IMA 2 - Usability testing report

Duolingo

Kai Dell, Nichole Peters, Victor Basha, Kara Kelley

October 11, 2020

Introduction

DuoLingo is a mobile app that is intended to provide learners with the tools needed to practice and learn one of the supported languages. The app is adaptive and customizes lessons based on a user's proficiency and performance. Interactions are varied and include listening, vocabulary, reading, and even speaking practice. The learning experience begins with an assessment of current proficiency, progresses through tiered vocabulary and grammar activities, reinforcement activities, and includes gamification elements that reward progress.

The user experience study was conducted by Kai Dell, Kara Kelley, Nichole Peters, and Victor Basha. Data was gathered by evaluating how well users competed 8 pre-defined tasks in the app. After observation, interview, and questionnaire responses from 8 participants were analyzed.

What the app does well
- Polished graphical user interface
- Intuitive workflow for progression
- Simple language, lesson, and exercise selection
- Easy to set up account

What the app could do better:
- Provide persistent controls that allow access to profile and other options
- Profile editing and customizable options are less intuitive

Overall, participants were largely impressed with the simplicity of the app and features.

Methodology – Nichole

Sessions

Each participant was selected at random by each team member to allow for the most accurate representation of the population. Every team member utilized the provided script to test their participants and questionnaires to determine the test's success. The participants were timed and evaluated using timers and charts to determine error rate and length of time between tasks. They were assigned a total of eight tasks to complete and given a pretest explanation to ensure understanding. Once the timed portion of each test was completed, they were asked a series of satisfaction questions regarding their experience. Additionally, they were asked about satisfaction and usability of the

application in its entirety at the completion of the test.

Participants

The participants of this usability test included eight people from various backgrounds, but the distribution managed to balance out well. A perfect mix of male and female participants and different races were tested. Many of the participants were college graduates utilizing an equal distribution of operating systems for this test. Below is a table of the participants demographics to clearly visualize our eight participants over the 4-day testing period.

Participant Demographic Table					
Participant	Age	Gender	Race	Education	App Platform
1	27	Male	Caucasian	College Grad	Android App
2	28	Female	Caucasian	College Grad	Android App
3	31	Female	Asian	College Grad	IOS App
4	47	Male	Native American	College Grad	IOS App
5	17	Male	Bi-Racial	High School	IOS App
6	63	Female	Caucasian/Latin	Some College	Android App
7	41	Male	Caucasian	College Grad	IOS App
8	32	Female	Caucasian	College Grad	Android App

Evaluation Tasks/Scenarios

The tasks were decided together as a group before we recruited our participants, and they are included in the list provided. We chose tasks that would explore the entire app in all its facets and examine its usability with a new user base. The first five tasks were predetermined and were the bulk of the steps needed to interact efficiently with the app. We decided on an added three tasks to further explore the apps offerings. The tasks we added were to complete one exercise, create a profile, and finally turn on notifications.

- Task 1. Complete the steps necessary to create a new account. Stop after your account has been created.
- Task 2. Now, find one language you like to learn and add it to your profile / account.

- Task 3. Change the language to other language you like to learn.
- Task 4. Turn off and turn on auditory sound.
- Task 5. Go to the setting / preference and change one setting option (any).
- Task 6. Create a profile
- Task 7. Complete 1 exercise
- Task 8. Turn on notifications

Results

Task Completion Success Rate

The success rate for each of the eight tasks was 100%. All the participants successfully completed each task given to them.

Task Completion								
	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8
P1	1	1	1	1	1	1	1	1
P2	1	1	1	1	1	1	1	1
P3	1	1	1	1	1	1	1	1
P4	1	1	1	1	1	1	1	1
P5	1	1	1	1	1	1	1	1
P6	1	1	1	1	1	1	1	1
P7	1	1	1	1	1	1	1	1
P8	1	1	1	1	1	1	1	1
Success	8	8	8	8	8	8	8	8
Completion Rates	100%	100%	100%	100%	100%	100%	100%	100%

Learnability: Time to complete a task successfully

Each participant was timed (in seconds) while completing each of the assigned tasks. Tasks that were more difficult had a longer average time of completion. Task 7 asked participants to complete one exercise and was the most difficult task with an average completion rate of 248 seconds. The times for Task 7 ranged from 86 to 520 seconds. Task 1 required the participants to set up an account and had the next highest average of 168 seconds with a range of 60 to 300 seconds. Task 6 was to create a profile and took an average of 145 seconds with a range of 27 to 360 seconds. Tasks 2, 3 and Tasks 8 were the easiest tasks with averages of 27, 30, and 34 seconds, respectively. Task 2 asked learners to select a language and the range for this task was 10 to 45 seconds. For Task 3, participants were told to change the language to another and the times ranged from 3 to 120 seconds. Participants were asked to turn on notifications for Task 8 which ranged from 20 seconds to 60 seconds. For Tasks 4 and 5 participants were told to

turn the audio settings off/on and then to change one setting. The range for Task 4 was 10 to 77 seconds while the range for Task 5 was 9 to 115 seconds. The relation between Tasks 4 and 5 was reflected in an average completion time of 47 seconds for both.

Time on Task (in seconds)

	P1	P2	P3	P4	P5	P6	P7	P8	Total	Average
Task 1	180	300	60	120	120	240	189	136	1345	168
Task 2	34	10	30	45	30	45	10	10	214	27
Task 3	30	20	15	3	25	120	15	15	243	30
Task 4	60	77	45	30	35	50	66	10	373	47
Task 5	20	115	80	45	30	60	15	9	374	47
Task 6	120	240	360	120	120	130	46	27	1163	145
Task 7	180	240	420	180	180	520	176	86	1982	248
Task 8	30	50	30	20	30	60	33	21	274	34
Total	654	1052	1040	563	570	1225	550	314		
Average	82	132	130	70	71	153	69	39		

Errors

The team members noted the number of clicks each participant made during the task sessions. From these numbers, the lowest number for each task became the baseline. The number of errors were counted as the number of clicks over the established baseline.

Participants encountered the most errors during Task 7: Complete an exercise. There 92 errors made during this task and the average was 11.5. Most of these errors were encountered with participants 6, 7, and 8. The next highest error rate was with Task 1: Complete a Profile. The average error rate was 7.1 with a total of 57 errors. Both tasks included a lesson in the foreign language that was being studied. Tasks 2 (Add a language) and 5 (Change a setting) had the least amount of errors with 1 error (0.1 average) and 7 errors (0.9 average). Task 3: Change the language had 18 errors with an average of 2.3. Task 4: Change auditory setting had 30 errors with a 3.8 average. Task 6: Create a profile had 41 errors with an average rate of 5.1. Task 8: Turn on notifications had 20 errors with a 2.5 average. The participants encountered no critical errors on any tasks and were able to successfully complete all 8 tasks.

Errors (# of clicks)										
	P1	P2	P3	P4	P5	P6	P7	P8	Total	Average
Task 1	8	12	14	13	5	9	15	21	97	12.1
Task 2	1	1	1	1	1	1	1	2	9	1
Task 3	2	2	3	2	3	3	7	12	34	4.3
Task 4	5	5	5	5	4	5	15	2	46	5.8
Task 5	3	3	3	3	3	4	2	2	23	2.9
Task 6	13	16	14	17	15	17	12	9	113	14.1
Task 7	18	19	23	21	21	45	40	49	236	29.5
Task 8	1	2	4	3	3	2	8	5	28	3.5
Total	51	60	67	65	55	86	100	102		
Average	6.4	7.5	8.4	8.1	6.9	10.8	12.5	12.8		

Errors (# over base)										
	P1	P2	P3	P4	P5	P6	P7	P8	Total	Average
Task 1 (5)	3	7	9	8	0	4	10	16	57	7.1
Task 2 (1)	0	0	0	0	0	0	0	1	1	0.1
Task 3 (2)	0	0	1	0	1	1	5	10	18	2.3
Task 4 (2)	3	3	3	3	2	3	13	0	30	3.8
Task 5 (2)	1	1	1	1	1	2	0	0	7	0.9
Task 6 (9)	4	7	5	8	6	8	3	0	41	5.1
Task 7 (18)	0	1	5	3	3	27	22	31	92	11.5
Task 8 (1)	0	1	3	2	2	1	7	4	20	2.5
Total	11	20	27	25	15	46	60	62		
Average	1	3	3	3	2	6	8	8		

Summary of Data

A summary of the data gathered during the usability test is presented below. The tasks with longer times for completion (Tasks 1, 6 and 7) were also the tasks with the highest averages of errors. The task with the least average of errors (Task 2) had the shortest average time for completion. The remaining four tasks (3, 4, 5, 8) did not show a strong relationship. For example, Tasks 4 and 5 both took an average of 47 seconds to complete but Task 4 had an error rating of 3.8 while Task 5 had an error rating of 0.9.

Summary

	Task Completion	Learnability (Time on Task)	Errors
Task 1	8	168	7.1
Task 2	8	27	0.1
Task 3	8	30	2.3
Task 4	8	47	3.8
Task 5	8	47	0.9
Task 6	8	145	5.1
Task 7	8	248	11.5
Task 8	8	34	2.5

Questionnaires

Task Ease of Use Questionnaire

The participants were asked to rate the ease of completing each task with 1 meaning super easy, 3 meaning not easy or hard, and 5 meaning super hard. The totals were added and averaged per task and per participant.

The majority of the participants, 88%, agreed that creating an account (Task 1), adding a language (Task 2), changing the language (Task 3), and turning on notifications were the easiest of the tasks. These tasks had average ratings of 1.5 to 1.75. The participants were less in agreement on Tasks 4, 5, 6, and 7 but had higher ratings of difficulty with a range of 2.13 to 2.25. A lesser majority of participants, 75%, rated the tasks of creating a profile (Task 6) and turning on the auditory sounds (Task 4) as easy. Only 63% of the participants agreed that changing a setting (Task 5) and completing an exercise (Task 7) were easy.

Per participant, an-all easy score would be 8 and an all-difficult score would be 40. The total per participants ranged between 12 and 21. The total group of participants agreed that the tasks were easy overall with an average per participant ranging from 1.75 to 2.63.

Task Questionnaire											
	P1	P2	P3	P4	P5	P6	P7	P8	Total (per task)	Average (per task)	% of easy ratings
Task 1	1	1	1	2	1	2	1	3	12	1.5	88%
Task 2	2	1	1	3	1	2	1	1	12	1.5	88%
Task 3	2	1	1	2	1	2	3	2	14	1.75	88%
Task 4	2	2	1	2	2	3	3	2	17	2.13	75%
Task 5	3	3	2	2	2	3	1	2	18	2.25	63%
Task 6	2	2	1	2	2	3	1	3	16	2	75%
Task 7	2	1	2	4	3	4	2	2	20	2.5	63%
Task 8	1	1	2	2	1	2	2	3	14	1.75	88%
Total (per participant)	15	12	11	19	13	21	14	18			
Average (per participant)	1.88	1.50	1.38	2.38	1.63	2.63	1.75	2.25			

* Scale: 1 (super easy) – 3 (not easy or hard) – 5 (super hard)
* Average (mean) rating per task = total of each task / 8 (# of participants)
* Average (mean) rating per participant = total per participants / 8 (# of tasks)
* Percent of easy ratings = total of [1] and [2] ratings / 8 (# of participants

Satisfaction Questionnaire

The participants were asked questions regarding their satisfaction with different aspects of the application. The scores were totaled and averaged. The participants were asked to rate on a scale of strongly disagree (1) to strongly agree (7).

For the purposes of determining the percentage that agreed, ratings of 5, 6 and 7 were added together and averaged by the total number of participants. All of the participants (100%) agreed that the application was simple to use (Q2) and that they were comfortable using it (Q6). They all agreed that they could effectively (Q3) and efficiently (Q5) use the app and recover from mistakes while using it (Q10). Many of the participants (88%) agreed that the app gave clear instructions for fixing errors (Q9). easy to understand information (Q13) and that the information was effective (Q14) and organized (Q15). Many participants, 88%, also agreed that they were satisfied with the app (Q19), with the ease of using the app (Q1/Q7), and that they could quickly complete tasks (Q4). Only 75% agreed that the app had all the functions they needed (Q18). The least percentage of agreement (63%) was in productivity (Q8), clear information for using the app (Q11) and the ease of finding information (Q12). Only 63% agreed that they liked using the interface (Q17) but 88% agreed that it was pleasant (Q16).

The items that received the lowest averages were clear information (4.43), ease of finding information (4.71), and likability of the interface (Q17). The highest agreement rating was for how simple the app was to use which had an average of 6.29. Effectiveness of the app in completing tasks and comfort using the app had the next highest agreement rating of 6.14.

The overall satisfaction ratings by participants ranged from a low of 3.95 to a high of 6.42.

Satisfaction Questionnaire

	P1	P2	P3	P4	P5	P6	P7	P8	Total (per question)	Average (per question)	% of agree ratings
Q1: Overall, I am satisfied with how easy it is to use this language learning app.	6	5	7	6	7	7	4	6	48	6.00	88%
Q2: It was simple to use this language learning app.	7	7	6	6	7	6	5	7	51	6.29	100%
Q3: I could effectively complete the tasks using this language learning app.	7	7	6	6	6	6	5	6	49	6.14	100%
Q4: I was able to complete the tasks quickly using this language learning app.	7	7	5	7	6	4	6	7	49	6.00	88%
Q5: I was able to efficiently complete the tasks using this language learning app.	7	7	6	5	6	5	5	7	48	5.86	100%
Q6: I felt comfortable using this language learning app.	7	7	7	4	6	5	7	6	49	6.14	100%
Q7: It was easy to learn to use this language learning app.	7	6	7	6	6	4	5	6	47	5.86	88%
Q8: I believe I could become productive quickly using this language learning app.	4	5	7	6	7	4	4	5	42	5.29	63%
Q9: This language learning app gave error messages that clearly told me how to fix problems.	6	5	4	5	5	6	5	7	43	5.14	88%
Q10: Whenever I made a mistake using the language learning app, I could recover easily and quickly.	5	6	6	6	6	5	6	7	47	5.71	100%
Q11: The information (such as on-line help, on-screen messages and other documentation) provided with this language learning app was clear.	5	6	4	6	5	4	1	6	37	4.43	63%
Q12: It was easy to find the information I needed.	5	4	6	5	6	4	3	6	39	4.71	63%
Q13: The information provided for this language learning app was easy to understand.	6	5	6	6	6	5	3	7	44	5.29	88%
Q14: The information was effective in helping me complete the tasks.	4	6	6	5	7	5	5	7	45	5.43	88%
Q15: The organization of information on the language learning app screens was clear.	7	6	6	6	7	5	2	6	45	5.57	88%
Q16: The interface of this language learning app (buttons, menus, graphics) was pleasant.	6	6	7	6	7	6	4	7	49	6.00	88%
Q17: I liked using the interface of this language learning app (buttons, menus, graphics).	4	7	5	4	7	6	1	7	41	4.86	63%
Q18: This language learning app has all the functions and capabilities I expect it to have.	6	6	6	4	6	5	2	6	41	5.00	75%
Q19: Overall, I am satisfied with this language learning app.	7	6	6	6	6	6	2	6	45	5.57	88%
Total (per participant)	113	114	113	105	119	98	75	122			
Average (per participant)	5.95	6	5.95	5.53	6.26	5.16	3.95	6.42			

* Scale: 1 (strongly disagree) – 4 (do not agree or disagree) – 7 (strongly agree)

* Average (mean) rating per question = Total per question / 8 (# of participants)
* Average (mean) rating per participant = (total per participant for all questions) / 19 (# of questions)
* Percent of agree ratings = total of [5, 6, 7] ratings / 8 (# of participants)

User experience

Most users felt the app was easy to use and provided a positive experience. Functionality and navigation were largely intuitive with the learning path easy to follow. The user interface is pleasing, and the inclusion of gamification elements makes using the app feel a bit less like a chore. One user was critical of the simplified controls and felt the bright colors and characters made it feel childish. Nevertheless, all users successfully completed the proscribed tasks with few errors and within a reasonable period of time. One user noted that although the app was easy to use and intuitive, there is no way of evaluating if the app is effective at teaching the language and if the results of the language assessment is accurate. However, it was noted that the reminders and notifications for daily practice should make the learning more effective.

Recommendations

The recommendations section provides recommended changes and justifications driven by the participant success rate, behaviors, and comments. Each recommendation includes a severity rating. The following recommendations will improve the overall ease of use and address the areas where participants experienced problems or found the interface/information architecture unclear.

Change	Justification	Severity
• Labels for Settings	• One of the complaints voiced most often was that the tabs that directed users to the settings were not labeled, which made it difficult to perform several tasks required for this test.	• High
• Limit Pop-ups	• One complaint was that Duolingo enables too many pop-ups when using their app. This included ads, in-lesson pop-ups with motivational messages like "You're getting good at this!", or warnings that the questions are about to get harder. This seemed to bother, as it apparently broke their concentration when focusing on the lessons.	• Medium

Conclusion

Implementing the recommendations and continuing to work with users (i.e., real lay persons) will ensure a continued user-centered website.

This test showed that Duolingo users found the app to be easy to use when taking tests, and well organized, but a little lacking in terms of direction

when navigating the app. Most users were able to easily complete lessons, however, they had some trouble getting to the app's settings. While most said the app was easy to use, there were complaints that they were not really learning the language by using the app, so it may take a different test to determine the success of this app as a language learning method.

Attachments