

**TOWARDS IMPROVING PERSONAL ASSISTANTS AND
EDUCATIONAL SOFTWARE: HOW QUESTIONS ARE ANSWERED
AFFECTS LEARNING**

John Leddo¹, Tianle Chen², Aditya Menachery³, Jai Agarwal⁴ and Taruna Agarwal⁵

¹Director of Research at MyEdMaster

^{2,3,4,5}researchers at MyEdMaster

MyEdMaster, LLC
Herndon, Virginia, USA

DOI: 10.46609/IJSSER.2021.v06i02.019 URL: <https://doi.org/10.46609/IJSSER.2021.v06i02.019>

ABSTRACT

Using technologies like artificial intelligence (AI), machine learning, natural language processing and voice recognition, personal assistants and other query-answering systems have made enormous progress in understanding people's questions, retrieving relevant information and using that information to answer people's questions. What these assistants appear to neglect is providing users with different ways of answering their questions in order to improve understanding and task performance on the part of the user. The present study investigated whether such variability in responses to questions makes a difference in people's performance on a task that requires use of those answers. 17 middle and high school students were asked, as part of a course on infectious diseases, to develop a plan for creating a vaccine. They were given a list of 13 questions they could ask and receive answers to. Half of the students had only informational answers available to them and the other half could choose from informational answers, real-world analogies, goal-based or causal-factor answers. Results showed that students who were given the choice of answer type to their questions received scores on their vaccine development solutions that were twice as high as the students who received informational answers only. Results suggest that personal assistants, including those embedded in educational technology, could be made more effective by offering a variety of ways to answer questions rather than a single format.

Key Words: Personal assistants, educational software, question answering

Introduction

Personal assistants are quite pervasive these days, performing a wide range of tasks such as waking people up in the morning or answering questions. Some of the most common personal assistants are Siri (made by Apple), Alexa (made by Amazon), Google Assistant (made by Google), Cortana (made by Microsoft) and Bixby (made by Samsung).

Question answering is one of the most popular features used in personal assistants. There are several steps involved in answering questions. The first step involves the use of speech recognition and natural language processing (Copeley, 2020) to understand the question that is being asked. Given that there are multiple ways to ask a question as well as different speaking voices, artificial intelligence (AI) and machine learning (ML) are often used to understand to improve the comprehension of questions (Marr, 2018).

Once the question has been processed an answer is searched for and created (Mutchler, 2019). This can be done using different processes. The assistant can access a database of questions and answers and match the question to that and retrieve the corresponding answer. The assistant can access the Internet, look for websites such as Wikipedia that have information indexed to the question and retrieve text from that site. Google also has a concept map it uses as a source of information. Once the appropriate information is retrieved, text to speech technology converts the answer into speech that is presented to the user.

While personal assistants have come a long way in terms of understanding human speech, processing what has been said, retrieving relevant information to answer questions and then generating answers to those questions, which are then spoken back to the user, there seems to be a final step in the chain that has received little attention. The step we speak of is making sure the user understands the answer that is given. Given that information is useless until someone learns it, this is perhaps the most important step. Because the present project team is interested in educational applications of personal assistants, it is this final step that is of greatest interest to us.

Just as there are multiple ways a person can ask a question, there are multiple ways a question can be answered while still making the same general point. While the component technologies that make up personal assistants have evolved greatly toward understanding the different ways the same question can be asked, it seems that personal assistants have devoted little to no attention to addressing the best way to answer questions to make sure that the questioner understands the answer. This flies in the face of the fact that empirical evidence suggests that people do have preferred modes of receiving information (John et al., 2016).

This seems to be a non-trivial issue. Virtually everyone has been in the position of being misunderstood by those he or she has spoken to. Teachers learn that some ways that they explain concepts to students will confuse them, while other ways are understood. The present research seeks to extend this train of thought by exploring whether personalizing how questions are answered can improve the utility of personal assistants, particularly those embedded in educational systems. This is of special interest to us as we are developing artificial intelligence-based electronic textbooks, that among other features, contain a personal assistant (Leddo et al., 2019). In our electronic textbook technology, a student reading a lesson or watching a video can activate a personal assistant and ask questions about the subject matter. As with personal assistants cited above, our existing technology uses a combination of voice technology, natural language processing, AI and machine learning to understand and answer the questions being asked. However, students asking similar questions (those that resolve to the same basic question, like "What is a complex number?") receive identical answers.

We hypothesize that different ways of answering questions could lead to differences in how well people understand the answers and hence learn the desired material. The present study explores this hypothesis. However, before expending the resources to create AI and machine learning-based personal assistants to adapt how questions are answered to each individual learner/user, we first explore the more fundamental question as to whether different ways of answering questions even matters.

To address this question, an experiment was run in which participants were asked to perform the written task of describing how they would create a vaccine for the coronavirus. Since none of the participants already knew how to do this, they were given the opportunity to ask questions and receive answers to help them with their process. The question-and-answer process was delivered in a frequently asked questions (FAQ) format, delivered on a website. In one condition, participants received only an informational answer to their question. In another condition, participants could select either an informational answer, one expressed in terms of a real-world analogy, one expressed in terms of cause and effect and one in terms of the goal that needed to be achieved. We hypothesize that participants allowed to select the type of answer they want would produce better solutions for how to create a vaccine than those who receive only informational answers to their questions.

Method

Participants

The experimental participants were 17 middle and high school students enrolled in an

an online 10-week course in infectious diseases. The course used Covid-19 as the test case, and it was taught by a medical doctor. Of the 17 participants, nine were assigned to the single answer condition and eight were assigned to the multiple-answer condition.

Software Used

The software used in this experiment was a one-page website coded with JavaScript, HTML, and Node.js(a library in JavaScript). There were two versions of the website, one for each experimental condition. Each version of the website directed the participant to describe how he or she would create a vaccine for Covid 19. A large text-box was provided for the participant to enter his or her solution.

Beneath the textbox was a list of 13 questions that the student could click on to get answers that would be helpful in creating the vaccine. The questions were:

What is a vaccine?

What are vaccines made of?

What are the ingredients in a vaccine that are needed to make it safe and effective?

What are the stages of development of a vaccine?

Who is involved in vaccine production and marketing?

How are vaccines stored?

How many doses of vaccine do we need?

Can we use the oral route (by mouth) or injection route of vaccine administration?

Do vaccines have an expiration date?

Can people with egg allergies get vaccines?

What is the FDA?

How are vaccines distributed?

Who can administer a vaccine?

In the informational answer only condition, when a participant clicked on the question, the informational version of the answer appeared. In the multiple-answer version, the question appeared along with each of the four types of answer types. The participant then clicked on his or her preferred answer type. Once clicked, the answer text appeared, and the other answer types became de-activated so that the participant could see only one answer to any question. This was done to control for the amount of information a participant got for each question asked.

Using Node.js to create a server, we were able to save all the information to a text file. The information was added as soon as the student clicked the done button. The information saved

was what questions were clicked on, what answer types were chosen (in the multiple-answer type condition) and the participants' solutions to creating the vaccine.

Procedure

The 17 participants performed the task online in groups of 8 and 9. The class they were part of was an online 10-week course that covered infectious diseases using Covid-19 as the test case. The students were taught by a medical doctor. In one of the weekly sessions, the instructor explained to the students/participants that they will be taking an online test and briefly talked about how to access the link to the online program. They were told that they would be asked to write about how they would create a vaccine for Covid-19. They were not informed that they would be part of an experiment nor were they told that there would be two versions of the software and that other participants might be viewing a different version than what they were.

The topic of creating a vaccine was chosen as it was something participants would have no prior exposure to and had not previously been covered in the course they were taking. Therefore, it was reasonably expected that any information they would need to fulfill the task would have to be gleaned from the answers to the questions contained in the software. Moreover, because the task was both new and difficult, it was expected that there would be no ceiling effects on the quality of the participants' solutions, thus creating an opportunity to test for the effects of different answer types of performance. Participants were given up to 3 hours to complete their task, although none required the full amount of time, suggesting that none were rushed and were able to fully complete the task. Since participants worked remotely and online, the instructor neither interacted with the participants during the experiment nor was there any way for the instructor to observe individual participant efforts and thus know which solution was being created by which student or in which condition.

Results

The software output two files, one from each condition of the experiment. Each file contained two types of data: the solution that each participant created and the questions that each participant sought answers to from those available. In the multiple answer condition, the type of answer chosen was also listed next to its associated question.

The two files were merged into a separate file for scoring. The separate file contained only the solutions created. Any information that might identify the name of the authoring participant was removed (some participants put their names at the beginning or end of their solutions). No substantive changes were made to the solutions themselves. Moreover, the two files were combined in a way that made it impossible to discern which conditions the authors of the

solutions were in. However, the solutions were numbers 1 to 17, so that they could be matched with the conditions they came from and the types of answers the participants requested for those in the multiple answer condition.

The file was then given to the instructor who graded each individual response on a scale of 0 to 100 points. The rubric was created at the same time the task was created in order. The rubric was as follows:

DATA TO BE INCLUDED IN THE ANSWERS -

Introduction - 5 points

Vaccine ingredients +components - 10 points

Phases of vaccine testing - 20 points

Licensing procedure (obtaining a license from FDA) - 5 points

Vaccine storage - 10 points

Vaccine transport - 5 points

Vaccine expiration date and doses - 10 points

Mode of vaccination (injection, oral etc) - 5 points

Vaccine administration (camps?, physician offices etc) - 5 points

Vaccine distribution by the manufacturer - 10 points

Vaccine marketing -10 points

Conclusion - 5 points

Once the scoring was completed, the scores were reviewed to see if there were any outliers. One score from the single answer condition was deemed an outlier as it differed from the mean by more than two standard deviations and differed from its nearest score by approximately two standard deviations, whereas all other scores were within one standard deviation of the mean. Accordingly, that score was dropped from the dataset and all remaining analyses are based on a sample size of eight per condition.

The mean solution scores for the two conditions are 41.6 for the multiple answer condition and 21.5 for the single answer condition. In spite of the relatively small sample size, this mean difference of approximately 2:1 was statistically significant, $t = 2.07$, $df = 14$, $p = .029$, one-tailed. These results suggest that people who were allowed to select how they wanted their questions answered scored approximately twice as highly as those who received answers labelled as informative only.

In order to determine whether the observed difference was due to the fact that choices were offered and not some other variable, additional analyses were performed. One factor that needed to be ruled out was whether the availability of answer choices simply induced participants to ask more questions and therefore obtain more information than did participants who had no answer choices. In order to do this, the total number of questions asked by participants in both conditions was tallied and compared.

A total of 13 questions were available for participants to ask. All participants availed themselves of the opportunity to ask questions and the minimum number of questions asked by any participant in each condition was seven and the maximum was 13. The mean number of questions asked by participants in the single answer condition was 11.4 and the mean number of questions asked by participants in the multiple answer condition was 11.6. This difference is not statistically significant.

The next factor to look at was whether participants in the multiple answer condition even took advantage of the opportunity to have their questions answered in different ways. Of the eight participants in the multiple answer condition, five requested more than one type of answer to the different questions (only one answer per individual question could be selected). The other three requested informational answers only. The sample size here was too small to determine if there were any performance differences between those who requested more than one type of answer and those who used only the informational answers. However, this may not be important since the present hypothesis addresses the availability of different answer choices, not what answer choices lead to better performance for what people. It is still possible that informative answer choices work better for and will be chosen by some people.

A final question of interest was whether there were any systematic preferences for answer choice types among the people in the multiple answer choice condition. Because of the small sample size, we consider any trends observed to be exploratory and additional research should be performed. Still, some definite patterns emerged. Of all answer choices selected, approximately 68% were informative, 21% real-world analogy, 4% were causal, and 7% were outcome, suggesting that nearly a third of the requested answers were not the informative ones.

In order to shed further light on whether there was any pattern to these preferences, we looked at the types of answer choices chosen to see if they related to the types of questions being asked. This review provided some interesting observations that should be the subject for additional research. Of the 13 questions available to participants, some were clearly factual in nature, such as “What is a vaccine?” or “Who can administer a vaccine?” Others involved procedures, such as “What are the stages of development of a vaccine?” and “How are vaccines stored?” Some were more causal in nature, such as “What are the ingredients in a vaccine that are needed to make it safe and effective?”

When choosing answers to factual questions, participants overwhelmingly selected informative answers, which were chosen 85% of the time. The remaining 15% of the time, real-world analogies were chosen as answers, while causal principle and outcome answers were not chosen at all. On the other hand, for procedural and causal questions, all types of answers were chosen. For procedural questions, the percentages of answer choices chosen were: informational, 53%; real-world analogy, 33%; causal principle 2%; outcome 12%. For causal questions, the percentages of answer choices chosen were: informational, 69%; real-world analogy, 10.3%; causal principle, 10.3%, outcome, 10.3%. This suggests that the optimal way to answer a participant’s question to maximize learning involves both characteristics of the learner and characteristics of the question being asked.

Discussion

The results of the present experiment convincingly demonstrate that a “one-size-fits-all” approach to answering questions is not as educationally effective as allowing people to have different options for how to receive answers to their questions. People who were given the option to choose the types of answers to their questions scored, on average, twice as highly in their solutions as those who received only informational answers. The implication for this is that developers of general personal assistants or educational software should include multiple formats for presenting information to users. This principle could extend to frequently asked questions sections of websites, user guides, etc.

A secondary finding of the present study relates to the answer choice preferences that people exhibited. In all cases, the informational answer choice was the most preferred. However, this was not consistent across all question types. For factual questions, informational answers were chosen 85% of the time, making it the overwhelming favorite. This makes sense given that, when one wants to know a fact, a straightforward answer is sufficient. However, when procedural or causal questions are asked, other types of answers increase in their popularity.

For procedural questions, informational answer choices were preferred just over half the time, with real-world analogies coming in second, being preferred roughly one third of the time. For causal questions, informational answers were preferred nearly 70% of the time with the other answer types being preferred roughly equally.

Of course, the above percentages have to be considered exploratory in nature, given that they are based on very limited data. However, these percentages do suggest that not only do people prefer variety in the ways they want to have questions answered, but that these preferences may be linked to the type of question being asked. Additional research is needed to shed light on this relationship.

Additional research is needed, as well, to elucidate the relationship between how questions are answered and how well people learn based on the interaction of question and answer types. The present study demonstrates that giving people options for how questions are answered greatly improves (in this case, doubles) their performance on a task that requires use of the answers. However, it may not be the case that simply relying on people to be their own judges of what answer types will be most beneficial to them will lead to optimal learning/performance. Further research can shed light on what types of answers work best for what types of questions and for which people. Any number of variables such as question type, subject matter type, answer type and user characteristics (e.g., extent of prior knowledge of the subject) may be relevant. It is likely to be the case that optimizing performance involves an interaction of all of these variables.

Conclusion

The present research may have far reaching implications for how technologies and services that provide information to people should operate. Media that provide people with answers to Frequently Asked Questions could offer multiple or integrated answer formats to boost understanding. Personal assistants and educational software should incorporate data analytics and machine learning to optimize how answers are given to questions posed by users/students based on question type, content type and user characteristics. The expected results of this should be users who better understand and are better able to use the information they are provided.

References

- Copley, M. (2020, November 12). Understanding the Basic Architecture of Siri. Retrieved January 17, 2021, from <https://www.colocationamerica.com/blog/the-architecture-of-siri>
- John, A., Shahzadi, G. & Khan, K.I. (2016). Students' Preferred Learning Styles and Academic Performance. *Sci.Int.(Lahore)*,28(4),337-341.

Leddo, J., Guo, Y., Liang, Y., Joshi, R., Liang, I., Guo, W. and Bailey, S. (2019). Artificial Intelligence and Voice-powered Electronic Textbooks. *International Journal of Advanced Educational Research*, 4(6), 44-49.

Marr, B. (2018, October 05). Machine Learning In Practice: How Does Amazon's Alexa Really Work? Retrieved January 17, 2021, from <https://www.forbes.com/sites/bernardmarr/2018/10/05/how-does-amazons-alexa-really-work/?sh=128422e21937>

Mutchler, A. (2019, October 04). What are Virtual Assistants? Retrieved January 18, 2021, from <https://voicebot.ai/2019/10/05/what-are-virtual-assistants/>