Alexa, Can You Help Us Solve This Problem? How Conversations With Smart Personal Assistant Tutors Increase Task Group Outcomes

Rainer Winkler
Institute of Information Management
University of St. Gallen
St. Gallen, Switzerland
rainer.winkler@unisg.ch

Matthias Söllner
Information Systems and Systems Engineering
University of Kassel
Kassel, Germany
soellner@uni-kassel.de &
Institute of Information Management
University of St. Gallen
St. Gallen, Switzerland
matthias.soellner@unisg.ch

Maya Lisa Neuweiler
Institute of Information Management
University of St. Gallen
St. Gallen, Switzerland
maya.neuweiler@student.unsig.ch

Flavia Conti Rossini
Institute of Information Management
University of St. Gallen
St. Gallen, Switzerland
flavia.contirossini@student.unisg.ch

Jan Marco Leimeister
Institute of Information Management
University of St. Gallen
St. Gallen, Switzerland
janmarco.leimeister@unisg.ch

KEYWORDS

Smart Personal Assistant; computer-supported collaboration; Amazon Alexa; experiment

Table 1: Overview of Hypotheses

Hypothesis	Description
Hypothesis 1	H1. Groups interacting with SPA tutors achieve better task outcomes compared to groups interacting with human tutors within a problem-solving context.
Hypothesis 2	H2. Groups interacting with SPA tutors achieve higher levels of collaboration quality compared to groups interacting with human tutors within a problem-solving context.
Hypothesis 3	H3. Collaboration quality mediates the positive relationship between type of tutor and task outcome.

ABSTRACT

Despite a growing body of research about the design and use of Smart Personal Assistants, existing work has mainly focused on their use as task support for individual users in rather simple problem scenarios. Less is known about their ability to improve collaboration among multiple users in more complex problem settings. In our study, we directly compare 21 groups who either use a Smart Personal Assistant tutor or a human tutor when solving a problem task. The results indicate that groups interacting with Smart Personal Assistant tutors show significantly higher task outcomes and higher degrees of collaboration quality compared to groups interacting with human tutors. The results are used to suggest areas for future research in the field of computer-supported collaboration.

1. INTRODUCTION

Intelligent systems have become ubiquitous in modern life and are increasingly helping us perform everyday tasks in ways that they could not previously. Specifically, the popularity of Smart Personal Assistants (SPAs), such as Amazon's Alexa, Google's Assistant, Apple's Siri and other systems, has been steadily growing over the past few years [3]. SPAs are computer programs that attempt to simulate conversations of human beings via voice and text in- and output in order to help users perform their tasks [13]. People are using SPAs on standalone devices (e.g. Google's Home or Amazon's Echo), on their phones (e.g. Apple's Siri), on personal computers (e.g. Microsoft's Cortana), and on internet of things devices (e.g. Samsung's Smart TV) to complete tasks such as setting alarm clocks, checking news and weather forecasts, turning on lights, making reservations in restaurants, etc. With rapid advancements in artificial intelligence, such as speech recognition and natural language processing, SPAs have the potential to go beyond simple use cases by helping multiple users in more complex problem scenarios (e.g. collaboratively cooking a complex meal or learning a new language) [18]. These developments turn SPAs from simple question-answer tools to powerful tutors [15]. Until now, less is known about the use of SPAs in these settings [15]. Thus, in our study, we analyze *whether SPA tutors are able to increase task group outcomes and collaboration quality compared to human tutors*. Therefore, we conducted a laboratory experiment with state-of-the-art SPA support. Groups of three participants had to conduct a 30-minute problem task receiving help from either a SPA or a human tutor. Building on the extensive work in human-SPA interaction, our work has implications for the CHI community since it shows how SPAs can be successfully used beyond traditional use cases in order to improve collaboration among group members and thereby being a worthwhile alternative to cost-intensive human tutors. Further research is needed to confirm these results and expand our work to similar scenarios.

Table 2: Group Task

Task
Increasing traffic flow has led to a significant increase in the number of traffic congestions, accidents and road casualties in Zürich. The city of Zürich asks you to build a focus group and find the best solution for them. What should the city do about it?

Figure 1: Example Dialogues with Amazon's Alexa Echo Dot (above) or a human tutor (below).

2. BACKGROUND

Smart Personal Assistants (SPAs) are computer programs that attempt to simulate conversations of human beings via voice and text in- and output trying to support users in performing their tasks, such as Amazon's Alexa, Google's Assistant or Apple's Siri [13]. Knote et al. [8] structure SPAs into five archetypes: adaptive voice (vision) assistants, chatbot assistants, embodied virtual assistants, passive pervasive assistants, and natural language assistants. Natural language assistants encompass more sophisticated speech recognition and spoken language understanding capabilities than any other type of SPA. Recent studies [10] indicate that these kind of SPAs can also be used as tutors assisting multiple users in solving more complex tasks. For example, Tegos and Demetriadis [17] demonstrated that SPAs are able to positively influence groups' dialogue behavior leading to substantially improved individual and task group outcomes. Moreover, studies in the field of computer tutoring proved that computer tutors can achieve similar learning gains as human tutors [9]. In some specific and narrow contexts, SPAs are even able to outperform human tutors [9] (see hypothesis 1, Table 1). Effective collaborative problem-solving (CPS) does not happen naturally. CPS is a complex process that can be defined as "... a joint activity where two or more people work together to progress through a series of cognitive states to solve a problem" [5]. CPS incorporates the components of cognition found in individual problem-solving (cognitive dimension) in addition to the components of collaboration (social dimension) [2, 12]. The cognitive dimension consists of taking appropriate actions to solve the problem and the social dimension consists of maintaining a shared understanding across the group members [5, 12]. Computers have often tried to trigger collaboration within problem-solving activities [6]. Harsley et al. [4] found out that students were able to significantly increase their programming skills and perceived SPAs as helpful within a collaborative problem-solving scenario. Similarly, Mcalister [11] demonstrated that group members interacting with SPAs show higher argumentation skills leading to better task outcomes. Moreover, Tarouco et al. [16] logged and analyzed student-SPA dialogues within a collaborative learning scenario and figured out that students tend to perceive SPAs as their collaboration partners (see hypotheses 2 and 3, Table 1).

3. METHOD

To test our hypotheses, we conducted a laboratory experiment where groups of three solved a problem task while receiving help from either SPA tutors or human tutors. Treatment Group 1 (TG1) was able to communicate with the SPA tutor via an Alexa Echo Dot device, while TG2 was able to communicate with the human tutor located in the same room (see Figure 1 for exemplary experimental setups). The human tutor can be considered as an expert (a teacher with many years of tutoring experience). The groups were not allowed to use the internet or any other aid.

We were able to recruit a total of 63 participants (T1=11 groups, T2=10 groups, mainly undergraduate students, m=36, w=27, age=21.29), who were randomly assigned to one of the subgroups. Participants received a baseline monetary reward with an optional additional reward depending on their task outcome.

Table 3: Exemplary Dialogue of Students Interacting with Tutor and Design Features

Dialogue	DF
Student: Alexa, open assistant **Alexa:** Hi there! I am Alexa, your assistant. I will help you to conduct the problem task on the piece of paper in front of you. You can skip to the next step by simply saying "next". If questions arise, say "help" and I can help you with two kinds of questions. First, I can help you with content-related questions. Second, I can also rephrase the problem-solving step by saying "repeat". Alright, let's start with the first step. Please try to identify the core of the problem within one sentence and note it on the piece of paper in front of you.	**Problem-solving step:** Problem identification
Student: Finished. Next **Alexa:** Ok, great. Now try to identify 3-4 causes of your identified main problem.	**Function:** Next
Student: Ok. Help. What can be a possible cause for this problem? **Alexa:** One cause can be the increasing distraction by using smartphones. What can be other causes?	**Function:** Help
(continuing interaction)	
Student: Can you repeat the problem step? **Alexa:** In this step, you should try to think of solution alternatives that will help your city to fight the causes. You should therefore not just think of any solution but rather mapping them to your identified causes!....	**Function:** Repeat **Problem-solving step:** Problem exploration
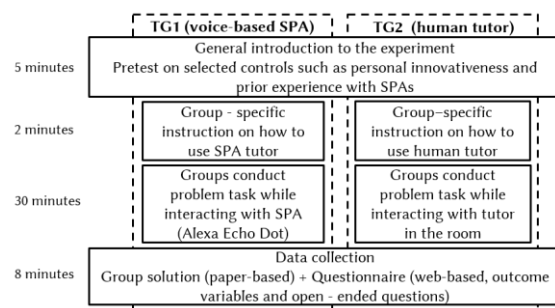
Figure 2: Experimental Procedure

We adapted the task from Gallagher et al. (Gallagher et al. 1992) that best meets the requirements of a problem task (ill-structured, open-ended, realistic) proposed by Jonassen (Jonassen 2000) (see Table 2).

We developed a voice-based Smart Personal Assistant with the help of Amazon's Alexa Skill Development Kit 2.0 with nodeJS. The interaction model consisted of two different logics (proactive and reactive). The SPA tutor proactively guided participants through the task by providing them five different problem-solving steps adapted from Kim and Hannafin [7].The participants were also able to ask questions (reactive) (see exemplary dialogues in Table 3 and Figure 1).

The procedure is divided into four components. Figure 2 outlines the procedure including activities done and the time spent in each component. For measuring task outcome, two experienced raters evaluated the group results using an evaluation framework aligned to Kim and Hannafin [7]. The framework helped to evaluate how well groups applied each of the five problem-solving steps on a scale from 1 to 7 which then resulted in the average task outcome. The task outcomes were rated individually, blinded, and independently from each other. Furthermore, we checked for interrater agreement with the help of a Pearson correlation (interrater agreement = 0.9699, $p < 0.05$). For measuring collaboration quality, we analyzed the cognitive and social component of collaborative problem-solving: For the cognitive component, we analyzed all 21 video recordings regarding (1) the frequency of applied turns and (2) the homogeneous content-specific contribution [19]. For the social component, we analyzed the video recordings regarding (3) the number of on-task asked questions and (4) the number of used repairs [14] (see Figure 3). Table 4 provides further information on the four measurements. We applied a mediation model that follows Baron and Kenny's [1] causal step approach investigating collaboration quality as mediator for the relationship between the type of tutor and task outcome.

4. RESULTS AND DISCUSSION

Table 5 shows the descriptive statistics for collaboration quality and task outcome for both treatment groups (TG1= SPA tutor, TG2= human tutor). Groups interacting with a SPA tutor showed both a higher collaboration quality and task outcome. Figure 4 depicts the model and summarizes the relationships between the dependent variable (task outcome), independent variable (type of tutor), and mediator (collaboration quality). Human tutor groups were on average 0.62 points lower in terms of task outcome than SPA tutor groups (c= 0.62, p= 0.0436) (path c). Therefore, H1 can be supported. Groups interacting with SPA tutors record a significantly better collaboration quality than groups interacting with human tutors (a= 1.09, p=0.0348) (path a). Therefore, H2 can be supported. Last but not least, we regressed the dependent variable (task outcome) on both the independent (type of tutor) and mediator variable (collaboration quality) to examine if the quality of collaboration
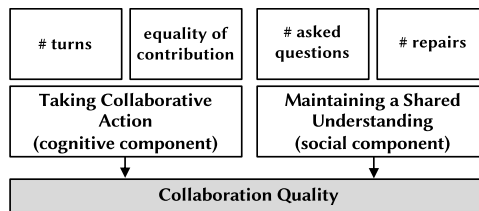
Figure 3: Components and Measurements affecting Collaboration Quality

Table 4: Description of Measurements

Measurement	Description
# turns	Number of turns within the collaboration.
equality of Contribution	Homogeneity of contribution per participant
# asked questions	Count of answered questions within the whole task
# repairs	Count of successfully detected and repaired misunderstandings within the whole task.

Table 5: Descriptive Statistics

	Collaboration Quality (M)		Task Outcome (Y)	
	$\bar{M}$	SD	$\bar{Y}$	SD
SPA Tutor (n = 11)	5.25	1.06	3.32	0.74
Human Tutor (n = 10)	4.30	1.29	2.71	0.54
All groups combined	4.80	1.24	3.03	0.71

Collaboration quality was ranked on a scale of 1-10, task outcome on a scale of 1-7
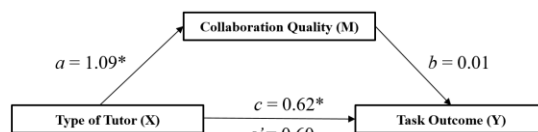

Figure 4: Path diagram with path coefficients.
*p<0.05

weakens the relationship between the type of tutor and task outcome (path *b* and *c*). The regression output affirmed that the direct effect between type of tutor and task outcome loses its significance (c = 0.62, p = 0.0436, c' = 0.60, p = 0.0824). According to Baron and Kenny [1], this suggests that collaboration quality mediates the relationship between type of tutor and task outcome. However, this mediation effect is not significant. H3 can therefore not be supported. These findings are in favor of most past studies indicating that technology is able to support collaboration [6].

Specifically, we noticed that groups interacting with SPA tutors put greater effort in the achievement and maintenance of a shared understanding of the problem at hand and had a more homogeneous contribution towards solving the task. In addition, groups interacting with a SPA tutor took turns (i.e. attempts to gain conversational floor) more frequently and their conversational flow was smoother as they were interrupting each other less. This confirms the results of Tegos and Demetriadis [17], who found that SPA interventions effectively scaffolded participants' discussions. One reason for this effect might be that participants feel more confident and comfortable when interacting with SPA tutors compared to human tutors since SPAs are not judging them nor building up any pressure. This can also be seen in a qualitative measurement at the end of the experiment. Some participants (n=8) mentioned that they prefer SPAs to human tutors because they felt much more in control of the situation. For example, one student commented "I like Alexa. Compared with a human tutor, I felt no pressure on successfully conducting the task and can decide for my own when I need help."

5. LIMITATIONS, FUTURE RESEARCH AND IMPLICATIONS

Our study has some limitations. The primary limitation of our study is the small sample size of 21 groups resulting in a rather small mediating effect. Despite the small sample size, the tests were able to show that there are significant differences between SPA and human tutors. It would be interesting to see if further research can confirm these results. Moreover, we investigated the effect of SPAs in a very narrow context (for collaboratively solving a problem task). It would also be interesting to see if SPAs can add value to other kinds of collaborations. This work has implications for the CHI community because it shows how SPAs can be used beyond their classical use case of one-to-one assistance for rather simple tasks. To our knowledge, this is among the first study that directly investigated the effects and differences of SPAs compared to human tutors in collaborative problem-solving settings. We plan to conduct more laboratory and field experiments in order to investigate the design and use of SPAs in this area.

REFERENCES

[1] Reuben M. Baron and David A. Kenny. 1986. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 51, 6, 1173–1182. DOI: https://doi.org/10.1037//0022-3514.51.6.1173.

[2] Pierre Dillenbourg. 1999. *Collaborative Learning: Cognitive and Computational Approaches (Advances in learning and instruction series)*. Elsevier Science, Inc., PO Box 945, Madison Square Station, New York, NY 10160-0757.

[3]   eMarketer. 2017. *Alexa, Say What?! Voice-Enabled Speaker Usage to Grow Nearly 130% This Year* (May 2017). Retrieved 24.12.18 from https://www.emarketer.com/Article/Alexa-Say-What-Voice-Enabled-Speaker-Usage-Grow-Nearly-130-This-Year/1015812.

[4]   Rachel Harsley, Barbara Di Eugenio, Nick Green, Davide Fossati, and Sabita Acharya. 2016. Integrating Support for Collaboration in a Computer Science Intelligent Tutoring System. In *Intelligent Tutoring Systems. 13th International Conference, ITS 2016, Zagreb, Croatia, June 7-10, 2016. Proceedings*, Alessandro Micarelli, John Stamper and Kitty Panourgia, Eds. Lecture Notes in Computer Science, 9684. Springer International Publishing, Cham, s.l., 227–233. DOI: https://doi.org/10.1007/978-3-319-39583-8_22.

[5]   Friedrich Hesse, Esther Care, Juergen Buder, Kai Sassenberg, and Patrick Griffin. 2015. A Framework for Teachable Collaborative Problem Solving Skills. In *Assessment and Teaching of 21st Century Skills*, Patrick Griffin and Esther Care, Eds. Springer Netherlands, Dordrecht, 37–56. DOI: https://doi.org/10.1007/978-94-017-9395-7_2.

[6]   Heisawn Jeong and Cindy E. Hmelo-Silver. 2016. Seven Affordances of Computer-Supported Collaborative Learning: How to Support Collaborative Learning? How Can Technologies Help? *Educational Psychologist* 51, 2, 247–265. DOI: https://doi.org/10.1080/00461520.2016.1158654.

[7]   Minchi C. Kim and Michael J. Hannafin. 2011. Scaffolding problem solving in technology-enhanced learning environments (TELEs): Bridging research and theory with practice. *Computers & Education* 56, 2, 403–417. DOI: https://doi.org/10.1016/j.compedu.2010.08.024.

[8]   Robin Knote, Andreas Janson, Matthias Söllner, and Jan M. Leimeister, Eds. 2019. *Classifying Smart Personal Assistants: An Empirical Cluster Analysis*. Proceedings of the Hawaii International Conference on System Sciences (HICSS).

[9]   James A. Kulik and J. D. Fletcher. 2016. Effectiveness of Intelligent Tutoring Systems. *Review of educational research* 86, 1, 42–78. DOI: https://doi.org/10.3102/0034654315581420.

[10]   Rohit Kumar, J. Beuth, and and Carolyn P. Rosé. 2011. Conversational Strategies that Support Idea Generation Productivity in Groups. In *Connecting Computer-Supported Collaborative Learning to Policy and Practice*. International Society of the Learning Sciences (ISLS), 398–405.

[11]   Simon Mcalister, Andrew Ravenscroft, and Eileen Scanlon. 2004. Combining interaction and context design to support collaborative argumentation using a tool for synchronous CMC. *Journal of Computer Assisted Learning* 20, 3, 194–204.

[12]   OECD. 2017. PISA 2015 collaborative problem-solving framework. In *PISA 2015 Assessment and Analytical Framework*. PISA. OECD Publishing, 131–188. DOI: https://doi.org/10.1787/9789264281820-8-en.

[13]   Margaret Rouse. 2016. *What is chatbot? - Definition from WhatIs.com* (2016). Retrieved December 24, 2018 from https://searchcrm.techtarget.com/definition/chatbot.

[14]   Nikol Rummel and Hans Spada. 2005. Learning to Collaborate: An Instructional Approach to Promoting Collaborative Problem Solving in Computer-Mediated Settings. *Journal of the Learning Sciences* 14, 2, 201–241. DOI: https://doi.org/10.1207/s15327809jls1402_2.

[15]   Isabella Seeber, Eva Bittner, Robert O. Briggs, Gert-Jan de Vreede, Triparna de Vreede, Doug Druckenmiller, Ronald Maier, Alexander B. Merz, Sarah Oeste-Reiß, Nils Randrup, Gerhard Schwabe, and Matthias Söllner. 2018. Machines as Teammates: A Collaboration Research Agenda. In *Proceedings of the 51st Hawaii International Conference on System Sciences*. IEEE Press Books. DOI: https://doi.org/10.24251/HICSS.2018.055.

[16]   Liane M. R. Tarouco, Clóvis Silveira, and Aliane L. Krassmann. 2018. Collaborative Learning with Virtual Entities. In *Learning and Collaboration Technologies. Design, Development and Technological Innovation*, Panayiotis Zaphiris and Andri Ioannou, Eds. Lecture Notes in Computer Science. Springer International Publishing, Cham, 480–493. DOI: https://doi.org/10.1007/978-3-319-91743-6_35.

[17]   Stergios Tegos and Stavros Demetriadis. 2017. Conversational agents improve peer learning through building on prior knowledge. *Educational Technology & Society* 20, 1, 99–111.

[18]   Alexandra Vtyurina and Adam Fourney. 2018. Exploring the Role of Conversational Cues in Guided Task Support with Virtual Assistants. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, New York, New York, USA, 1–7. DOI: https://doi.org/10.1145/3173574.3173782.

[19]   Armin Weinberger and Frank Fischer. 2006. A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education* 46, 1, 71–95. DOI: https://doi.org/10.1016/j.compedu.2005.04.003.