a story. We had the reporting step in mind during the context and pilot phases. We asked ourselves, what information do we need to answer the evaluative questions? The results of the population survey as well as the DET scores and other valuable information gathered from population were critical in making inferences about the product viability.

## **Recommendations**

From the CPR process, recommendations for product improvement were shaped. While the DET is a well-rounded product, there are several recommendations that should be considered to make the product better. When evaluating the DET, our group approached it from multiple perspectives including the user experience, educational value, and future implications. Based on these lenses, we have identified four recommendations for the DET: enhance the certificate the user is awarded after the exam, create Section 508 of the American with Disabilities Act test items, provide instructions to the Duolingo English Exam in a user's native language, and provide percentage breakdowns on the report of the test.

### *Certificate Improvements*

Once a user completes DET and their exam has been reviewed by the Duolingo reviewer, the user will receive an email notifying them of their certificate. To access the certificate, a link is sent to the user. On this link, the certificate includes the following information: a photo of the user from the exam, the name of the user, the date the exam was taken, and the level assigned to the user (beginner, intermediate, expert, etc.). Along with the level, Duolingo provides context about each of the levels. One improvement that should be considered is that it seems the level information pre-populates in a section with a character limit. In the example of the certificate above, the description of being an expert in English cuts off. The user and anyone with the link to the certificate is left without a full description of the level. Additionally, going more in depth on what a level means and sharing how it relates to the Duolingo app data would be helpful. Also, an improvement that could be considered is listing all of the levels a user can earn. One way of doing this is to create an artifact to accompany the user certificate. This way users, as well as organizations that the user may share the certificate with, will have more information about the levels. Furthermore, the artifact can provide organizations information about the DET as a whole. Another recommendation for the user certificate is providing a secure URL. The way the DET certificate is shared currently is through a URL. This URL has no safeguards so anyone with the URL can access a user's certificate. Depending on how the user feels about this, it would be an improvement if the URL had a captive portal or some type of safeguard to protect the information.

### *Create Accessible Test Items*

While the DET does a great job with offering a variety of types of questions, it could be improved by taking in Section 508 of the American with Disabilities Act into consideration to create test items that are accessible. Throughout the DET, there are various types of questions such as choosing the best word from a drop down menu, speaking about a topic, and selecting words from a list that are English words. While this variety makes the DET unique, it could be improved by considering different formats that are accessible. For the items that ask users to choose the best word from a drop down menu, Duolingo should consider enabling keyboard shortcuts and arrow key usage to make this type of question more accessible. Many users who use screen readers rely heavily on keyboard shortcuts in the same way that other users utilize a mouse. Furthermore, the speaking of topic items would be

difficult for users who are deaf. Offering an alternative question type such as an essay would be ideal. One way Duolingo could help users who may have disabilities is to ask in the beginning while setting up the test if they require an accommodation. Duolingo staff could then reach out to the user before the test or perhaps Duolingo would provide contact information for the user to contact Duolingo directly in order to assist the user.

#### *Provide Instructions to the DET in User's Native Language*

During our evaluation of DET, a group member served as a pre-test proctor, sitting in the room to assist the user to get the test set up. While a primary purpose of this was to make sure the users were set up properly, another reason for the pre-test proctor is because the DET has all instructions in English. Depending on the English level of the user, this may be a hindrance. Due to the specific instructions such as having no one in the room as well as taking photos of government issued identification, it is imperative the user understand the process in order to complete the DET without fear of it not passing audit review. If DET offered instructions in multiple languages, it could guarantee that users understand the protocol and policies of the test. As discussed above, we believed the test would allow users to set up the test in their native language. But, it did not work during the learners' try-out testing when students tried to select a language other than English.

#### *Provide Percentage Breakdowns on the Report of the Test*

Since the DET has many different types of items, it would be helpful for a user to receive a breakdown of the percentage correct in each different type of question. One item type asks the user to talk about a specific incident, such as a time they had to wait for something. Based on the information in our evaluation, we believe that the speaking part of the exam is not factored into the overall score in the DET. Having a breakdown of the item type would help both the test taker as well as an organization receiving the score to make an accurate judgement about someone's English abilities.

### **Conclusions**

The Duolingo English Test, when paired with the Duolingo English app, could provide ELL instructors multiple types of data for their students. Students who take the DET will be provided a digital artifact and encounter multiple question types to test their English proficiency. Before considering a full adoption, an evaluation is recommended using the framework Context, Pilot, and Reporting (CPR).

### **References**

- Duolingo (2016). *About Us*. Retrieved from <https://www.duolingo.com/info>
- Duolingo (2016). *Duolingo English Test Scores*. Retrieved from <https://www.duolingo.com/scores>
- Duolingo (2016). *Frequently Asked Questions*. Retrieved from <https://testcenter.zendesk.com/hc/en-us/categories/201147826-Frequently-Asked-Questions>
- Educational Testing Service (2016). *TOEFL iBT® Test Fees*. Retrieved from <https://www.ets.org/toefl/ibt/about/fees/>
- Rodriguez, S. (2014). Duolingo offers language-certification tests via mobile devices. *Los Angeles Times*. Retrieved from <http://www.latimes.com/business/technology/la-fi-tn-duolingo-test-center-20140722-story.html>
- Ye, F. (2014). *Validity, Reliability, and Concordance of the Duolingo English Test*. Retrieved from <https://s3.amazonaws.com/duolingo-papers/other/ye.testcenter14.pdf>