Usability Evaluation of Language-Learning Software

Report

Christopher Jones
Bournemouth University
Human Factors in Computer Systems
Word Count: **2931** *(Including References)*

0 Abstract

Difference in languages remains one of the largest cultural barriers that the population encounters around the world. From work classes, to studying abroad, to using premium computer applications, there are a variety of methods that can facilitate people with learning foreign languages. With an abundance of methods to choose from, it can be challenging to decide which learning approach should be performed.

This report investigates the usability of the Duolingo system. By using a rich picture and multiple human factor approaches, I will determine how successful mobile applications are at teaching foreign languages.

Results were generally positive, with a SUS score of 72.5 and users praising it for its ease of use and gamifying elements. There were several areas of concern, that being accessibility and quality of life features that the platform lacks. These lack of features caused many members with physical disabilities to not have a straightforward experience.

With a vast majority of participants enjoying the product, it became more understandable as to why Duolingo is as popular as it is right now. However, the mobile-based platform is not a perfect method of delivering educational content, with areas of the app deemed not as effective compared to using alternative language-learning strategies.

Stakeholders will need to rectify those issues, whilst continuing to iterate on the product. By doing so, they will make sure that their customers are satisfied now and for the foreseeable future.

1 Introduction

The challenge that stakeholders confront is to design a dependable product for users of all skill levels to use regularly, showing no user experience concerns. The application should have as fewer perceivable barriers as possible, whilst being accessible to all consumers.

With a product like Duolingo, the fear of a stagnated or diminishing active user base (*Lardinois, 2018)* is a major problem that stakeholders encounter during their development lifecycle. Therefore, all the human factor approaches will be centred on evaluating the incentives with using the system on a regular basis.

AIM: Produce an in-depth evaluation on a language-learning system (Duolingo). Determining whether using mobile-based applications is the optimum method to learning a foreign language.

OBJECTIVES:

- *Analyse the UX techniques and conduct thorough accessibility tests*
- *Measure the usability of the Duolingo system.*
- *Identify potential usability issues within the system.*
- *Employ a mixture of investigative and generative methods/tools to develop an understanding of user needs within the system.*
- *Identify useful advancements that can be made on the overall application*

2 Duolingo

Learning a new foreign language can be a very challenging exercise, with only 34.6% of UK citizens capable of speaking multiple languages (*Eurostats, 2018).* Thus showing that the methods put into place to teach individuals can improve, and Duolingo is one of many attempting to do so.

Duolingo is a freemium language-learning platform, well renowned for its gamifying elements included in each course. They have taken major steps into producing a fun and universally accessible learning experience.

Stakeholders of Duolingo have claimed that their ambition with the platform is to have the same capabilities as a one-to-one trainer, but admits that there is still a long way to go to achieve that goal (*Communicaid Limited, 2018*). Analysing Duolingo as a whole can determine those potential weaknesses in the system and what can be improved.

3 Methods

3.1 Rich Picture

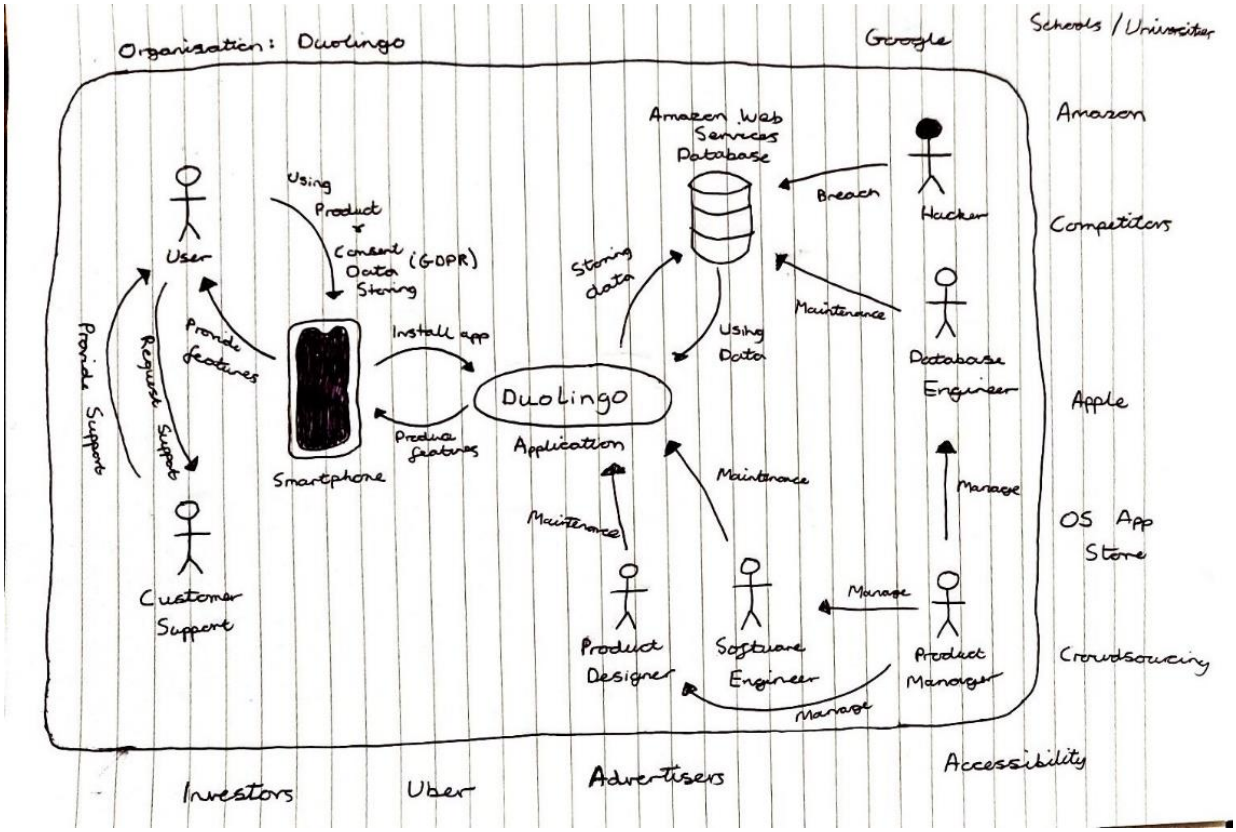


Figure 2: Usability evaluation of the Duolingo app

Figure 2 contains the rich picture, a detailed visual representation of the defined system. It is used to explore and exemplify the problem situation through a simple diagram. Thus will help open a greater discussion, whilst giving a shared understanding of the problem in hand.

3.2 System Usability Scale

The System Usability Scale (SUS) is a simple, ten-item technique for globally assessing the usability of software. This reliable tool provides a universal outlook on Duolingo's app usability (*Brooke, 1996*).

Adopted largely by various international industries to test numerous systems, it has become an industry standard due to its extremely brisk and straightforward practice.

However, shortcomings with the system are recognised, including the inability to provide an accurate judgement on a product's weaknesses. Therefore, this method is adopted instead to complement additional usability measuring tools.

A total of ten participants, at an age range of 20-30 took part in the ten-item questionnaire inspired by John Brooke (*Brooke, 1996, p.4*). Participants are asked to score the items with one of five responses that range from *Strongly Disagree* to *Strongly Agree*. This scales all values from 0 to 4 (with four being the most positive response).

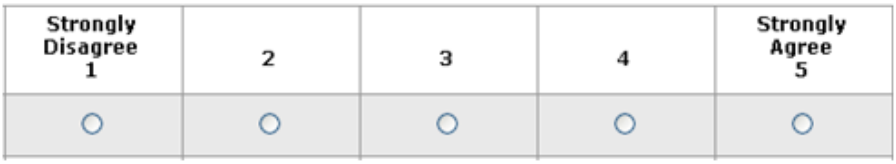


Figure 3: SUS Response Format

For each of the ten questions, the scale is converted into single numbers, representing the usability score. Additionally, the results yield a single numbered score that serve as the overall usability of the system. The final SUS scores are compared in relation to the average SUS scores produced as shown in Figure 4.

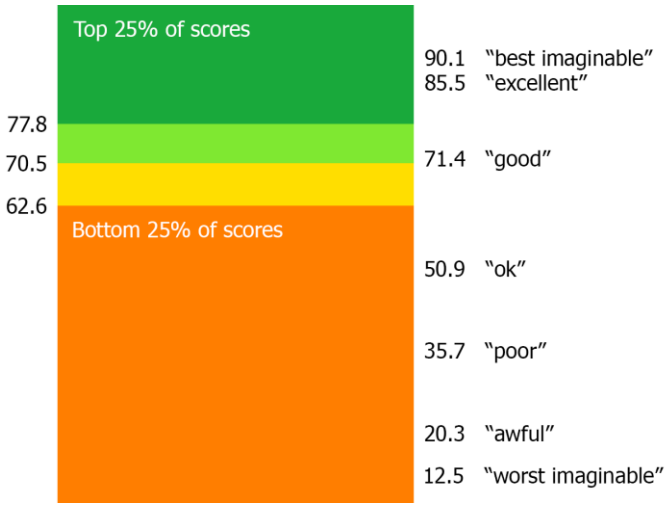


Figure 4: SUS Percentile Ranking

3.3 User-Centred Design

One key element of Duolingo in their business is the value of each customer, to analyse the customer profitability and maximising their user lifetime cost.

Increasing or maintaining the user base is a necessary task to meet stakeholder expectations with the product. Therefore using the User-Centred Design (UCD) process will analyse the needs, wishes and limitations that users face. UCD is a common design industry practice and when adopted leads to an increase of product usefulness and usability (*Vredenburg et al, 2002*).

Such testing is necessary as it is often strenuous for designers to understand intuitively what first-time users of their design experiences, and what each user's learning curve may look like. By testing the product with end-users, this will help bring customers at the centre of the experience whilst saving resources.

UCD is by no means a flawless model, containing certain limitations like the process being relatively time-consuming. So striking a balance between user and designer opinions is essential as basing designs solely on what consumers want can be very problematic long-term.

I interviewed ten users (aged between 20 and 30) of diverse proficiencies a series of questions regarding their tasks and goals. Their results are accumulated into an eight field table similar to *Figure 5*, coining a simple unique value proposition that summarise the entire process. The eight fields are the following:

- *Business*
- *Users*
- *Problems*
- *Motives*
- *Fears*
- *Solutions*
- *Alternatives*
- *Competitive Advantages*
- *Unique Value Proposition*

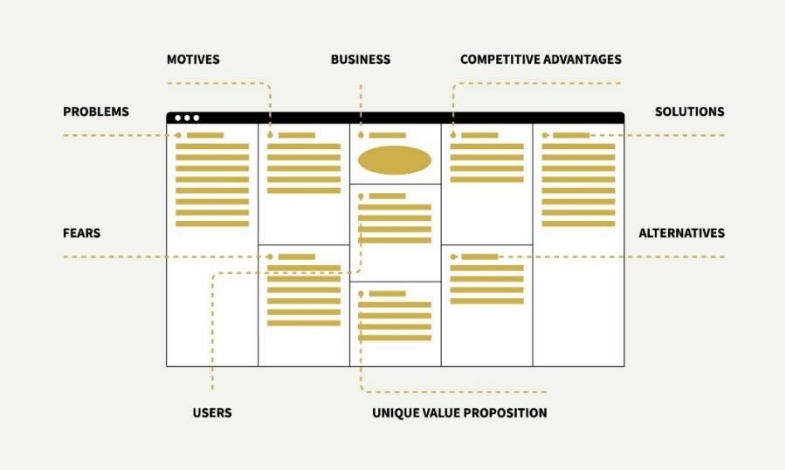


Figure 5: UCD Structure

3.4 Nielson's Heuristic Evaluation

Unlike the other usability testing methods that count on users to test the product, heuristic evaluation instead relies on experienced evaluators. This method requires an expert of the system to review the product on predefined characteristics, identifying the common usability issues found in the design.

Largely used due to how inexpensive the technique is, Heuristic Evaluation can be used to obtain useful feedback in the early stages of development. This model will be used to complement other usability testing methodologies, with no requirements for advanced planning necessary.

However, the evaluator must have some form of experience to use this method effectively, which can potentially be costly long-term. Therefore, deadlines have been put into place to make sure that the evaluation doesn't take longer as people expect.

The evaluator will follow Jakob Nielsen's 10 general principles for usability found in his guide (*Nielsen, 1995*).

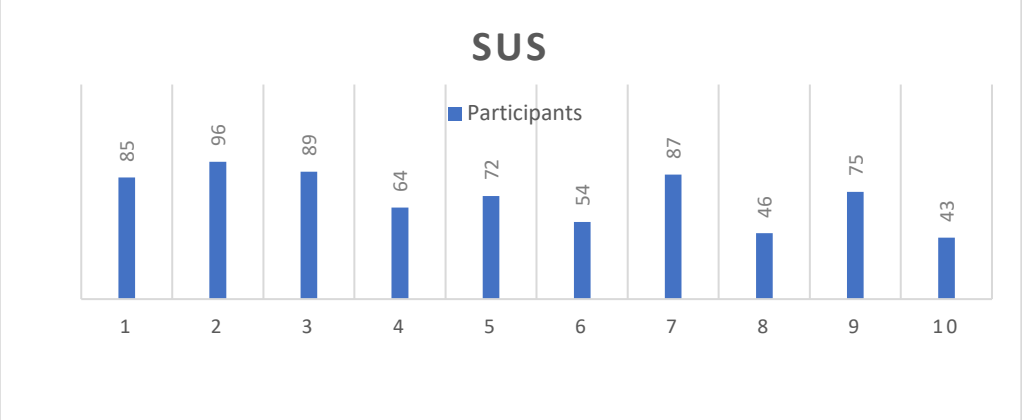
The evaluator will use their judgement and expertise to conduct the Heuristic Walkthrough, recording any issues that they may encounter and explaining what heuristics have been violated and why. Information will be collected and recorded in a radar chart, marked with severity ratings to enable rapid deployment of resources and fix the problems identified.

4 Results

4.1 System Usability Scale

Interpreting the usability score requires the participant's scores for each question to be converted into a new integer. The scores are added together and then multiplied by 2.5 to convert the original scores of 0-40 to 0-100. The scores should only be considered in terms of their percentile rankings.

(a) System Usability Scale (SUS)



(b) Box-and-Whisker plot of the user scores (10 SUS statements)

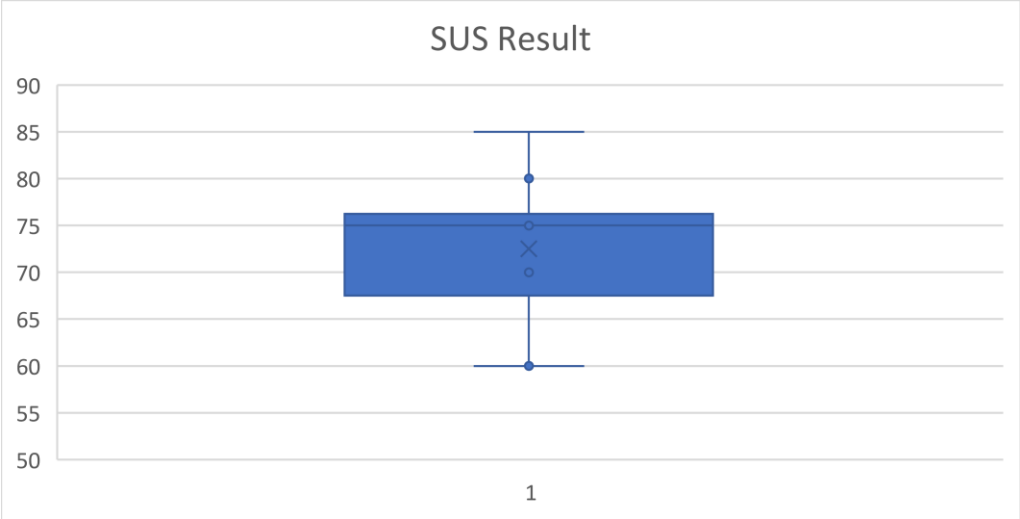


Figure 5: System Usability Scale (SUS) results per question

4.2 User-Centred Design

<table>
<tr>
<td rowspan="2">3. Problems

- User is planning on going on holiday, so will need to learn the basics of a foreign language quickly.

- User wants to challenge themselves and learn a new language from scratch.

- User that has spoken a particular language at an advance level wants to challenge themselves.

- Users with physical disability to learn a language with no barriers.

- Users to continue learning a language offline.

- Using mobile resources can become a major distraction

- Mobile rarely providing any real speaking/listening experience.

- Becomes more about the gamifying scores and less about learning the language.</td>
<td>4. Motives

- Low cost

- Ease of use and access

- Wants to learn a new foreign language

- Advance users to challenge themselves</td>
<td>1. Business

Duolingo

Freemium language-learning platform, featuring over 80 different language courses.</td>
<td>8. Competitive Advantages

- Free to start
- Gamifying aspects
- Learn a language anytime and anywhere.
- Covers a wide range of contexts.
- Friend features</td>
<td rowspan="2">6. Solutions

- Provide quick and simple courses that teach users the basics of a foreign language.

- Give users in-depth courses that challenges both new and experienced users with a particular language.

- Provide accessibility features for the physical disabled, which includes a colour-blind mode.

- Give users the ability to save specific courses so they can work offline.

- Provide gamifying features to provide user incentives.

- Duolingo Live feature that gives users 1-1 lessons on the app

- Use Duolingo in tandem to other applications/methods.</td>
</tr>
<tr>
<td>5. Fears

- Language courses are ineffective.

- App causes users to start spending money to get the full experience.

- Lack of context in each course.

- Lack of incentives while using produce

- Very little grammatical instructions

- Can get distracted easily

- Lack of repetitive drills</td>
<td>2. Users

Everyone
Language-learning business that gives everyone an opportunity to learn languages, particularly those that couldn’t afford to do so.

9. Unique Value Proposition

Duolingo – A free language learning platform, where users can develop their language skills wherever they go, whilst competing with friends and family.</td>
<td>7. Alternatives

- Alternative language-learning applications (Babel, Rosetta Stone)

- Personal language schools/courses.

- Direct travelling method

- Audiobooks

- Language-learning books

- More comprehensive programs

- Language exchanges</td>
</tr>
</table>

Figure 6: UCD Results Table

4.3 Nielson's Heuristic Evaluation

Positive Findings

FINDING NO.	HEURISTIC COMPLIED	TASK/ISSUE/DESCRIPTION
1	1	Informs users when a course is completed/points have been given to the user
2	3	Members have a "home" link in the bottom-left hand corner to return back to the home page.
3		User signs them in automatically when opening the app and have already signed in before
4	4	If a user is not signed into the site, they are taken to the registration page
5	5	Error messages are written in plain language and easy to understand
6	7	Inactive menu items greyed out

Negative Findings

FINDING NO.	HEURISTIC VIOLATED	TASK/ISSUE/DESCRIPTION
7	1	Not every display begins with a title or header that describes the contents.
8		No feedback when keys are pressed
9		Not all the icons are concrete and familiar
10	4	Not all icons are labelled
11	5	Sound not used to signal an error
12		Error messages don't suggest the cause of the problem
13		Error messages don't indicate what action the user needs to take to correct the error
14	6	Difficult to find how to sign out of user account. – Forcing users to go through the settings to sign out.
15	7	Colour-coding not found throughout the system

5 Discussion

5.1 System Usability Scale

Figure 5(a) shows the SUS score of each test user; most scoring better than the benchmark. Overall the Duolingo application has obtained an average SUS score of 72.5, well above the benchmark (68) showing a score falling between "good" and "excellent".

Figure 5(b) shows a Box-and Whisker diagram of the 10 SUS statement scores. The diagram displays the median and distribution of the scores. The mean is 72.5, which is an impressive result.

In addition, I can see that the highest average score is 85 for question 1, showing that the users have evaluated the Duolingo app to not be a complex system. Due to its simplicity, it makes is a valuable platform for newcomers to begin participating in their online courses.

5.2 User-Centred Design

Figure 6 describes complications that are encountered while using Duolingo in the perspective of users. UCD revealed many queries and resolutions that were never considered in stakeholder's point of view.

Users discovered 5 main issues that they encountered while using the product on a daily basis, most were centred on accessibility and quality of life. Developers should commit more resources into producing features that will directly improve the experience of the app, these features include implementing text-to-speech and a colour-blind mode for individuals.

The target audience identified for this application is everyone because the platform is centred on profiting through ad revenue, creating a large user base due to its simplicity and affordability. Having such an extensive target audience signifies that the product needs to contain components that will satisfy everybody, which is a very unrealistic expectation.

Alternative methods that users adopted include language exchanges and external classes. UCD results discovered that the leverage Duolingo has over its competitors are in regards to price of entry and the portability aspect that the smartphone provides. It however lacks the real speaking/listening experience that the alternatives provide.

There were many anxieties that users felt while using the product. Examples being the smartphone's easily distracted nature and the application's lack of ability to teach the users, which is not as effective as users anticipated. This shows the disparity that users want the application to achieve compared to what the application can ultimately produce.

Stakeholders should work on producing more features that manages users of all proficiencies when utilising the product, whilst making courses repetitive and beneficial.

5.3 Nielson's Heuristic Evaluation

Heuristic evaluation revealed several findings hindering user progress while performing tasks on the Duolingo app. In view of new customers, it can prove to be difficult to perform simple tasks like signing out of the application. Most of the issues can be resolved by implementing major changes to the user interface of the system.

Furthermore, refactoring the error messages that appear to notify the user the actions that can be made to solve the problem would be a huge benefit. Lack of colour-coding on the site proves to be a hindrance for colour-blind individuals.

Providing users with feedback when a key is pressed (vibration on the smartphone) will contribute a lot to partially blind users. Currently there is no indication that a user input has been made apart from the page changing.

In order for Duolingo to become one of the best language-learning methods for all users, these negative findings need to be resolved. Thus, providing a product that has as fewer perceivable barriers as possible.

6 Conclusions

This study on the effectiveness of Duolingo answered many important questions. Duolingo is a very good product and using mobile-based applications has shown to be a competent method of learning.

However, the product isn't the optimum method. The disparity between what the system can provide the user, and what the user demands the system to accomplish has proven to be a big issue. Improvements are needed in many areas of the app if the company wants to grow into a greater, more ambitious product.

The key aspects of reform are the user interface and lack of accessibility features, as the obstacles put into place while performing simple tasks can disapprove new customers. The stakeholders must address these areas as soon as possible if they want to satisfy a large number of current and future users.

Also, it needs to be noted that the analysis of this product is limited due to time and money constraints; a greater depth of understanding and evaluation can occur with the utilisation of their competitors. Only then can a full appreciation of the business's current advantages and disadvantages can occur.

As long as user expectations are kept to a realistic level, mobile-based applications are a great secondary resource for newcomers to start learning foreign languages. Alternative methods that have been around longer than Duolingo still provide features that apps lack, like the 1-1 interaction with foreign individuals.

Duolingo has strong future prospects in the area of language-learning if it continues down its current path though. Stakeholders should continue to iterate and produce more content for their existing users, while maintaining the low price of entry for new ones.

7 References

Lardinois, F., 2018. Duolingo hires its first chief marketing officer as active user numbers stagnate but revenue grows [online]. TechCrunch. Available from: https://techcrunch.com/2018/08/01/duolingo-hires-its-first-chief-marketing-officer-as-active-user-numbers-stagnate/ [Accessed 6 Dec 2018].

Eurostats, 2018. 65% know at least one foreign language in the EU [online]. Ec.europa.eu. Available from: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/EDN-20180926-1?inheritRedirect=true [Accessed 5 Dec 2018].

Communicaid Limited, 2018. Language Training Apps - Do They Work? [online]. Communicaid.com. Available from: https://www.communicaid.com/business-language-courses/blog/do-apps-spell-the-death-of-traditional-language-training/ [Accessed 9 Dec 2018].

Brooke, J., 1996. [online]. Hell.meiert.org. Available from: https://hell.meiert.org/core/pdf/sus.pdf [Accessed 10 Dec 2018].

Vredenburg, K., Mao, J., Smith, P. and Carey, T., 2002. A Survey of User-Centered Design Practice [online]. Cse.chalmers.se. Available from: http://www.cse.chalmers.se/research/group/idc/ituniv/kurser/09/hcd/literatures/Vredenburg%202002.pdf [Accessed 12 Dec 2018].

Nielsen, J., 1995. 10 Heuristics for User Interface Design: Article by Jakob Nielsen [online]. Nielsen Norman Group. Available from: https://www.nngroup.com/articles/ten-usability-heuristics/ [Accessed 11 Dec 2018].