

Which virtual personal assistant understands better?

Siri, Alexa, or Cortana?

Ahmed Alqurashi

Crescent View Academy

Which virtual personal assistant understands better? Siri, Alexa, or Cortana?

Abstract

The purpose of this experiment is to compare the abilities and understanding of virtual personal assistants (VPAs) and investigate which of them gave better understanding through three software; Alexa, Siri and Cortana. These virtual assistants help people and make their life easier by answering questions and performing some digital actions through voice queries. In this experiment, I asked each virtual personal assistant fifty-seven questions under seven categories. The results of this project will help users know that these virtual personal assistants are different software, and know which one of them is better. So, these results will help them decide which device they will prefer to buy since VPA is one of the main features of nowadays personal devices.

Background:

Nowadays people rely on technology a lot. Virtual personal assistants (VPAs) are one of the most used applications in technology every day. Virtual personal assistants like Siri, Cortana, and Alexa are online assistant's applications that help people and make their life easier. These virtual assistants are not like search engines, yet they will set reminders, send messages, tell directions, and make calls. These assistants are very useful, however, it is still hard for people to know how they understand us. These cognitive devices help people in many different aspects such as shopping, travel, and entertainment ... etc. (Sathi 29-39)

The purpose of my research will investigate which virtual personal assistants will give better understanding. My experiment will examine the three major Virtual personal assistants (Alexa, Siri, and Cortana) to see if there is a difference in their answers and which one gives more accurate answers in many different categories, that are : Health category, Marketing and shopping category, Entertainment and travel category, Timing and reminders category, Math category, Connected questions category, and Non classified category.

Virtual personal assistants are online assistants that we use almost every day. Alexa, Siri, and Cortana are all VPAs. Amazon made Alexa in 2014. Cortana was made much before by Microsoft in 2014. Siri is a little younger and was made in 2011 by Apple.

Apple's Siri

Siri is one of the applications created by Apple. Siri started as a virtual personal assistant application on the iPhone 4S, iPod touch 5, iPad touch 3, iPad Air and all of the iPad minis. It was introduced to the world in 2011. It experienced many developments. In iOS 6 Siri was able to provide information about sports, movies, and restaurants, and was able to open apps. In iOS 7 it was able to have better voices and languages. In iOS 8 it could listen to hey Siri (“In short”). Even though Siri uses information from Google, yahoo, and yelp, it is not considered a search engine. Siri is available on Apple devices and Apple TV. The price for these devices is between 200 and 600 dollars. Siri is good on the go because it can be in your pocket. It can give you locations. Some questions that you ask Siri will end up with Siri giving you a bunch of links (Gustavo 241–250). The Siri icon appears in the right side of figure 1.



Figure 1: virtual personal assistant applications

Amazon Alexa

Alexa is a VPA created by Amazon. It was named after the ancient library Alexandria. It was created in 2014. Alexa is a touch free device that has speaker. It can play music, read the news, and set alarms. Alexa is able to connect to a phone as well as show directions and estimate traffic times. Also, it can be connected to smart home system and control the lights by voice. Since 2017, over 5000 workers work on Alexa and its programs. The Alexa devices cost about 50-150 dollars (Gustavo 241–250) (“Amazon Alex”). Alexa’s icon is show in the middle of figure 1.

Microsoft Cortana

Cortana is another VPA that was created by Microsoft in 2014. Like Alexa and Siri, she can make phone calls, set directions, read the news, and send messages. Also, she can set timers and reminders. Cortana is available in many devices, such as 10 windows, 10 Mobile, Windows Phone 8.1, Microsoft Band, and the Xbox One. These devices cost around 200-1000 dollars. Cortana comes in many languages like English, Spanish, Portuguese, French, Japanese, and Chinese (Goksel-Canbek 592-601) (“Cortana is your truly personal digital assistant”). Cortana icon appears in the left side of figure 1.

There are many researchers that studied virtual personal assistants (Rammohan) (Gustavo 241–250) (Dunn). Some of them did not cover all the virtual personal assistants, and others covered a small number of categories. In contrast, this study will test three virtual personal assistants (Apple Siri, Amazons Alexa, and Windows Cortana) in seven different categories: Health category, Marketing and shopping category, Entertainment and travel category, Timing and reminders category, Math category, Connected questions category, and Non classified category. Also, when I reviewed the literature I found that no one used the statistical method since all of them describe the answer and mentioned the best answers. In this study, all the answers will be evaluated statistically and the mean for the answers for each device will be shown.

Rammohan conducted a study to compare Apple Siri, Google Now, Windows Cortana and asked them questions about health. They asked the three VPAs five questions about Allergy and Asthma Emergencies. They found out that none of them realized the questions were emergency questions. Google Now answered all the questions, however, Siri answered 2 questions and Cortana answered all the questions and gave search information answers (Rammohan).

Gustavo did a study (241–250) that compared the virtual personal assistants in four categories: shopping and buying assistant, care assistant, travel and entertainment assistant and administrative assistant. They found that Cortana was best in the shopping category. The worst in this category was Google assistant. In the travel and entertainment category, Siri and Cortana were the best. The worst were Alexa and Google assistants. In the administrative assistant

category, Siri and Cortana were best, and the worst was Google assistant. Finally, they mentioned that these systems need to be improved and developed (Gustavo 241-250).

Dunn asked the devices many questions then he mentioned which device was best in every question. He mentioned that Siri won when the question was “Do I have any new texts?”. However, Cortana won when he said “what’s happening near me?”. Then Alexa won when he asked “Play me some new music.” Google assistant won when he asked ”what’s the news today?”. Finally, he mentioned that there was no winner when he asked “who won the premier league?”

Research question:

This study was designed to answer one main research question, and under this question there were two sub questions:

- 1) Is there a difference between Siri, Cortana, and Alexa in their understanding?
 - a. Which one of the three virtual personal assistants understands better and gives the most accurate answers in all categories?
 - b. Which one of the three virtual personal assistants understands better and gives the most accurate answer in the:
 1. *Health category.*
 2. *Marketing and shopping category.*
 3. *Entertainment and travel category.*
 4. *Timing and reminders category.*
 5. *Math category.*
 6. *Connected questions category.*
 7. *Non classified category.*

Method:

The purpose of this experiment is to compare the abilities and understanding of virtual personal assistants (VPAs) and investigate which of them gave better understanding through three software; Alexa, Siri and Cortana. These virtual assistants help people and make their life easier by answering questions and performing some digital actions through voice queries. In this experiment, I asked each virtual personal assistant fifty-seven questions under seven categories. The results of this project will help users know that these virtual personal assistants are different

software, and know which one of them is better. So, these results will help them decide which device they will prefer to buy since VPA is one of the main features of nowadays personal devices.

Variables:

1. **Dependent variables:** *Independent variables:* Alexa vs. Siri vs. Cortana
2. **Dependent variables:** The scores for their responses.
3. **Control variables:** The experiment conducting time was important. I controlled this variable by repeat the experiment in different times during the day. For example, when Siri was asked for a cheap pizza place open now and it was 6:30 AM, she didn't give an answer. Thus, we asked her the same question again at 4:00 PM and she answered it perfectly.

Hypothesis:

There was a difference between Alexa, Siri, and Cortana in their understanding.

Materials:

- [1] Apple iPhone 7 Plus (for Siri)
- [2] Amazon Dot Speaker (for Alexa)
- [3] Windows Acer 10 (for Cortana)



Figure 2: Devices used in the experiment

Questionnaire:

There were 57 questions developed for this study. All 57 questions were assorted under 7 categories. The first category was “the timing and reminders category”. There were 8 questions in this category. These questions are asking about time and setting reminder and creating calendar meeting. The second category was “the health category”. There are six questions in this category. These questions are testing what the virtual personal assistants would do with health issues such as: I’m not feeling good; sick; I have asthma. The third category was “the connected questions category”. There were 3 questions in this category. These questions are asking about the second question of a connected question for example what time is it in San Francisco? “What is the weather there?” The next category is “the marketing and shopping category”. It has 8 questions. This category is asking about prices and shopping information. The fifth category is “the travel and entertainment category”. It has 9 questions. It asks about ticket prices and movie times and directions. The sixth category is “the math category”. It has 8 questions. It asks about math solutions and calculations. The last category is “the non classified category”. It has 15 questions. It has random topics like jokes and translations.

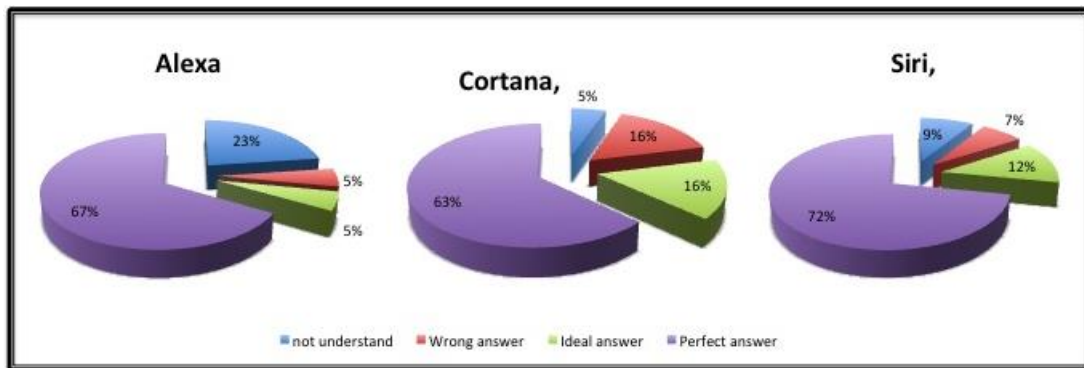
Procedure:

1. Each one of the VPAs was asked 57 questions from 7 categories.
2. The answers were recorded. Each one was evaluated from 0 to 3:
 - If they did not understand, the answer was evaluated by (0).
 - If the VPAs understood the question but gave a wrong answer, the answer was evaluated by (1).
 - If the VPAs understood the question and gave an ideal answer, the answer was evaluated by (2).
 - If the VPAs understand the question and gave a perfect answer, the answer was evaluated by (3).
1. Two methods were used to collect the data. The first method was the quantitative method by evaluating each question. The second method was the qualitative method by interviewing the VPAs and recording the answers and observed their response.

3. To analyze these data, the scores were calculated for each category and for all categories. Also, the means were calculated for each category and for all categories. The means were used to answer all the research questions.

Data and summary:

The first table shows the repeated responses of the three VPAs. As we can see, Siri answered not understood answers 5 times and 4 times gave wrong answers and 7 times gave ideal answers and answered perfectly 41 times. Then Cortana with 3 miss understood answers and 9 wrong answers as well as 9 ideal answers and 36 perfect answers. Alexa answered 13 times with misunderstood answers and 3 wrong answers as well as 3 ideal answers and 38 perfect answers.



As we can see in the table below, all the VPAs have their highest scores in their perfect answers. Siri’s lowest was the wrong answer. Cortana’s lowest repeated response was misunderstanding. Alexa’s lowest were the wrong answer and the ideal answer.

For the not understand the lowest one was Cortana, and the highest one was Alexa. For the wrong answer the lowest one was Alexa, while the highest one was Cortana. For the Ideal answer the lowest was Alexa and the highest was Cortana. For the perfect answer the lowest was Cortana and the highest was Siri.

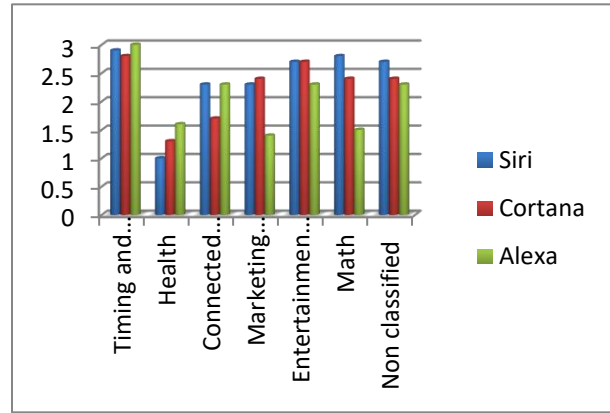
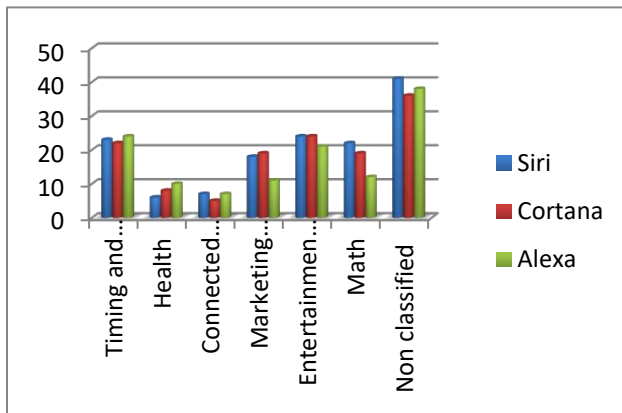
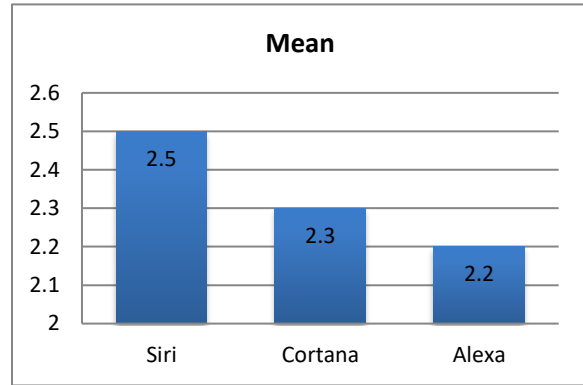
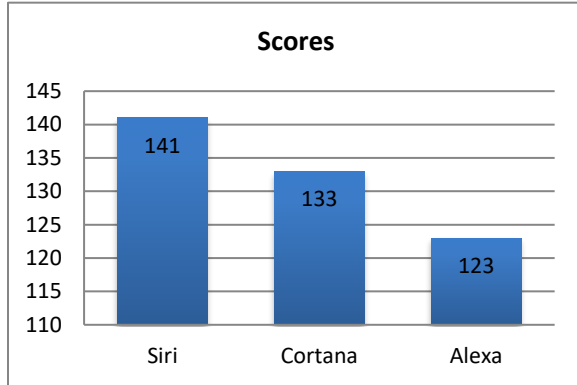
Type of answer	Siri	Cortana	Alexa
Not understand	5	3	13
Wrong answer	4	9	3
Ideal answer	7	9	3
Perfect answer	41	36	38

As we can see in the second table below the result showed that Siri's software had the highest score between all the VPAs in general. Siri got 141 points that considered 83% of all the 171 total points. Siri had the highest mean (M=2.5) over all the VPAs and Alexa had the lowest score and mean (S=126, M=2.2). In contrast, Cortana had the median score and mean (S=133, M=2.3). Separately, All three VPAs did great understanding the first category, which is the Timing and reminders category and all three VPAs performances was more than 92%. In contrast, Siri and Cortana did the lowest performances in the Health category and the same thing with Alexa in the Marketing and shopping category.

Category	Siri		Cortana		Alexa	
	Score	Mean	Score	Mean	Score	Mean
Timing and reminders category	23	2.9	22	2.8	24	3
Health category	6	1	8	1.3	10	1.6
Connected questions category	7	2.3	5	1.7	7	2.3
Marketing and shopping category	18	2.3	19	2.4	11	1.4
Entertainment and travel category	24	2.7	24	2.7	21	2.3
Math category	22	2.8	19	2.4	12	1.5
Non classified category	41	2.7	36	2.4	38	2.3
Total	141	2.5	133	2.3	123	2.2

In details for the each category separate, the results showed that Alexa in the Timing and reminders category recorded the highest by answering all the questions for this category perfectly, So Alexa had a mean of 3 in this category. Second it was Siri with a mean of 2.9, then Cortana with a mean of 2.8. In the second category, which is the Health category, Siri had the lowest mean, which was 1. Then Cortana with a mean of 1.3 and then Alexa with a mean of 1.6. In the third category, which is the connected question category, Siri and Alexa have better means (M=2.4) than Cortana's mean (M=1.7). In the fourth category, which is the marketing and shopping category, Cortana was best with a mean of 2.4. Siri was just behind with a mean of 2.3. And last is Alexa with a mean of 1.4. In the fifth category Siri and Cortana are tied at first place with a mean of 2.7. Then it's Alexa with a mean of 2.3. In the math category Siri was best with a

mean of 2.8. Followed by Cortana that has a mean of 2.4. Then finally Alexa with a mean of 1.5. In the last category Siri was best with a mean of 2.7. Then Cortana with a mean of 2.4. And last is Alexa with a mean of 2.3.



Conclusion:

This study was designed to answer one research question and two sub questions which where:

- 1) Is there a difference between Siri, Cortana, and Alexa in their answers?

The answer of this research question is yes. There are many reasons for that. First, the results showed that there were different between their means. Second, from my observation I saw a difference in their answers. So, my hypothesis was correct.

- a) Which one of the three virtual personal assistants understands better and gives the most accurate answer in all categories?

The answer of this research question is Siri. The results showed that Siri has the highest mean, and she gave the least misunderstood answers.

b) Which one of the three virtual personal assistants understands better and gives the most accurate answer in each category?

To answer this research question, the highest mean score was used to decide which one of the three virtual personal assistants understands better and gave the most accurate answer in each category. Alexa gave a better understanding in the Health category and Timing and reminders category and Cortana gave better answers in the Marketing and shopping category. Siri gave better answers in the Math category and the Non-classified category. Jointly, for the Entertainment and travel category, Siri with Cortana gave better answers and Siri and Alexa gave better answers in the connected questions category.

From the results I saw that all of the three virtual personal assistants were great in the timing and reminding category, however, there were very bad in the health category. In my opinion, the virtual personal assistants should be improved in the health area because it is very important. What caused Siri to win was the timing and reminders category and the non-classified category and the math category. The results show what made Alexa lose was the marketing and shopping category because she is connected to Amazon and most of the time she orders from Amazon. She answered a lot of wrong questions from the math category. I think it's a big problem because math skills are important features of VPAs.

In conclusion, studying VPAs is very important, and this topic needs to get discussed more and more. My results will help the reader understanding VPAs, which will help them make the correct choice between them, and researchers to continue working on this topic in future studies. I think there could be many future studies that could be done on this topic. In the future, I would add more VPAs to my study such as Google Home and Bixby. Also, I would time the speed of the VPAs' responses to see how fast they answer me.

Work cited

- Sathi, Arvind. "Cognitive Devices as Human Assistants." *Cognitive (Internet of) Things*, Palgrave Macmillan, 2016, pp. 29-39.
- Goksel-Canbek, Nil, and Mutlu, Mehmet. "On the track of Artificial Intelligence: Learning with intelligent personal assistants." *Journal of Human Sciences*, vol. 13, no. 1, 2016, pp. 592-601.
- Rammohan, Rajmohan, et al. "Smartphone Conversational Agents (Apple Siri, Google, Windows Cortana) and Questions about Allergy and Asthma Emergencies." *Journal of Allergy and Clinical Immunology*, vol. 139, no. 2, 2017.
- Gustavo, López , Luis Quesada, and Luis A. Guerrero. "Alexa vs. Siri vs. Cortana vs. Google Assistant: A Comparison of Speech-Based Natural User Interfaces." *International Conference on Applied Human Factors and Ergonomics*, Springer, Cham, 2017, pp. 241–250.
- “Amazon Alexa.” *Amazon developer*. www.developer.amazon.com/alexa.
- “Cortana is your truly personal digital assistant.” *Microsoft*. 2017. www.microsoft.com/en-us/windows/cortana.
- “IN SHORT.” *iMore*. Nov 2017. www.imore.com/siri.
- Dunn, Jeff. We put Siri, Alexa, Google Assistant, and Cortana through a marathon of tests to see who's winning the virtual assistant race - here's what we found. Nov 2016. *Business insider*. <http://www.businessinsider.com/siri-vs-google-assistant-cortana-alexa-2016-11/#the-setup-theres-no-perfect-way-to-evaluate-a-talking-ai-database-let-alone-four-of-them-but-i-tried-to-cover-as-many-fundamental-topics-as-i-could-1>