Evaluating the Experience of Dr. Bill

Rebekah Ku, Fernanda Moreno, Shivali Pandya
Spring 2020

RESEARCH QUESTIONS

The goal of Dr.Bill is to prepare students for interviews by simulating a real job interview with practice behavioral and technical questions tailored to the user's resume. Therefore, for this phase of the project, our team wanted to evaluate the practice interview (the middle) part of the experience. After brainstorming many research questions, we decided to evaluate the following research question and subquestions which focus on the aspects of the application which attempt to create a realistic job interview:

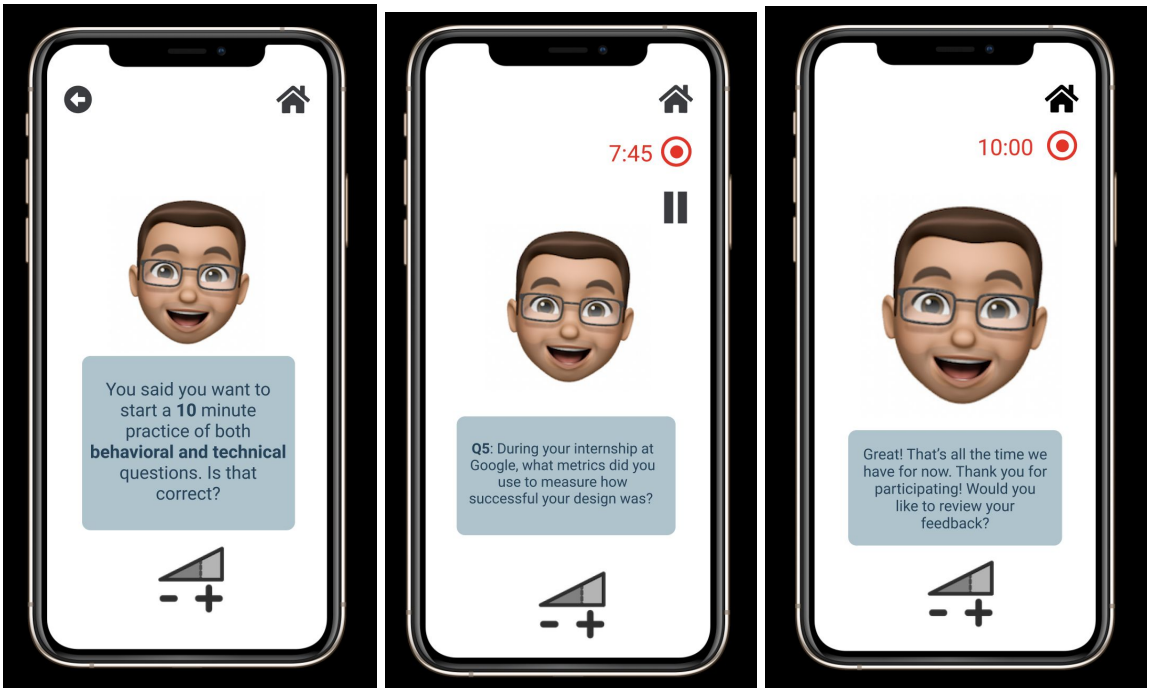
I. How does the voice interview help the user gain knowledge about a real job interview?
   A. How do the questions tailored to the user's resume help the user?
   B. Do the image and tone of Dr. Bill successfully portray a knowledgeable recruiter to users?

PROTOTYPE DESIGN

Link to our full Figma prototype:
https://www.figma.com/file/JDcz8OMdZxblJxUsTbUzqY/Interview-Screen-Eric?node-id=0%3A

Video of prototype interaction (4mins 26 secs):
https://drive.google.com/file/d/1RfdXn0dmY0PuI0I2wybp-aRBhreB1ch3/view?usp=sharing

Image highlights:



SCENARIO AND TASKS

The scenario we imagined for this particular case is as follows. Person A (our user tester) has an upcoming interview for a remote summer internship. Person A finds they have some spare time in between classes and wants to use that time to practice for their upcoming interview. Since it's hard and impractical to find another person to help Person A practice in such short notice, Person A decides to use Dr. Bill to help them practice for 10 minutes.

To answer our research questions, we gave the evaluators a short series of tasks to complete, as they opened the application and went through the practice interview:

1. Start a 10-minute mock interview.
2. Start an interview that consists of both behavioral and technical questions.
3. Confirm your interview.
4. Start the interview and answer all the questions. (Keep in mind that you don't have to answer them for real).
5. Pause the interview during the first question, and resume the interview again.
6. On question 4, click on the Home button but return back to the interview.

METHODOLOGY

We used Figma to create an interactive prototype of the practice interview. Dr. Bill would require incredibly powerful natural language processing capabilities, so our prototype relied heavily on the wizard of oz technique. In Figma, we designed our screens so that clicking on different spots would navigate the user to a specific screen. Because we needed more control of the prototype during the evaluation to make sure that the correct screens show up based on the evaluator's

responses, we had one team member share their screen with the evaluator through BlueJeans. We instructed the evaluators to say "click [button name]" if they wanted to click any of the buttons on the screen which we would then click for them. We collected data by observing reactions and comments made by the evaluators as well as from the heuristic evaluation sheet.

HEURISTICS

We began designing our heuristics by first reading the Nielsen Norman Group's ten usability heuristics and researching principles of voice interaction. We then adapted Nielsen's heuristics to answer our research question. To make the practice interview more realistic, Dr. Bill asks users questions based on their resume. This means that the application must be able to handle a large spectrum of user responses, so we included a heuristic for tolerating ambiguity. The system persona is another key aspect of the interview. Dr. Bill should closely replicate the expertise and sound of a recruiter; we created a heuristic for tone of voice. Lastly, because the application uses a more conversational style of voice interaction, we wanted to make sure that it was still easy for users to figure out what buttons and commands they could say during the interview to play, pause, or exit the interview. We chose ten heuristics that were given to the evaluators on the heuristic evaluation sheet with a place to leave comments, suggestions, and rate each heuristic on a severity scale.

Heuristic	Description
1. Visibility of system status	The system should always keep users informed about what is going on with appropriate feedback within reasonable time.
2. Match between system and the real world	The system should speak the user's language with words, phrases, and concepts familiar to the user.
3. User control and freedom	Users often choose system functions by mistake and users should easily be able to leave an unwanted state.
4. Tolerate ambiguity	The system should understand reasonable synonyms or paraphrased commands such as "stop interview" and "end interview".
5. Error prevention	System presents users with a confirmation option or displays users' selection before they commit to an action.
6. Recognition rather than recall	The system minimizes the user's memory load by making objects, actions, and options visible.
7. Tone of Voice	Dr.Bill's tone of voice is polite and appropriate for the situation. It simulates the tone of a real interviewer.
8. Discoverability	System voice commands are easy to discover. Button functionality is quickly identifiable.

9. Aesthetic and minimalist design	System displays information that is necessary and relevant to the current task. Buttons displayed enhance the interview.
10. Help and documentation	If the user needs help, information is easily accessible. It is concise and focused on the user's task.

HEURISTIC EVALUATION SHEET

Dr.Bill Heuristic Evaluation

Evaluator Name:

Device Used:

Task: Conducting a practice interview

Severity Scale

0 Not an issue
1 Aesthetic design flaw
2 Minor usability issue
3 Major usability issue
4 Catastrophic usability issue; must be fixed immediately

	Issue	Recommendation
1. Visibility of system status The system should always keep users informed about what is going on with appropriate feedback within reasonable time.		
2. Match between the system and the real world The system should speak the users' language, with words, phrases and concepts familiar to the user.		
3. User control and freedom Users often choose system functions by mistake and users should easily be able to leave an unwanted state.		
4. Tolerate ambiguity The system should understand reasonable synonyms or paraphrased commands such as "stop interview" vs. "end interview".		
5. Error prevention System present users with a confirmation option or displays users' selection before they commit to the action.		

Dr.Bill Heuristic Evaluation

Evaluator Name:

Device Used:

Task: Conducting a practice interview

Severity Scale

0. Not an issue
1. Aesthetic design flaw
2. Minor usability issue
3. Major usability issue
4. Catastrophic usability issue; must be fixed immediately

6. Recognition rather than recall

The system minimizes the user's memory load by making objects, actions, and options visible

○ 0 ○ 1 ○ 2 ○ 3 ○ 4

Issue

Recommendation

7. Tone of Voice

Dr. Bill's tone of voice is polite and appropriate for the situation. It simulates the tone of a real interviewer.

○ 0 ○ 1 ○ 2 ○ 3 ○ 4

Issue

Recommendation

8. Discoverability

System voice commands are easy to discover. Button functionality is quickly identifiable.

○ 0 ○ 1 ○ 2 ○ 3 ○ 4

Issue

Recommendation

9. Aesthetic and minimalist design

System displays information that is necesary and relevant for the user's current task. Buttons displayed enhance the interview.

○ 0 ○ 1 ○ 2 ○ 3 ○ 4

Issue

Recommendation

10. Help and documentation

If the user needs help, information is easily accessible. It is concise and focused on the user's task.

○ 0 ○ 1 ○ 2 ○ 3 ○ 4

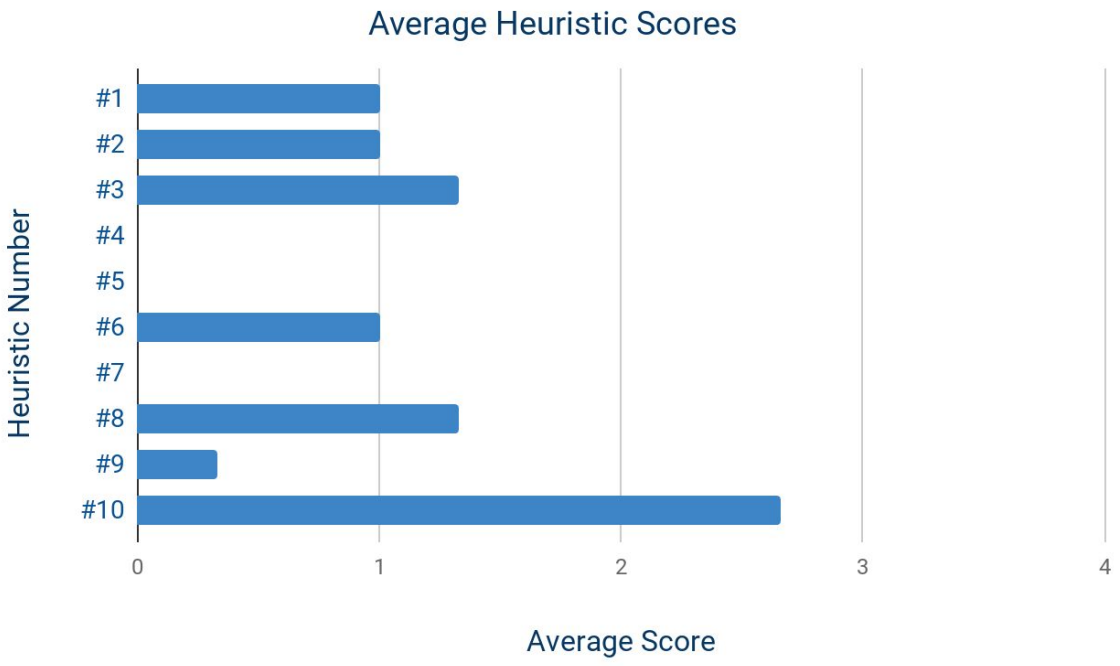
Issue

Recommendation

DATA

We reached out to three of our classmates and asked them to evaluate our Dr. Bill app prototype. Prior to each evaluation session, we asked our evaluators to send us their resume so that the mock interview session with Dr. Bill would include some questions that were tailored to the user's resume and the type of job they were interviewing for. The evaluation session took place via a BlueJeans call, with one of our team members sharing the screen of our prototype with the evaluator and another team member acting as the voice of Dr. Bill. The evaluators were given a list of tasks to complete as they navigated our prototype.

After each person tested our prototype, we asked them to fill out our heuristic evaluation sheet and send it over to us as soon as they could. The heuristic evaluation sheet listed the 10 heuristics we wanted to evaluate and asked the evaluators to rank the severity of the issue (from 0 being no issue to 4 being catastrophic usability issue) for each heuristic. The sheet also asked them to describe the issue and provide any recommendations for each heuristic. To visualize the data collected from each evaluator, we took the average of the severity scores for each heuristic and created a bar chart.



The chart above shows us that heuristic #10, Help and Documentation, was ranked as the most severe issue by our evaluators. It also tells us that our evaluators had no issues with heuristics #4 (Tolerate Ambiguity), #5 (Error Prevention), and #7 (Tone of Voice). Based on the issue rankings, descriptions, and recommendations, we learned that the primary issue with our prototype was the lack of a help button or screen, resulting in users being unsure of what to say or where to go if they needed help while navigating the app. Users also reported an issue with the User Control and Freedom heuristic (#3), and pointed out that our app may rely a little too much on voice control to navigate through it.

RECOMMENDATIONS

One of our users recommendations for future iterations is to include a help button with some sort of help documentation. Although our users didn't have any problems during the testing of the app, they still wanted something that they could reference in the case that they did run into problems. They also recommend that some actions, like quitting a mock interview, should also be available as buttons instead of voice commands. Another recommendation is to add a progress bar to the top of the interview screens to allow users to see their progress through the interview. Some users felt a bit lost during the interview and wanted to know how much of the interview was left.

Overall, we found that the prototype did well in helping users by having tailored questions. Users were surprised and pleased to see questions specific to them. In fact, one user even commented that one of the questions Dr. Bill asked was similar to a question they received in a previous interview. Additionally, although we told our user testers that they didn't have to answer the mock interview questions for real, many took the time to go through and thoughtfully answer the questions. Likewise, we also found that Dr. Bill successfully portrayed a knowledgeable recruiter to the user testers. User testers thought that Dr. Bill's voice was fitting and appropriate as a recruiter. However, one user noted that at certain screens, Dr. Bill's word choice seemed a bit too stiff and robotic, breaking Dr. Bill's image as a recruiter. As such, moving forward, it would be recommended to alter Dr. Bill's word choice, particularly on user confirmation screens, in such a way that feels less robotic and doesn't break the users' image of Dr. Bill as a recruiter.