
Short Research Papers

Does Duolingo Improve TOEIC Scores?

Jeff Broderick

Tokyo Denki University

This study investigates the effectiveness of Duolingo, a popular smartphone language learning app, for improving Japanese university students' TOEIC Bridge Test (BT) scores. Although past studies have claimed that approximately 30 hours of Duolingo use produces similar learning to a semester of college for beginners, this study found that Duolingo had little effect on learners' TOEIC BT scores.

The advent of new technology results in educators wondering whether it can help their students achieve their learning goals. One such technology is language learning applications (apps) for smartphones. Duolingo is one widely used language app, with over 300 million users worldwide learning 22 languages, according to the Duolingo website (Duolingo Research., n.d.).

Language apps can offer certain advantages over traditional study methods. Many of these features match what Sharples (2000) described as the desirable characteristics for mobile learning technology. Learners carry their devices at all times, meaning they can choose when and where they learn, leading to greater autonomy (Farmer & Nucamendi, 2012). Increasing learner autonomy is one of Dörnyei and Csizér's (1998) "Ten Commandments" of language teaching and leads to greater learner well-being (Oxford, 2016). Level testing means that lessons are offered at a difficulty tailored to the individual. They may utilize Spaced Repetition Systems (SRS) which help optimize retention and learning (Kang, 2016). The learning experience is gamified: users accumulate points, leading to immediate feelings of accomplishment. Some systems are integrated with social media, allowing users to share their progress or compete with friends.

Crucially, many are usable for free.

It is no surprise then, that language apps are extremely popular. But how effective are they? An oft-cited 2012 study by Vesselinov and Grego, commissioned by Duolingo, found positive results: adult beginner Spanish learners acquired a semester's worth of knowledge in an average of 34 hours of use. Student scores on a standardized WebCAPE Spanish test increased by an average of 8.1 points per hour of Duolingo use. Munday (2016) suggests that Duolingo should be a component of university language classes, while Krashen (2014) is more skeptical.

Daily practice and review may be the best way to learn a language. Many Japanese university students, however, receive English instruction for just a few hours a week. Given this reality, daily review using a language learning app could be a valuable supplement to classroom instruction, particularly if Vesselinov and Grego's (2012) results are typical. It is therefore important to test their findings in a university setting. Consequently, the current study was formulated around the following research question: Does Duolingo use lead to score increases for university students on a standardized English test, the TOEIC Bridge Test (BT)?

Methods

Participants for this study were students at a science and engineering university, enrolled in the author's compulsory first- and second-year oral communication English classes. Participants also took a reading and writing class (taught by other instructors). Contact hours were 90 minutes per week for each class, over 14 weeks.

The study began with 224 participants. They took the TOEIC BT in week 1 of the spring semester and again in week 14. The TOEIC BT is a one-hour multiple-choice listening and writing test, yielding a score from 0 to 100. In the first class, participants were asked to study English on their own using Duolingo, with Duolingo counting for 10% of their final English score. They took an initial in-app level assessment test and were directed to select the "casual" study option, which presents users with lessons of approximately 10 minutes' duration. Completing 60 lessons (i.e., 10 hours of study) over the semester would equate

to a full score of 10 points, so students were advised to aim for about one lesson per weekday over the semester.

Due to absences, not all students took both tests. Following Vesselinov and Grego (2012), students who used Duolingo for less than two hours were removed from the study. Any students studying TOEIC skills, either in other classes or independently, were eliminated. Students with more than 40 hours of use were removed as outliers, as were those with extreme test scores or changes in score (e.g., a student who slept through the first test; a student whose score dropped by 40 points). After data screening, 116 participants remained ($N = 116$; Year 1 $M = 38$, $F = 12$; Year 2 $M = 44$, $F = 12$).

Results and Discussion

A linear regression was performed to understand the effect of hours of Duolingo use on TOEIC BT score increase. To assess linearity a scatterplot of score increase against Duolingo use was plotted. Visual inspection indicated a weak linear relationship between variables. The residual plot indicated homoscedasticity of residuals, and the normal probability plot indicated an approximately normal distribution, with a slight tendency towards a negative kurtosis. Regressions with $n > 15$ are considered robust against non-normality, and as this study utilized 116 data points, analysis proceeded. The results of the regression are shown in Figure 1, below. Partial statistics for this test are presented in Table 1.

The prediction equation was as follows: TOEIC BT score increase = $0.998 + 0.121 \cdot \text{hours}$. Time spent using Duolingo significantly predicted score increase, $F(1, 115) = 4.937, p = .028$, accounting for 4.2% of the variation in score increase with adjusted $R^2 = 3.3\%$, a small size effect according to Cohen (1988). An hour of Duolingo use led to a 0.121 (95% CI, 0.013 to 0.230) increase in TOEIC BT score.

While most students' scores increased, a significant portion went down. Furthermore, the R^2 value for the regression is between 3~4%, indicating that only a small portion of the variance is attributable to Duolingo. According to the regression model, one hour of studying with Duolingo yielded an increase of 0.1 points on the TOEIC BT, which must be regarded as a very small increase.

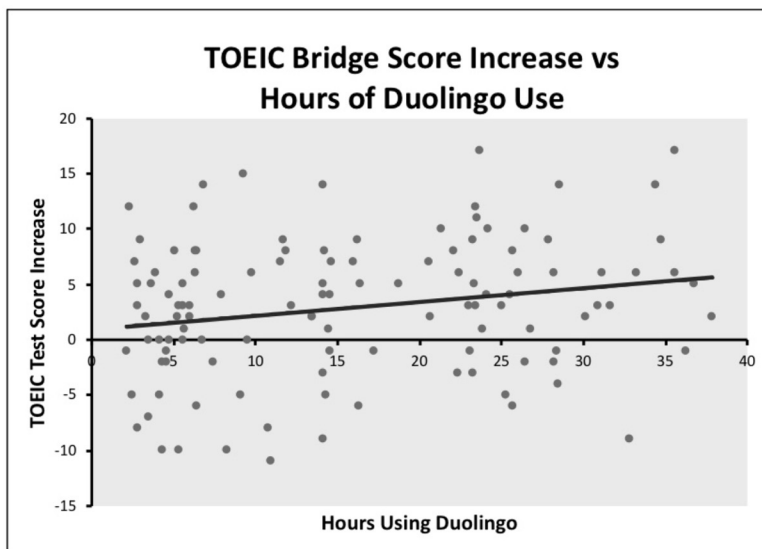


Figure 1. Regression results.

Table 1

Regression Model ($N = 116$)

	Lower 95%	Predicted	Upper 95%	P -Value
Y-Int.	-1.059	0.998	3.055	0.339
Coeff.	0.013	0.121	0.230	0.028
R^2	0.042		F	4.937
Adj R^2	0.033		Sig F	0.028

These results appear to indicate that while Duolingo has a positive effect, it is not an efficient way to learn the skills and knowledge applicable for the TOEIC BT. This may be because the level of grammar and vocabulary trained with Duolingo is too low to be useful on the TOEIC, which is oriented towards business and professional English.

A further possibility, difficult to test for in practical terms, is that most of the gains provided by Duolingo were negated by other factors. With no extrinsic reward for score increases, it may be that they simply did not try very hard on the

second test.

Participants in Vesselinov and Grego's 2012 study were adult learners with a desire to learn Spanish. Their motivation is almost certainly higher than Japanese students enrolled in a compulsory English course, and both this author's past experiences and past studies (e.g., Broderick, 2013; 2015) indicate that these students are not highly motivated.

Furthermore, correlation between TOEIC BT score gains and Duolingo use does not necessarily imply that the latter caused the former. In the current study, the observed gains may be related to a mediating factor such as motivation, as in the following way: students who are more motivated to study English used Duolingo more, and *at the same time* studied harder in their other English courses, and consequently may have benefited more from their efforts.

Conclusion

Academia, and the social sciences in particular, may be said to be in the midst of a replication crisis. It is therefore important to attempt to replicate studies, especially when those studies make extraordinary claims. The claims of Vesselinov and Grego (2012) regarding Duolingo's effectiveness are certainly striking. While these results may be applicable to motivated adult students beginning their study of a language they want to learn, the current study indicates that Duolingo is far less effective for Japanese university learners: an average of between eight and nine hours of Duolingo use would be required to gain just one point of improvement on the TOEIC BT.

The TOEIC BT might not accurately test the kind of knowledge imparted by Duolingo. However, to the extent that the TOEIC BT assesses grammar, listening, and vocabulary, Duolingo does not appear to effectively improve these skills. This may be a problem of motivation. Aspects of Duolingo that make it easy to use (gamification, multiple-choice answers, daily reminders, etc.) may also make it possible to use Duolingo with only half of one's attention. Students who use the app primarily to satisfy the requirements of a compulsory course, and not due to any intrinsic desire to learn, may use the app in this way, failing to internalize (i.e., acquire) the grammar and vocabulary they encounter in the

platform.

Past studies (Broderick, 2013; 2015) that analyzed the language learning motivation of students at this university found a lack of intrinsic motivation. English may be viewed not as a communication tool, but rather as a dry academic subject with little connection to their lives. For such students, Duolingo may be ineffective, even if it is easy to use and convenient. They may have little retention of the material.

The finding that Duolingo can effectively teach a language should be investigated in a wider range of learning environments, and also with a variety of implements such as WebCAPE to compare learner progress equally. Another avenue of investigation would be to test Duolingo against other smartphone applications and directly against traditional learning methods.

New technological developments tend to be adopted quickly for classroom use, and smartphone apps have been no exception. While it is certain that these applications are here to stay, as researcher-educators we need to assess their usefulness and applicability to our learners before relying on them in our classrooms.

References

- Broderick, J. (2013). Profiling the motivation of TDU English learners. *Bulletin of Tokyo Denki University, Arts and Sciences*, 11, 145–155.
- Broderick, J. (2015). Dynamics of English motivation and willingness to communicate. *Bulletin of Tokyo Denki University, Arts and Sciences*, 13, 195–202.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Psychology Press.
- Dörnyei, Z., & Csizér, K. (1998). Ten commandments for motivating learners: Results of an empirical study. *Language Teaching Research*, 2(3), 203–229.
- Duolingo Research. (n.d.). <https://research.duolingo.com/>
- Farmer, F., & Nucamendi, M. (2012). Towards a rationale for mobile learning. In J. E. Diaz-Vera (Ed.), *Left to my own devices: Learner autonomy and mobile-assisted language learning* (pp. 21–44). Emerald Group.

- Kang, S. (2016). Spaced repetition promotes efficient and effective learning: Policy implications for instruction. *Policy Insights From the Behavioral and Brain Sciences*, 3(1), 12–19.
- Krashen, S. (2014). Does Duolingo “trump” university-level language learning? *The International Journal of Foreign Language Teaching*, 9(1), 13–15.
- Munday, P. (2016). The case for using Duolingo as part of the language classroom experience. *Revista Iberoamericana de Educacion a Distancia*, 19(1), 83–101.
- Oxford, R. (2016). Toward a psychology of well-being for language learners: The ‘EMPATHICS’ vision. In P. D. McIntyre, T. Gregersen, & S. Mercer (Eds.), *Positive psychology in SLA* (pp. 10–87). Multilingual Matters.
- Sharples, M. (2000). The design of personal mobile technologies for lifelong learning. *Computers and Education*, 34, 177–193.
- Vesselinov, R., & Grego, J. (2012). Duolingo Effectiveness Study Final Report. Manuscript available through Duolingo.

Author bio

Jeff Broderick has taught English at Japanese universities since 2004 and at Tokyo Denki University since 2013. His research interests focus on motivation, willingness to communicate, and recently, technology. He is also interested in translation, history, and martial arts studies. <broderickjeff@mail.dendai.ac.jp>

Received: October 20, 2019

Accepted: September 9, 2020