

# IMMERSE



## FINAL REPORT

*Practice English through immersing yourself in your favorite TV shows*

***Designed By:*** Becky Weinstein and Kimberly Tran

***Developed By:*** Cathy Zhang and Andrea Collins

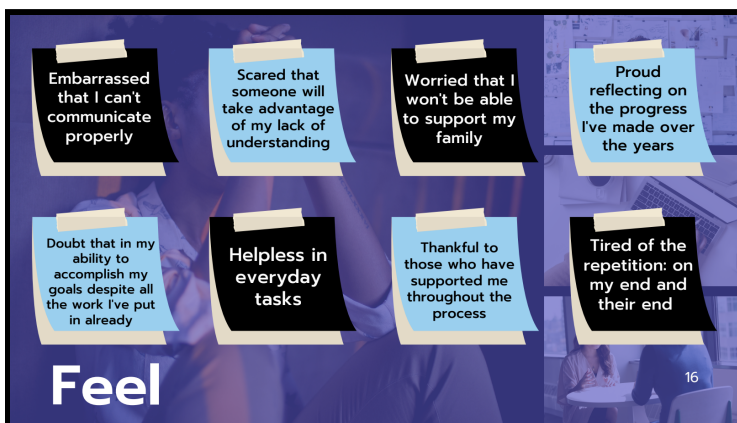
## Problem and Solution Overview

Speakers of English as a second language struggle with topic-specific vocabulary—such as going to a doctor’s office or filling out government forms—that is often not covered in English-language classes.

Immerse is an interactive content platform where users can play the role of a character in a TV episode, recording their character’s lines as they watch the video. Users receive immediate AI-generated feedback on their pronunciation, allowing them to learn new vocabulary and assimilate to new contexts, all while watching their favorite TV shows.

## Needfinding Interviews

Our first step was to chat with English-as-a-second-language (ESL) speakers to hone in on their most pressing pain points, as well as to learn about what solutions have worked well for them. Specifically, we talked to a variety of ESL speakers, covering a total of seven native languages, ages 20 to 50 years old, and over eight countries.



Highlighting a few of our interviewees, we spoke with Thuy, a mother from Vietnam who immigrated 10 years ago and avoids small talk because of insecurity in her ability to converse. We also spoke with Mevini, a graduate student



from Sri Lanka who moved to the U.S. for school and finds herself Googling phrases after conversations because she's reluctant to ask for clarification. Through our conversations, we developed a deeper understanding of and empathy for the ongoing hardships faced by ESL speakers, even after they have achieved "fluency."

And as we had more conversations, a few themes began to reveal themselves as prevalent across much of our user group.

#### Problems:

1. Contexts involving a specific set of vocabulary—such as going to a doctor's office, renting an apartment, or being pulled over by a police officer—cause uncertainty for our users, who feel topic-specific vocabulary is under-emphasized in English classes.
2. Pronouncing new vocabulary causes discomfort to our users, who fear embarrassment if they pronounce words incorrectly.

#### Insights:

1. Watching multimedia in English helps our users learn cultural references and slang, as well as hear pronunciation of new words.
2. Practicing new words out loud was found to be a key part of learning their pronunciation and correct usage.



## POVs and Experience Prototypes

After further unpacking our needfinding results and interviewing even more participants, we crafted point of view (POV) statements for our users.

1. We met Thuy, a mother who immigrated from Vietnam 10 years ago. We were amazed that after a decade of speaking English, Thuy is still uncomfortable with small talk. It would be game-changing to help her feel more comfortable in a variety of English-speaking situations.
2. We met Yuri, an international student from Japan who is still not completely comfortable with using English beyond the classroom. We were amazed how dependent she is on visual cues for verbal communication. It would be game-changing to help her to rely less on visual and physical cues.
3. We met Alejandro, who immigrated with his family as a child from Ecuador and often had to serve as his mother's translator and communicator. We were amazed to learn that he is exhausted from having to constantly support his mother as a translator and communicator. It would be game-changing to provide Alejandro with a resource to enable him to alleviate his mother's dependence.

Using these POVs, we then wrote "How might we?" (HMW) statements around the identified pain points of small talk, ability to converse in various contexts, and difficult-to-understand speech.

1. How might we make engaging in small talk more accessible and fluid for adult ESL speakers in their daily lives? (POV: Thuy)
2. How might we expose adult ESL speakers to various English-speaking contexts in which they may feel uncomfortable? (POV: Thuy)
3. How might we decrease the difficulty of understanding fast and/or hard-to-understand speech for adult ESL speakers? (POV: Yuri)

These HMWs helped inspire our solution brainstorming. As a team, we conducted multiple rapid-fire brainstorming sessions and ultimately narrowed down to three solution ideas.

1. Source short videos to help ESL speakers learn and understand colloquialisms (HMW #1)

2. Provide visual cues to ESL speakers during their conversations (HMW #1)
3. Teach ESL speakers conversational English through a game or gamified experience (HMW #2)

The last step of this process was to build experience prototypes to test our solutions' assumptions.

1. For solution #1, our assumptions were: 1) people feel engaged while watching short videos, and 2) people can learn from watching short videos. For this prototype, we interviewed Dora, a dermatologist in her 50s whose first language is Hebrew. We began by asking Dora two questions: 1) "Is 'gif' or 'jif' the correct pronunciation?" and 2) "Which is the correct meaning of 'poggers,' a common slang term?" Dora answered both questions incorrectly. We then showed her two short videos that we sourced from TikTok, a video of the pronunciation of "gif" and a video of a woman using "poggers" in a sentence, and re-asked the questions. Dora answered both questions correctly this time and said that on a scale of 1-5, she was a 4 for "How likely are you to retain this information?" and a 5 for "How much did you enjoy learning in this format?" Thus, our takeaway here was that people can successfully engage with and learn from content in a short video format. Were we to do this again, we would reconsider the videos we sourced, as the colloquialisms could have been more applicable to the target user.
2. For solution #2, our assumptions were: 1) visual cues are helpful during a conversation, and 2) visual cues are not distracting or stressful during a conversation. For this prototype, we interviewed Jenny, an engineer in her 50s whose first language is Mandarin. We hopped on a Zoom call with Jenny, during which Andrea began by verbally telling a story about cooking empanadas without any visuals, and then concluded by telling a story about cooking biryani while screen-sharing slides with images. Jenny shared that the visual references helped her make guesses when she didn't understand biryani ingredient names, as well as visualize and feel the sentiments Andrea was trying to convey. Thus, our takeaway here was that visuals during a conversation can assist with ESL speakers' comprehension. In reflecting on this test, static images were not as engaging or useful as perhaps multimedia would be.

3. For solution #3, our assumptions were: 1) people will be engaged playing a game to learn English, and 2) playing a game is an effective way to retain English. For this prototype, we interviewed Will, a refugee who immigrated from Vietnam and is in his 60s. For Will, we created a Google Form choose-your-own-adventure game that walked him through the process of ordering food at a diner. Will reflected how the visuals were helpful in understanding the words within the context. Furthermore, the ability to repeat the process of ordering food is beneficial for ESLs because there are different options which enable them to diversify their responses and also practice the same flow to become more comfortable with the situation. Thus, our takeaway here was that exposure and repetition allow ESLs to learn and retain English. If we were to test this again, we would implement more comprehension features to help reinforce vocabulary at the end of the task flow.

Conducting these experience prototypes provided us useful perspective into ESLs' learning experiences and illuminated the importance of visuals, interaction, and repetition in the English learning process.

## Design Evolution

### Solution

Based on all of the insights and feedback we received from needfinding interviews, experience prototypes, and heuristic evaluation feedback from peers and staff, we identified the following key findings. ESL speakers struggle with specific conversation topics that are not taught in class, face difficulty and can be unconfident with English pronunciation, and value interactivity in the English learning process. Our initial solution, which aligns with our low-fi prototype, was an interactive mobile application that allows users to practice topic-specific conversations with a bot. After receiving feedback about the room for novelty in our solution, we decided to revise our tasks and solution for the medium-fi prototype (final version of tasks are listed below). Our final solution, which aligns with our medium-fi and hi-fi prototypes, is an interactive mobile application where users can practice and receive AI-generated feedback on pronunciation and vocabulary, all while watching and playing the roles of characters in their favorite TV shows.

## Tasks

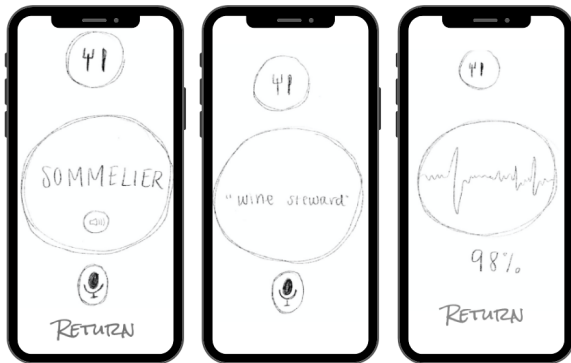
1. Simple: Allow the user to choose a topic to study
  - a. We chose this task because of the many times we heard about ESL speakers struggling with topic-specific conversations like going to the doctors' office or getting a traffic ticket and speaking to police officers.
2. Moderate: Allow the user to practice topic-specific vocabulary
  - a. We chose this task because a speaker must first have basic vocabulary in their toolkit in order to engage in topic-specific conversations.
3. Complex: Allow the user to practice and receive AI-generated feedback on pronunciation of topic-specific phrases and sentences
  - a. We chose this task because of how important yet difficult native pronunciation is for ESL speakers. It would be effective to use AI technology to provide active and actionable feedback to users, so that our app can help build ESLs' confidence in speaking English aloud.

## Major Design Changes (low-fi → med-fi)

### *Vocabulary-in-Context*

In our low-fi prototype, our vocabulary practice consisted of text-based and predetermined sets. Prototype testers felt that our implementation of the vocabulary task was not helpful for users to actually learn and retain vocabulary, so we redesigned our medium-fi prototype to place the emphasis on seeing vocabulary in the context of a sentence, as opposed to practicing vocabulary in isolation. In terms of UI, we used bolding and highlighting to emphasize vocabulary words in the context of sentences and called out vocabulary words and definitions in actual boxes on the screen.

## Low-Fi



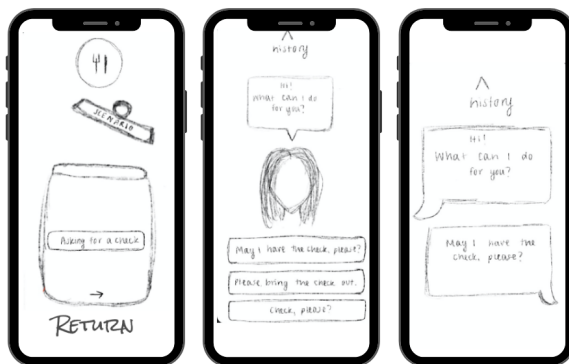
## Medium-Fi



### *Practicing Conversations Through Interactive Videos, Rather Than Templated Conversations*

In our low-fi prototype, users could select a topic (i.e. eating at a restaurant) to practice a conversation with a bot. We received feedback about room for novelty in how we carried out this task, so we made a major task and design change. In our medium-fi prototype, users instead select a topic to experience by watching a video. Rather than following a templated conversation with a bot, users play the role of a character throughout a video, recording their lines and receiving feedback on pronunciation.

## Low-Fi

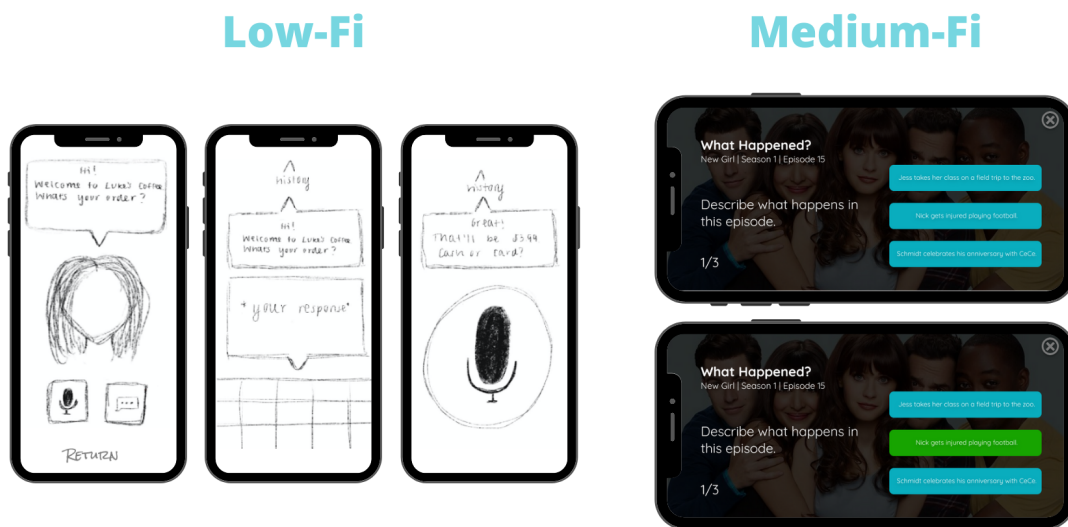


## Medium-Fi



## Assessing Comprehension Through Video Quizzes, Rather Than Free-form Bot Conversations

In our low-fi prototype, users were tested on their ability to follow a conversation by participating in a free-form conversation with a bot. Similarly, we received feedback about room for novelty in how we carried out this task and thus made a major task and design change. In our medium-fi prototype, users are instead tested on their understanding of a video by answering questions about plot and vocabulary. (Note that we ultimately removed this feature for our React Native implementation.)

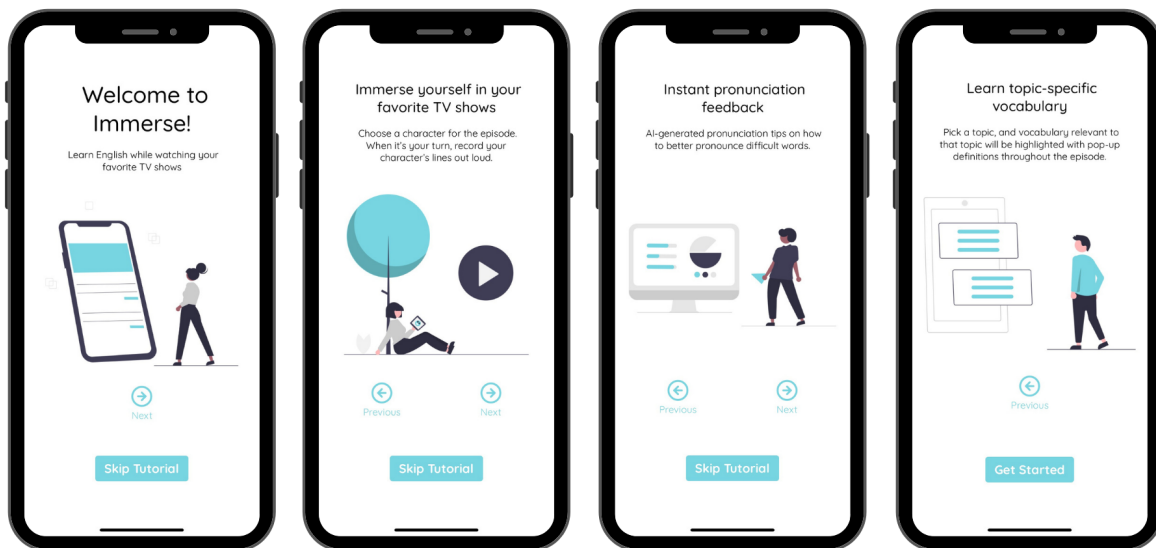


### Major Design Changes (med-fi → hi-fi)

#### Tutorial Screens

Our evaluators filed a severity 3 violation of *H10. Help and documentation*, noting that there is no onboarding screen for the app in general and what pronunciation exercises are. To address this violation and help onboard new

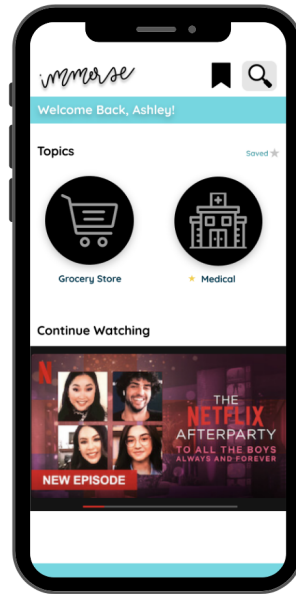
users to Immerse, we added three tutorial screens at the beginning of the app flow that introduce our main features.



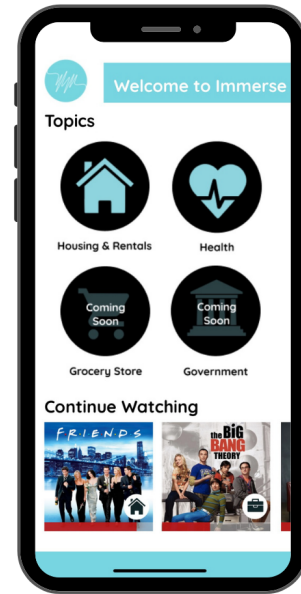
### Home Screen

Our evaluators provided feedback about minimalism and consistency across the app, noting severity 3 violations of *H8. Aesthetic and minimalist design* and *H4. Consistency and standards*. Specifically, they mentioned that text, buttons, and icons were inconsistent. To address the violations on the home screen, we made icons and text uniform in color and embellishment and removed icons and features (My List, Saved, search) that evaluators found unintuitive and busy.

Medium-Fi

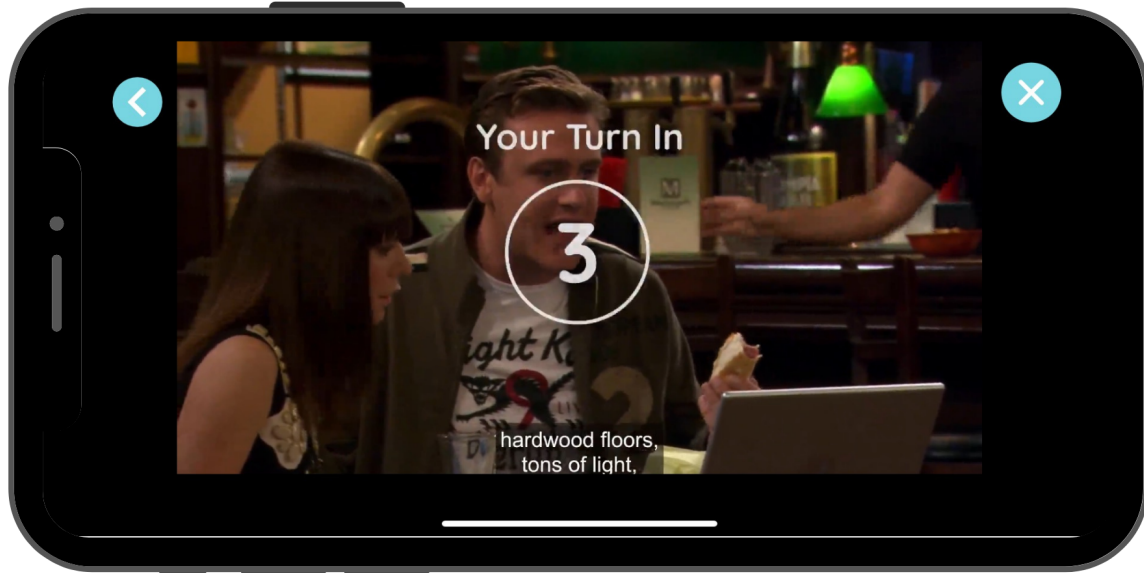


High-Fi



### Video Screen

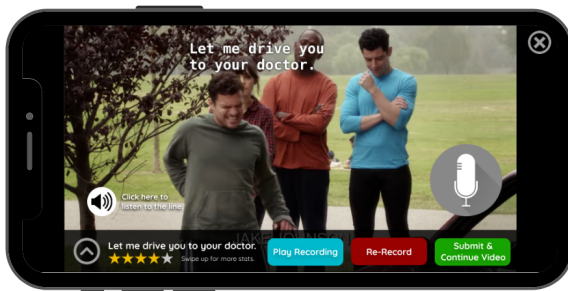
Our evaluators filed a severity 3 violation of *H1 - Visibility of system status*, noting that users have no forewarning of when a pronunciation exercise will be prompted, which can be stressful. To address this violation and help minimize any stress induced on the ESL learner, we included a 3-2-1 countdown before each pronunciation exercise as a heads-up to the user.



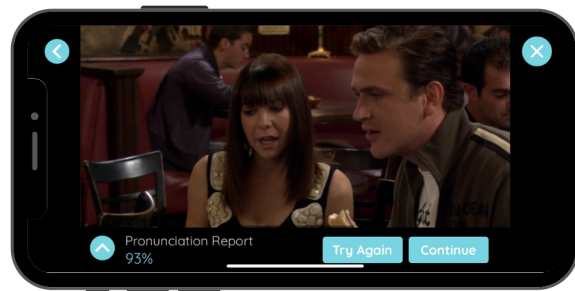
### Pronunciation Screen

The severity 3 *H8* and *H4* violations that applied to the home screen also applied to the pronunciation screen. In addition to general inconsistencies with text, buttons, and icons, evaluators mentioned that there were unnecessary button-label redundancies. We addressed these violations by making icons and buttons uniform in color and embellishment, removing redundant text from captions, and moving all interactive buttons that were scattered across the video screen to the bottom horizontal bar.

### Medium-Fi



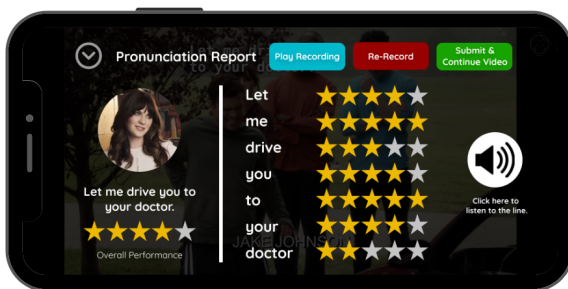
### High-Fi



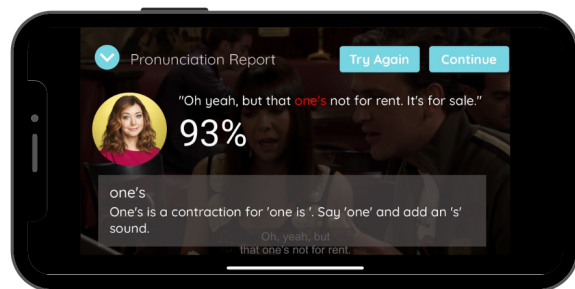
### Pronunciation Feedback Report Screen

Lastly, our evaluators filed a severity 3 violation of *H9 - Help users recognize, diagnose, and recover from errors*, noting that the existing pronunciation feedback report does not provide feedback or suggestions for improvement. Additionally, the UI is crowded and confusing, contributing to the screen's unhelpfulness. To address this violation and help users receive meaningful feedback via this screen, we redesigned the UI to provide a breakdown of the user's pronunciation performance by individual word, as well as actionable suggestions for incorrectly pronounced words.

### Medium-Fi



### High-Fi



## Hi-Fi Storyboard Walkthroughs

We designed our hi-fi prototype using Figma and BuilderX. Here are detailed storyboard walkthroughs of our hi-fi prototype.



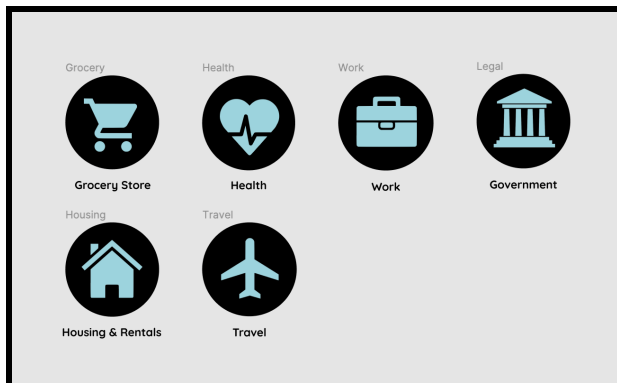
## Final Prototype Implementation

For our final prototype, we used a combination of Figma, BuilderX, React Native, iMovie, and Expo.

Figma allowed us to mock up much of the UI. Figma's shortcoming was the lack of an option to convert our designs into code. To address this, we used

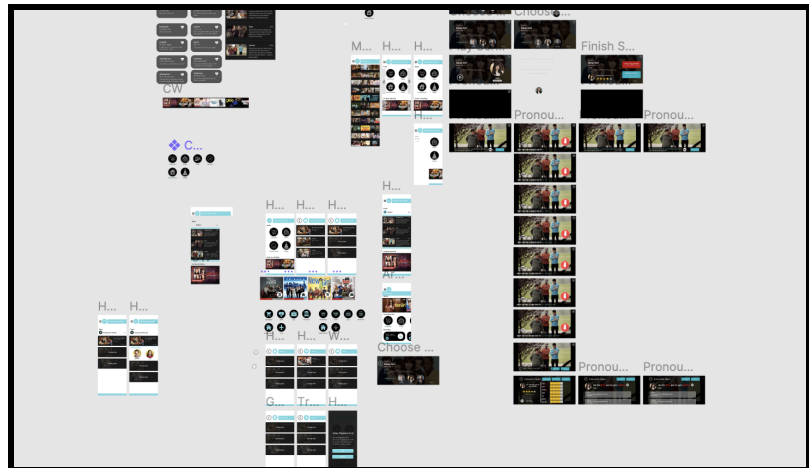
BuilderX—leveraging a combination of elements made in Figma and elements made natively in BuilderX—to assemble

most of our application's screens. BuilderX's drawback was in the primitiveness of the software, which from a UI standpoint is not as fleshed out as Figma and, to our detriment, didn't allow for joint editing.



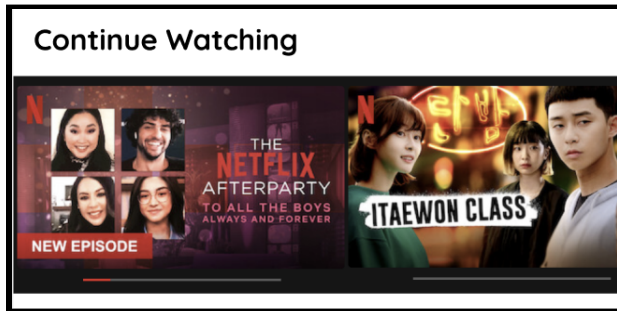
However, BuilderX was great for converting our screen designs into React Native code that we uploaded into Expo, which we used to build and publish our app. Expo was great as a tool, but a huge problem we ran into was a file size constraint, which meant we hit a wall when it came to trying to build our app. Our app, due to the quantity of media (images, videos, audio), was massive in its size, so the largest problem we encountered was in having to re-design our app to accommodate Expo's build limitations.

As far as shortcuts in our design, we used Wizard of Oz techniques on a few features. First, rather than using AI to generate pronunciation feedback, we provided the feedback ourselves. Second, the vocabulary pop-ups and line warnings seen throughout our videos are edited directly into our videos using



iMovie. Finally, the voice recording and speech recognition are imitated in our app, as opposed to actual implementation of these features.

Regarding hardcoded features, a “Continue Watching” section can be seen on our home page, featuring a scroll bar of available videos and a progress bar indicating the user’s previous place in these videos. Both the list of videos and the progress bar are hardcoded. In addition, the episode choices, episode descriptions, and character choices are all hardcoded.



## Summary and Next Steps

Throughout this quarter, we have learned the importance of conversation and iteration. We came into our project with certain conceptions about what was lacking from the language-learning space. This put blinders on our initial ideation, and it wasn't until we fully committed to our needfinding process, listening to the pain points and asks conveyed by our interviewees, that we were able to step outside the box and devise a solution we are all incredibly excited about. In this way, we leveraged the conversations we had to iterate through versions of our solution, ultimately reaching our final application design.

Of course, there is so much more we would love to implement on Immerse. We think there's huge potential for building out the number and diversity of contexts we offer our users, as well as the options within these contexts (i.e. number of episodes, character options, amount of vocabulary within each episode). On our current platform, several instances of a “Coming Soon” visual can be seen superimposed on icons indicating our available contexts and characters. Given more time, we would have loved to build out our platform to offer more of these options to our users to make the experience as helpful as possible in their English-learning endeavors.