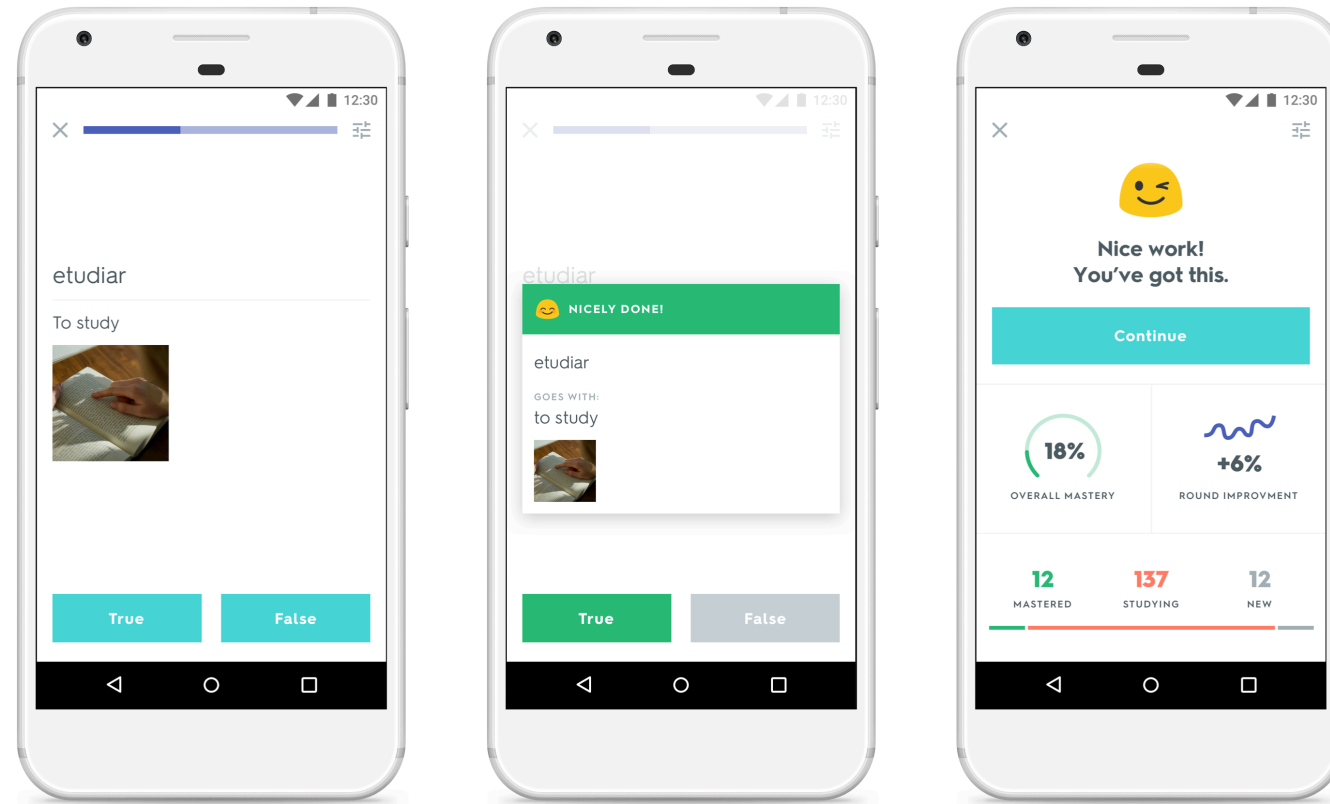


QuizBot: A Dialogue-based Adaptive Learning System for Factual Knowledge

Sherry Ruan¹, Liwei Jiang², Justin Xu¹, Bryce Tham¹, Zhengneng Qiu¹, Yeshuang Zhu³,
Elizabeth Murnane¹, Prof. Emma Brunskill¹, Prof. James Landay¹



Factual knowledge sometimes needs to be learned outside of a classroom setting



User interface of Quizlet

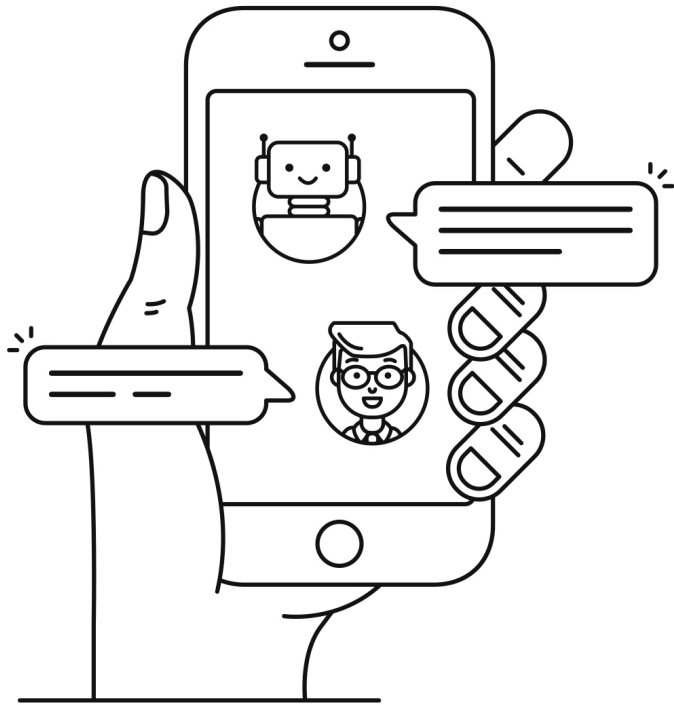
However, current electronic tools (flashcards) are passive and boring



We want to provide better automated ways to help students learn factual knowledge



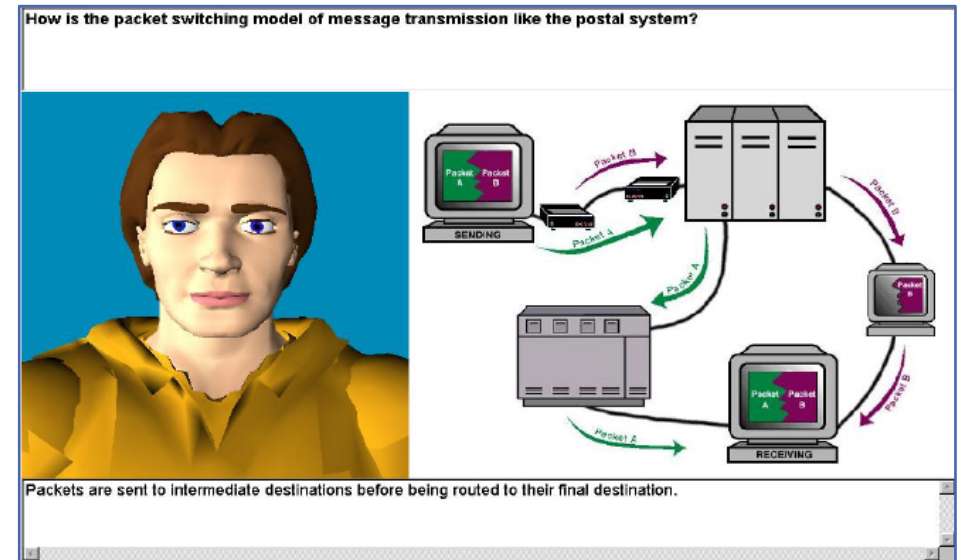
Chatbots are a promising solution for casual learning



- NLP grading algorithms are becoming sophisticated
- Having people interact can promote stronger learning than recall alone (Chi & Wylie 2014)
- Conversational UIs might be more interactive and engaging than flashcards

Chatbots are mostly used as formal study tools

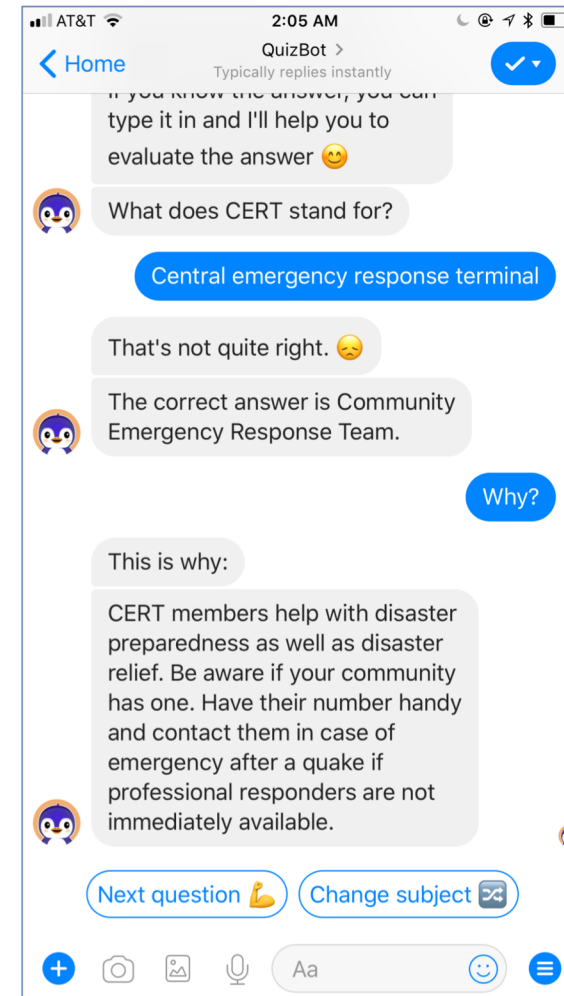
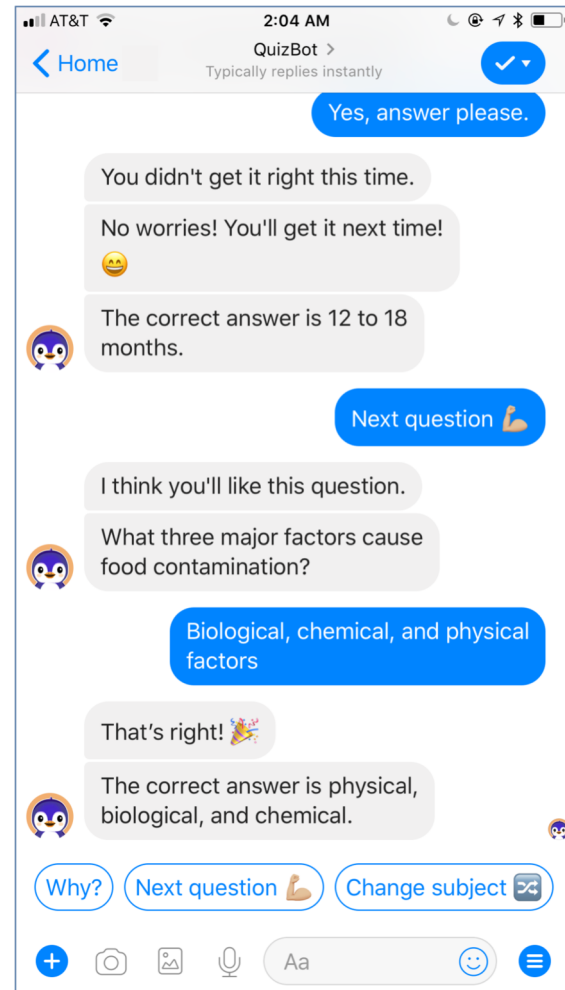
- Auto Tutor for computer literacy and physics (Graesser et al. 2004)
- Ms. Lindquist for high school algebra (Heffernan & Koedinger 2002)
- Piagetbot for college psychology (Heller & Procter 2007)



Interface of AutoTutor

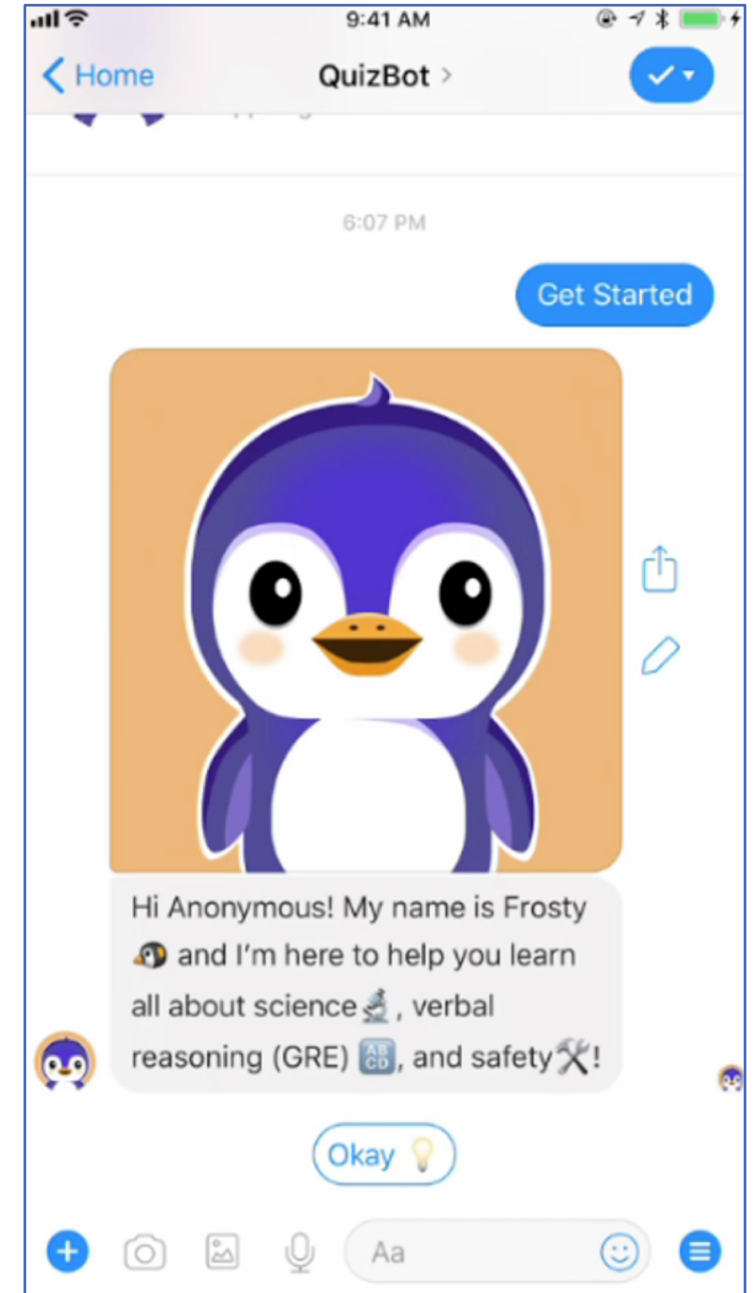
System Design and Implementation

QuizBot: conversational fact tutor



Key components of QuizBot

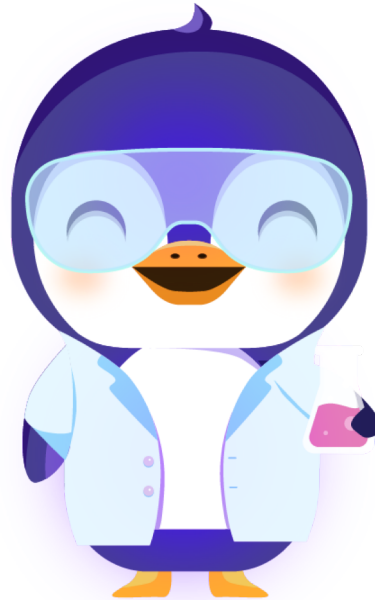
- Interactive chat interface
- Animated avatar with positive feedback
- NLP targeted feedback for typed answers (Arora et al. 2017)
- Ability to ask for hints/explanations
- Adaptive question sequencing algorithm (Lindsey & Mozer 2016)
- A mix of typing and button-based interactions



QuizBot teaches vocabulary, science, and safety



Vocabulary

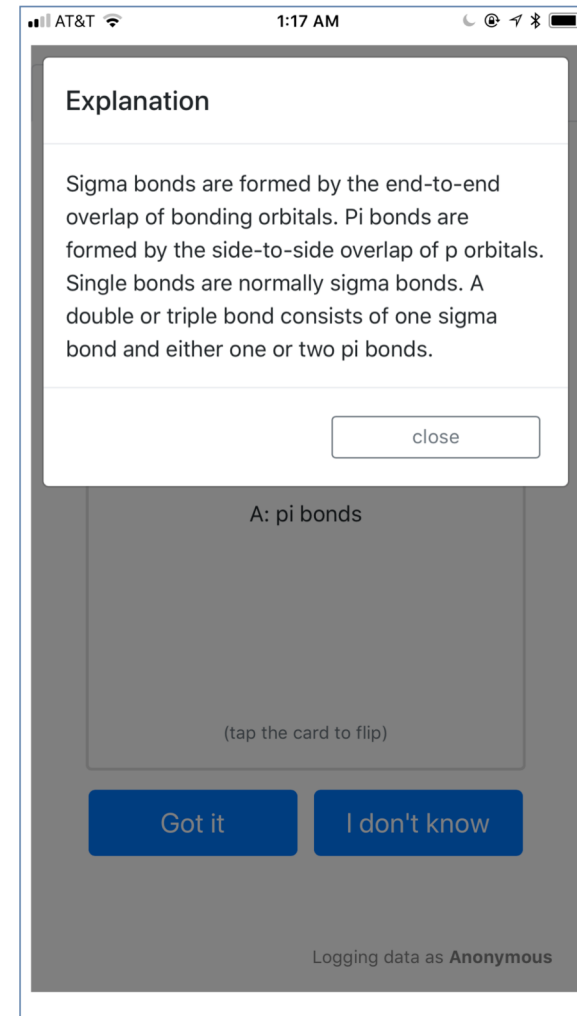
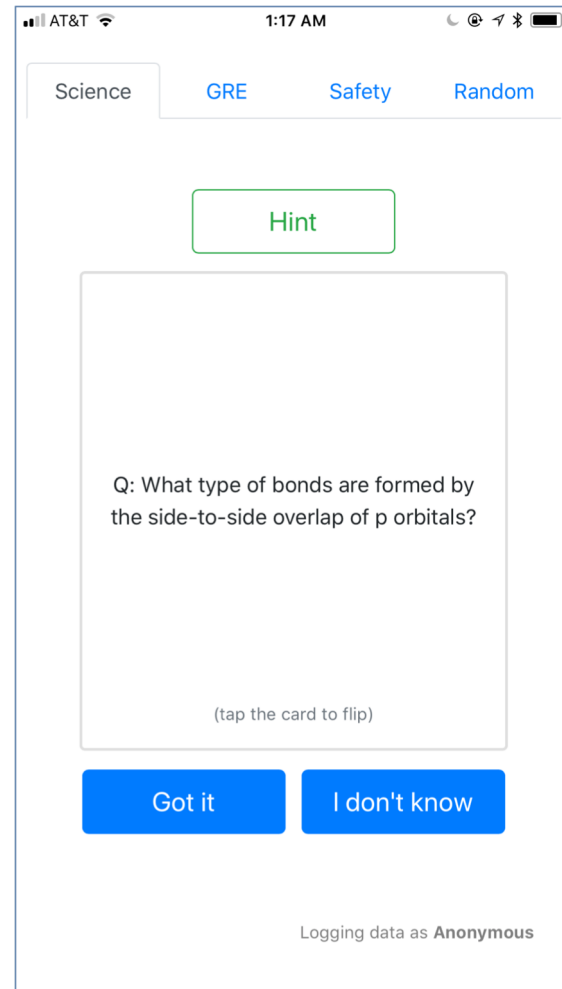


Science



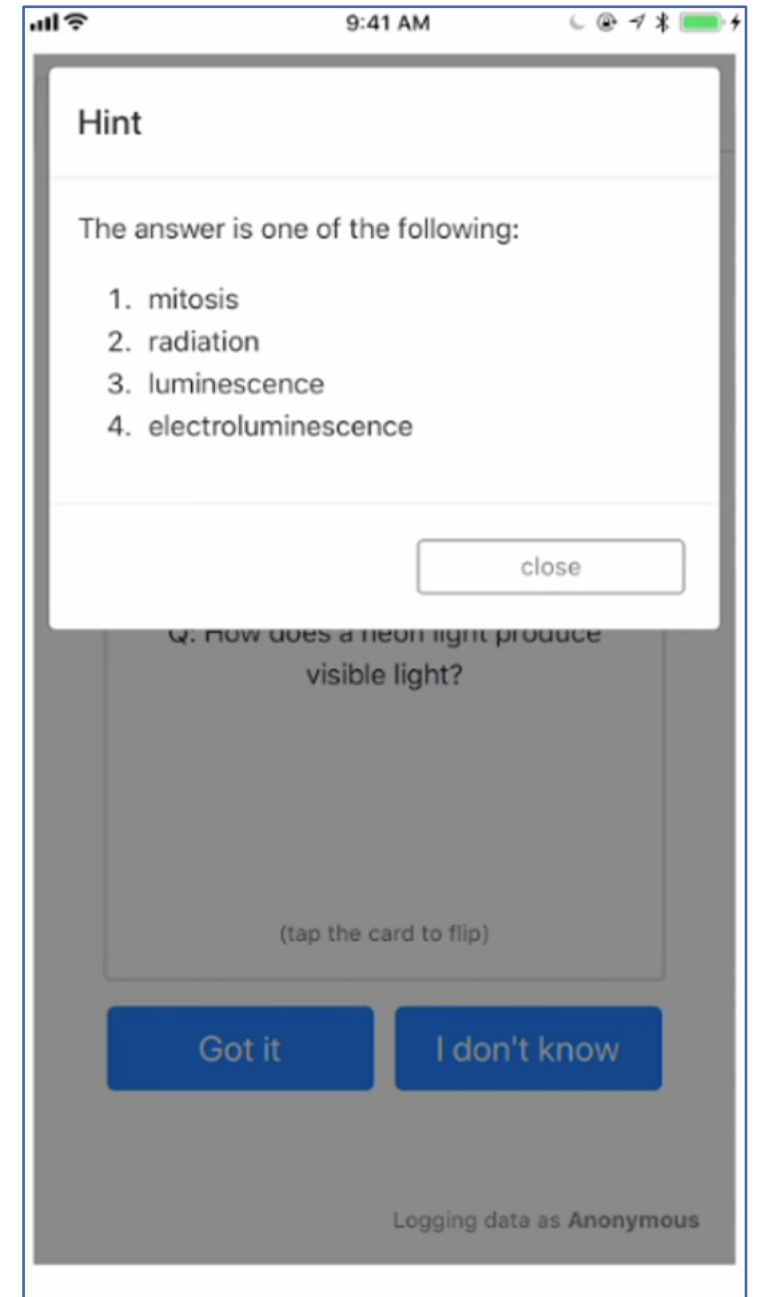
Safety

Evaluated QuizBot against a custom flashcard app



Key components of the flashcard app

- Tap-to-flip interface inspired by Quizlet
- Ability to toggle hints/explanations
- Same adaptive question sequencing algorithm
- Same factual question pool



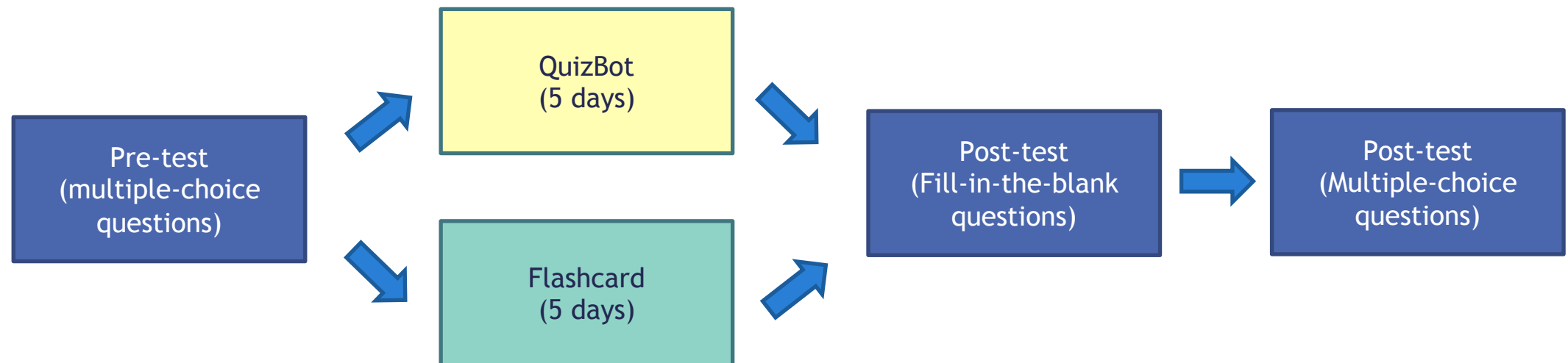
User Study

Research questions

- How engaging is QuizBot to learners in comparison to flashcards?
- How effective is QuizBot with helping learners per number of practice items compared with flashcards?
- Given voluntary usage, which system is more effective?

Study procedure

- The question pool of each app contained 48 questions, 16 for each subject
- Pre-test contained 96 multiple-choice questions from two question pools
- Users used both apps for 5 days under study specifications
- Post-tests only contained questions answered incorrectly in the pre-test. A fill-in-the-blank post-test was followed by a multiple-choice post-test



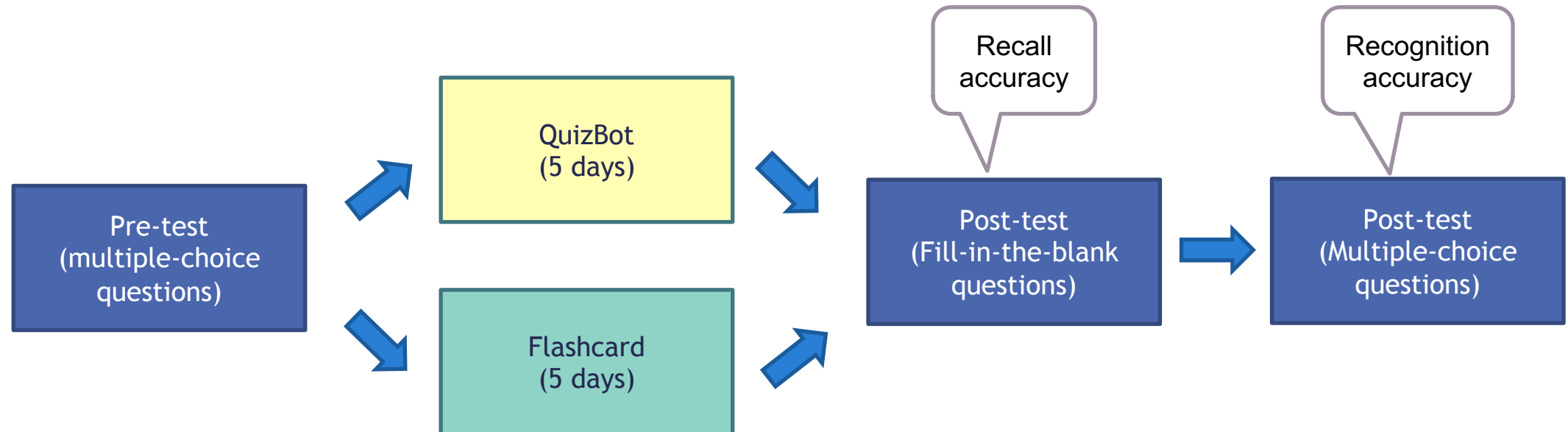
Study procedure

- Recall accuracy:

percentage of questions correctly answered on the fill-in-the-blank post-test

- Recognition accuracy:

percentage of questions correctly answered on the multiple choice post-test



Two within-subject studies

Study 1

Fix number of
questions practiced



Effectiveness

Study 2

Use both apps
voluntarily



Engagement

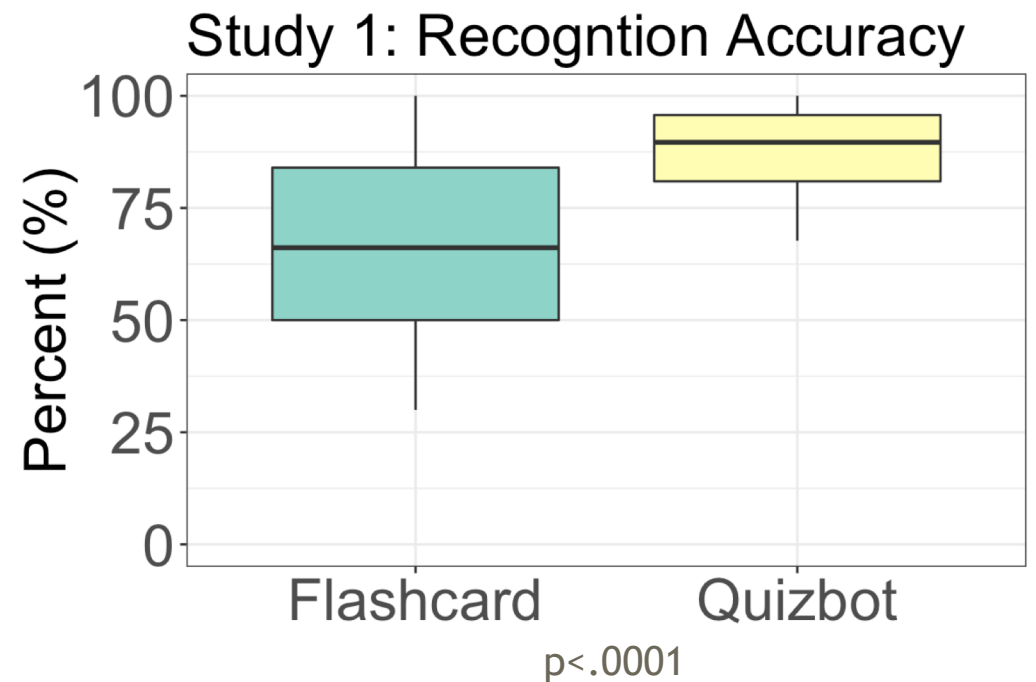
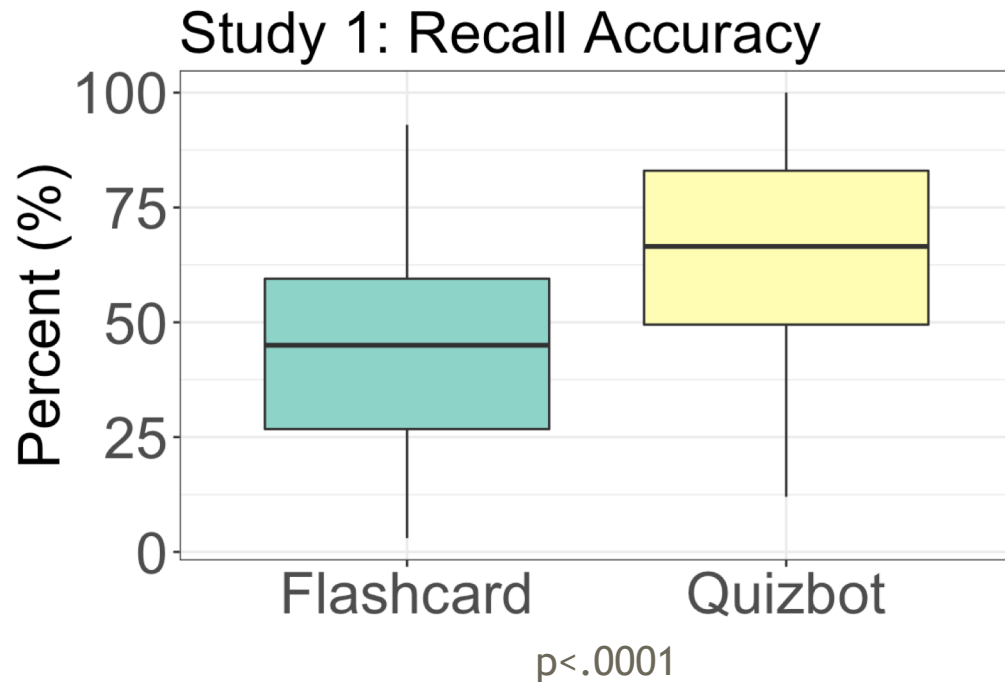
Study 1: fix number of questions practiced

- 40 university students and alumni from 11 different universities
- Users were required to practice 20 questions within each app in the first four days, and 16 questions in the fifth day
- Every question was practiced by every user exactly twice
- Each app sequenced 48 questions using the same algorithm

QuizBot more effective when usage was fixed

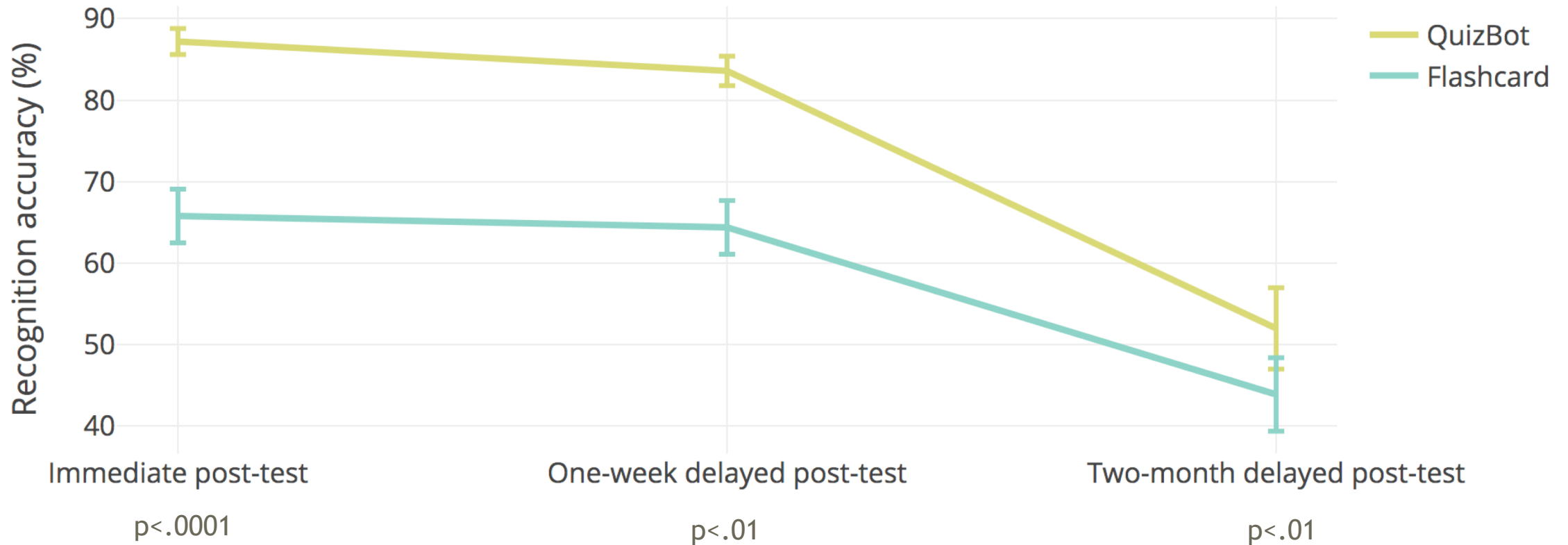
Given a fixed number of items practiced, with QuizBot people

- **recalled** over **20%** more correct answers
- **recognized** over **20%** more correct answers



QuizBot learning improvement persisted over time

Recognition accuracy of study 1 users over time



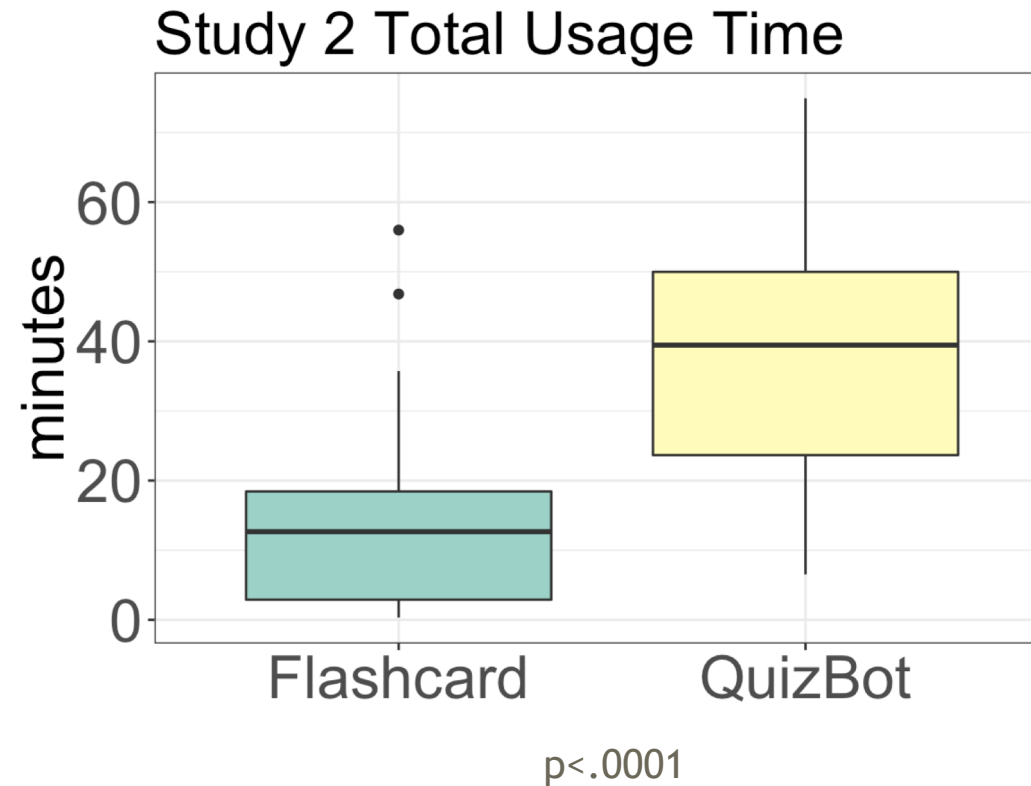
Error bars represent +/- 1 standard error

Study 2: use both apps voluntarily

- **36** university students and alumni from **8** different universities
- Users were asked to use both apps for 5 days **of their own volition**

QuizBot more engaging

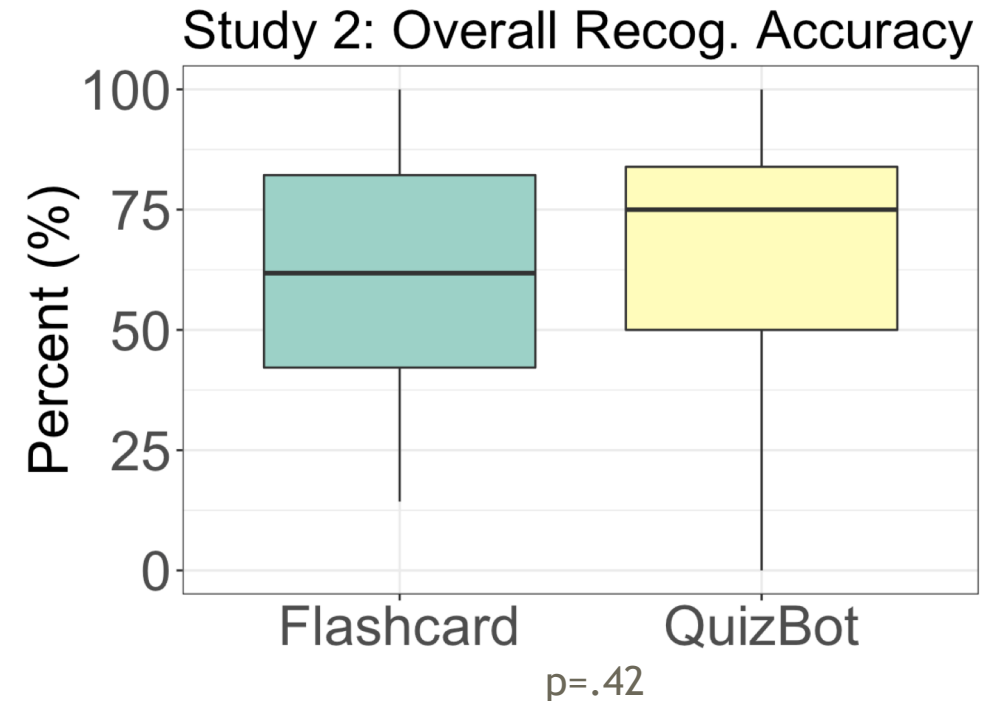
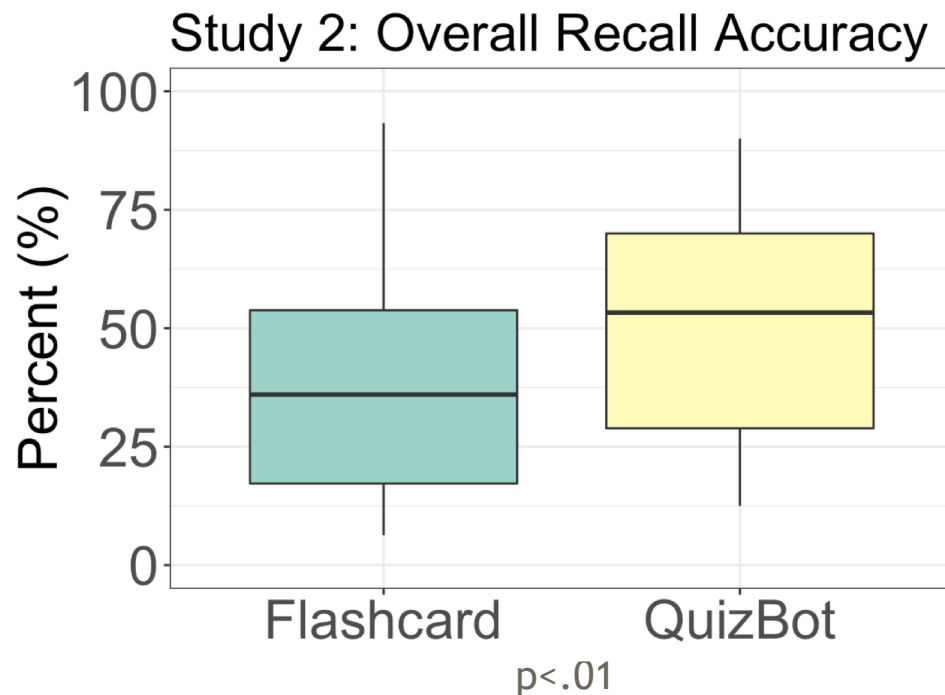
People spent **2.6x** more time with QuizBot than with the flashcard app of their own volition



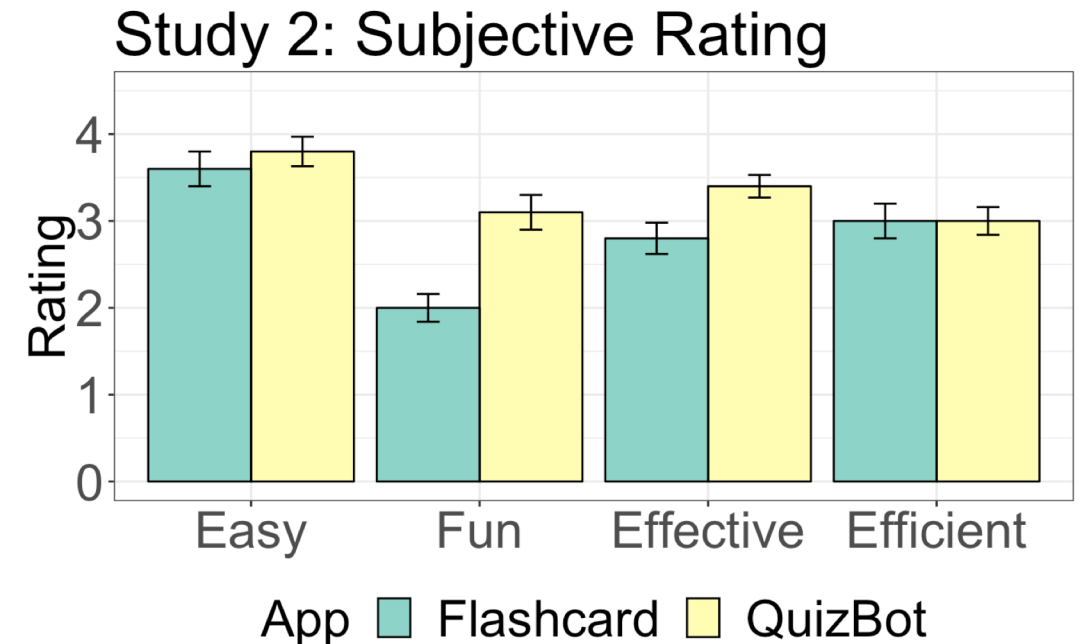
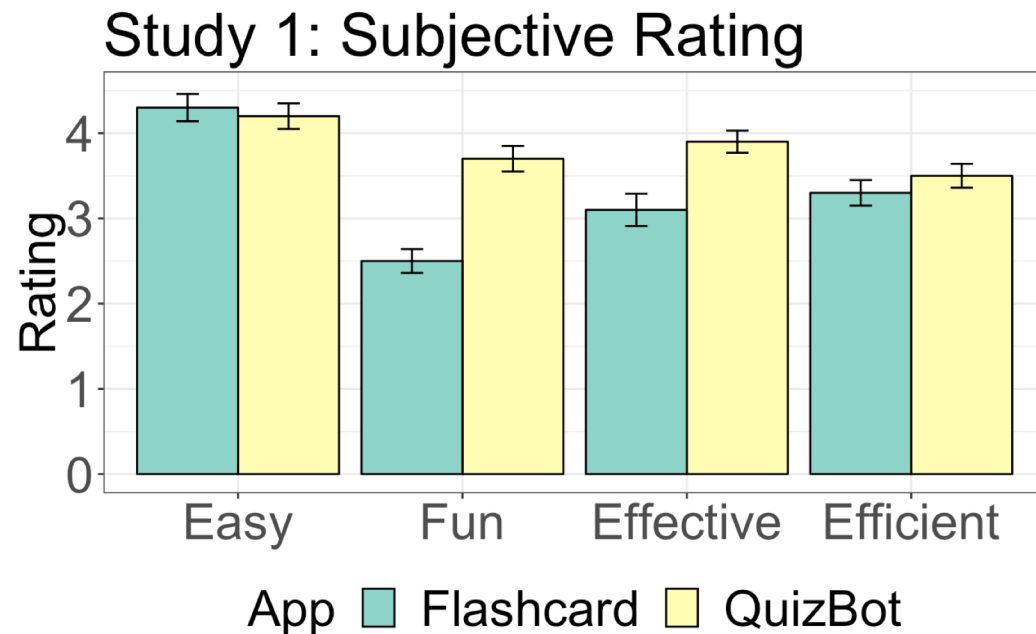
QuizBot more effective when usage was voluntary

Given voluntary usage time, with Quizbot people

- **recalled** over 12% more correct answers
- **recognized** a **similar** amount of correct answers

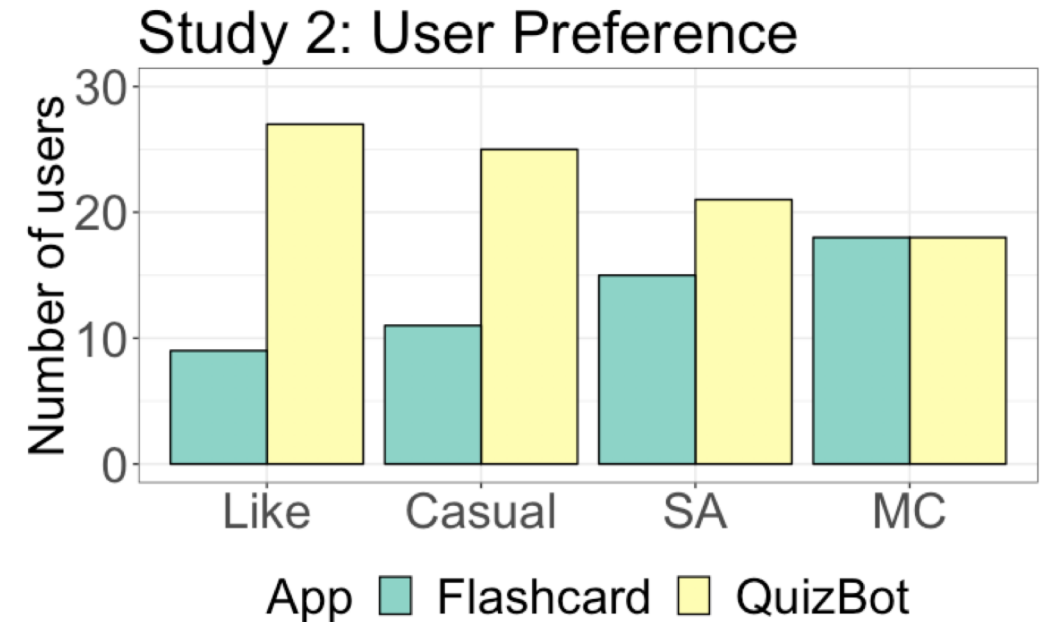
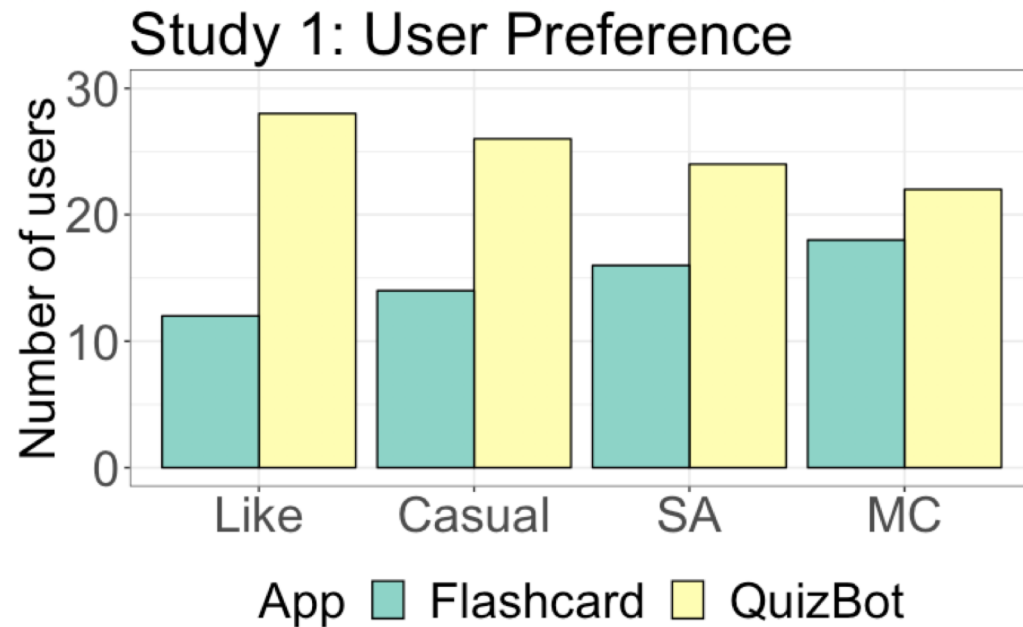


Participants self-reported QuizBot to be more engaging & more effective



Fun and effective results are statistically significant in both studies

Participants prefer QuizBot for casual learning



From left to right: which app users like better (Like) / prefer to use for casual learning (Casual) / for short-answer exams (SA) / for multiple-choice exams (MC)

Positive qualitative responses

- On communication and interactivity

“I could actually answer and type it out rather than just looking at flashcards.”
- On engagement and feedback

“QuizBot was more interactive and engaging. The feedback (positive or negative) was also motivating.”



Positive qualitative responses

- On conversations

“The penguin had a lot of personality, and I enjoyed trying to trick the chatbot.”

- On context of use

“It is something that I would definitely use if available and can easily be used during a bus ride etc.”



Suggestions for improvement

- On speed

*“I can go over questions and answers very fast.
Sometimes I have to wait for QuizBot to respond.”*



Directions for Future Work

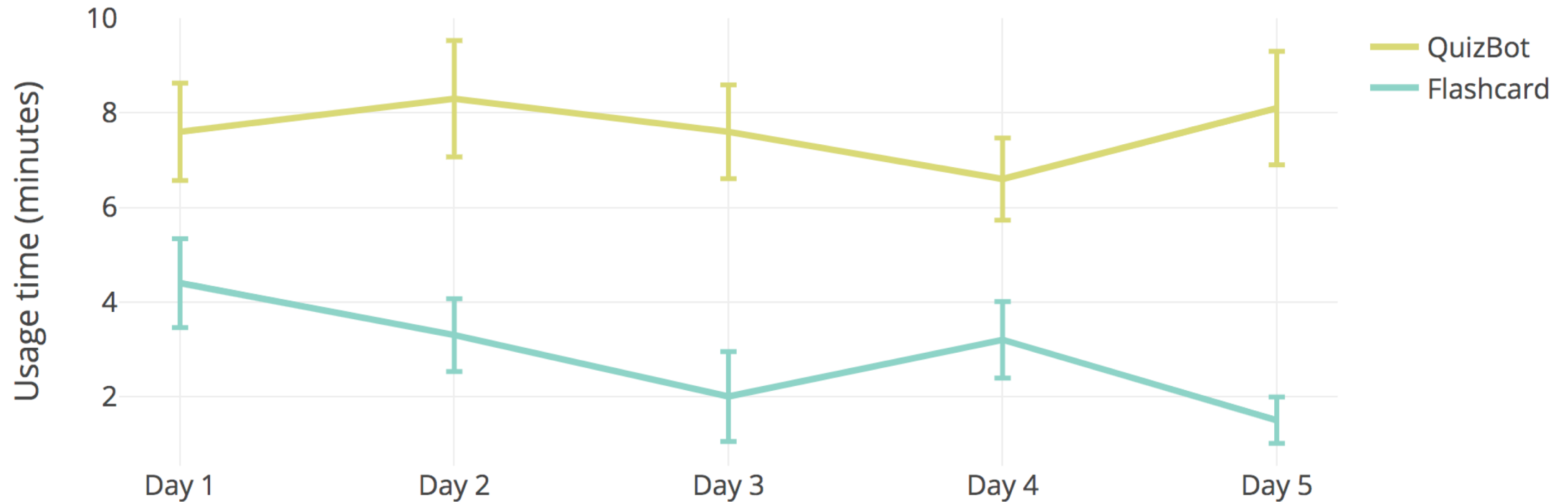
QuizBot is inefficient compared to flashcards

- Support speech in addition to keyboard based conversations (Ruan et al. 2018)
- Consider a flashcard system where learners have to type in the answer



Understand long term usage behaviors and attitudes

Average usage time of study 2 users from day 1 to day 5



Error bars represent +/- 1 standard error

Parse and understand user input better

- Randomly selected 11,000 (out of 43,956) consecutive conversational logs
 - contained 1052 questions and 144 questions were answered via typing
 - 139 questions were correctly graded by our algorithm (96.5% accuracy)
 - 1 was because the system could not handle a typo
 - 3 were because our algorithm penalized short answers and users tended to type terse answers
 - 1 was because the system was not perfect at understanding phrases
- Leverage richer natural-language input from learners to provide feedback of a higher granularity (Kolchinski et al. 2018)

Conclusion

- Designed and built the first chat-based learning system for factual knowledge memorization outside of classroom settings
- Evaluated its effectiveness and engagement through rigorous comparison studies with a traditional learning tool
- Offered design suggestions for building improved future educational chatbot systems

Thanks to

TAL Education Group for funding the project

Google for cloud credits

Anna Wang of Harker School for designing the graphics and the conversational visual style of QuizBot

Educational chatbots are beneficial, particularly for learning outside of traditional settings



 [@sherrysruan](https://twitter.com/sherrysruan) |  ssruan@stanford.edu |

 <https://hci.stanford.edu/research/smartprimer>

Sherry Ruan, Liwei Jiang, Justin Xu, Bryce Tham, Zhengneng Qiu, Yeshuang Zhu, Elizabeth Murnane, Prof. Emma Brunskill, Prof. James Landay

