

Hey Alexa, Please Help Us Solve This Problem! How Interactions with Smart Personal Assistants Improve Group Performance

Completed Research Paper

Rainer Winkler

University of St. Gallen
Müller-Friedberg-Str. 8,
9000 St. Gallen
rainer.winkler@unisg.ch

Maya Lisa Neuweiler

University of St. Gallen
Müller-Friedberg-Str. 8,
9000 St. Gallen
maya.neuweiler@student.unisg.ch

Eva Bittner

University of Hamburg
Vogt-Kölln-Strasse 30,
22527 Hamburg
bittner@informatik.uni-hamburg.de

Matthias Söllner

University of Kassel
Henschelstr. 4,
34127 Kassel
soellner@uni-kassel.de &
University of St. Gallen
Müller-Friedberg-Str. 8,
9000 St. Gallen

Abstract

Despite a growing body of research about the design and use of Smart Personal Assistants, such as Google's Assistant or Amazon's Alexa, existing work has mainly focused on their use as task support for individual users in rather simple problem scenarios. Little is known about their ability to improve collaboration among multiple users in more complex problem settings. Using a mixed-method approach, we investigate the value of Smart Personal Assistants as facilitators by directly comparing 40 groups of three who either interact with a Smart Personal Assistant or a scripted human facilitator in a complex problem-solving scenario. The results indicate that groups interacting with Smart Personal Assistant facilitators show significantly better task outcomes and higher degrees of collaboration quality. A conversation analysis approach further analyzes differences in human SPA communication. Our work contributes to computer-supported cooperative work research and emphasizes the value of conversation analysis approaches in HCI research.

Keywords: smart personal assistant; computer-supported cooperative work; Amazon Alexa; experiment, conversation analysis

Introduction

Intelligent systems have become ubiquitous in modern life and are increasingly helping us perform everyday tasks in ways they could not previously. Specifically, the popularity of Smart Personal Assistants (SPA), such as Amazon's Alexa, Google's Assistant, Apple's Siri and other systems, has been steadily

growing over the past few years (eMarketer 2018). SPAs are software agents that can automate and ease many of the daily tasks of their users by engaging with them via voice-based, natural language dialog (Pais et al. 2015). Globally, the market of SPAs is predicted to increase across different sectors from 3.0 billion U.S. dollars in 2017 to 15.7 billion U.S. dollars in 2021 (Statista 2019b). In the United States, 65 percent of respondents from the ages of 25-49 stated that they were consistent (at least once per day) users of smart personal assistants (Statista 2019a). 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 natural language processing, SPAs have the potential to go beyond simple use cases by helping multiple users in more complex problem scenarios (Vtyurina and Fourney 2018). For example, Graesser et al. (2018) designed a SPA that tried to help dyads through a complex simulation task. These developments turn SPAs from simple question-answering tools into powerful group facilitators (Seeber et al. 2018).

Effective collaboration between group participants does not happen naturally (Briggs et al. 2003). It is a complicated mechanism that needs guidelines in order to be effective (Briggs et al. 2013). Past literature proved that human facilitators are able to improve group performance (Oxley et al. 1996; Witte 2007). Some studies indicate that new emerging SPAs might also be able to compete with human facilitators or even outperform them in some specific contexts (Seeber et al. 2018; Vanlehn 2011). For example, Lucas et al. (2014) found out that group members increase their willingness to disclose when they interact with a SPA compared to a human. Interacting with a computer lowers evaluation fears and impression management (Lucas et al. 2014). Moreover, interacting with a SPA might also lead to a higher degree of agency of the group. The degree of agency is a key factor in how people experience interactions with technology and humans (Limerick et al. 2014). Although the current state of technology is not yet ready to keep up with full-fledged expert facilitators, recent technological developments in machine learning and natural language processing let us assume that SPAs are soon be ready to imitate human facilitators (Seeber et al. 2018). A better understanding of how new emerging SPAs can increase group performance helps to replace costly human facilitators in future use scenarios and helps to inform future designs of SPAs for group support. Thus, in our study, we answer the following questions:

RQ1: Are Smart Personal Assistant facilitators better able to increase group task outcomes and the collaboration quality compared to scripted human facilitators?

RQ2: How do groups interact with Smart Personal Assistant facilitators compared to scripted human facilitators?

To answer our research questions, we ground our work on research in the field of computer-supported cooperative work and collaborative problem-solving. Based on that, we conduct a laboratory experiment with state-of-the-art SPA support. Using a mixed-method approach, we complement the quantitative results (RQ1) with a qualitative conversation analysis approach (RQ2). Mixed-method research can develop rich insights into various phenomena of interest that cannot be fully understood using only a quantitative or a qualitative method. Currently, there is a dearth of mixed method research in IS (Venkatesh et al. 2013). Especially conversation analysis approaches are not often used in IS studies so far (Sarker et al. 2013). In our experiment, groups of three participants have to conduct a 30-minute problem task, receiving help from either a SPA or a scripted human facilitator. The scripted human facilitator provided exactly the same scope of information and functions as incorporated into the SPA facilitator. We decided to limit the human facilitators' ability to a script, because we see this study as one of the first direct comparisons between SPAs and human facilitators. This study should first discover potential differences between SPA-human and human-human interactions serving as a basis for answering further questions regarding the value of SPAs as facilitators. By scripting the human facilitator, we were able to control potential other factors (prior knowledge of the human facilitator, etc.). Our results show that SPA facilitators outperform scripted human facilitators and presents three interaction patterns where human-SPA interaction differs from human-human interaction. The results suggest that SPAs might be a suitable replacement for human facilitators and even show some benefits. Building on the extensive work in human-SPA interaction, our work has two main implications for the HCI community. First, it shows how SPAs can be successfully used beyond traditional use cases in order to improve collaboration among group members. This study is one of the first that indicate the advantages of an SPA compared to a human facilitator to set the ground for later hypothesis development in that area. Second, we emphasize the

importance of using conversation analysis methods to directly compare human-SPA and human-human interaction. This approach helps to uncover often hidden interaction patterns in HCI which are very important to inform future research in that area. The remainder of this paper is structured as follows. The next section explains and classifies Smart Personal Assistants, and describes computer-supported collaborative work and collaborative problem-solving. Section 3 describes the setup of the lab experiment, including sample description, task design, technological development of the SPA, and the measurement details. Finally, we report our results, discuss them, and close with limitations, a future outlook and a conclusion.

Theoretical Background and Hypotheses Development

In this section, we first define and classify the term Smart Personal Assistant. Additionally, we describe computer-supported cooperative work and collaborative problem-solving that serves as a foundation for our hypotheses.

Smart Personal Assistants

SPAs are software agents that can automate and ease many of the daily tasks of their users by engaging with them via voice-based, natural language dialog (Pais et al. 2015). According to this definition, SPAs comprise all types of voice-based software systems that enable humans to hold goal-oriented, natural conversations with computers such as Amazon's Alexa, Google's Assistant, Apple's Siri and other systems. In essence, the SPA's "brain" is typically a cloud-hosted service, that converts human voice to text, performs linguistic analysis and carries out the requested action, whereas the devices that run the agent's program have no embedded intelligence (Chung et al. 2017). SPA providers offer rich ecosystems with easy-to-use toolkits and intuitive interfaces that allow users to create their own skills, thereby increasing SPA providers' own business value. Compared to traditional user assistance systems, SPAs can be characterized with a rather high degree of interaction and intelligence (Maedche et al. 2016). They are able to react to users' utterances and can proactively guide users through a complex task. For example, Fast et al. (2018) designed a SPA that support users in conducting data science tasks (e.g., a predictive modeling task). Recent studies indicate that these kind of SPAs are also effective facilitators for assisting multiple users in solving more complex tasks. For example, Tegos and Demetriadis (2014) demonstrated that SPAs are able to positively influence groups' dialog behavior, leading to substantially improved individual and task group outcomes. Goda et al. (2014) showed that a preceding conversation with a SPA lead to an increase in the number of contributions that participants made to conversations and could increase the number of conversations in which the group members participated. Moreover, studies in the field of computer tutoring proved that computer facilitators can achieve similar learning gains as human facilitators (Kulik and Fletcher 2016).

Computer-Supported Cooperative Work

Computer-supported cooperative work (CSCW) aims to improve the productivity and effectiveness of multiple individuals working together in a conscious way (Mills 2003). CSCW consists of software tools and technology that supports a group of individuals working together and adopts either a technology-centric or work-centric viewpoint. A technology-centric viewpoint emphasizes designing computer technology to support groups working together. A work-centric viewpoint emphasizes designing computer systems to support group work (Grudin 1994). Among others, there are four important dimensions inherent to CSCW (Mills 2003). One important dimension of CSCW technology occurs along the time dimension: is there a requirement to support cooperative work that occurs simultaneously (synchronously) or separately (asynchronously) or both? Another decision relates to space: must the individual collaborators be physically located at the same site, such as a room or an auditorium? A third important dimension is group size: must the system support a small team or a whole department? A fourth dimension must consider the interaction style: does the group require support for planned or spontaneous interactions or both? Our work concentrates on simultaneous cooperative work, where collaborators are physically located in the same room, because this kind of group work is one of the most common collaborative working scenarios in practice and allows us to control interfering variables within our laboratory experiment (Winkler et al. 2019). Moreover, we concentrate on small groups where interactions are taking place spontaneously and are driven by the participants themselves.

Collaborative Problem-solving

Collaboration is used to describe a variety of behaviors and generally occurs when more than one person works on a single task towards a common goal (Hesse et al. 2015), and group members “must make a conscious, continued effort to coordinate their language and activity with respect to shared knowledge” (Roschelle and Teasley 1995). Collaborative problem-solving (CPS) is a specific form of collaboration, which is the responsive approach of a team towards jointly solving a problem by exchanging ideas (Hesse et al. 2015). Today’s workplaces highly demand people who can solve problems together with others (Nelson 1999; PISA 2017). Effective 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”. CPS incorporates the components of cognition found in individual problem-solving (cognitive dimension) in addition to the components of collaboration (social dimension, Hesse et al. 2015, Dillenbourg 1999). The cognitive component consists of taking appropriate actions to solve the problem (i.e., it encompasses the team members’ ability of problem definition, the understanding of the team members prior knowledge, the identification of gaps in what is known and what is required, the integration of these in order to generate solutions, and the monitoring of the group process towards achieving the common goals, Graesser et al. 2018), and the social component consists of maintaining a shared understanding across the group members (i.e., the coordination of actions in synchrony with other participants, Hesse et al. 2015; PISA 2017). Shared knowledge building often resonates with common ground, and mental models of a team (Roschelle 2013). In this context, the facilitator’s role is to provide ongoing and individualized feedback to the group members to improve their cognitive and social actions (Merrill et al. 1996). Computers have often tried to support groups within problem-solving activities (Jeong and Hmelo-Silver 2016). For example, Harsley et al. (2016) found out that student groups were able to significantly increase their programming skills while receiving help from a SPA facilitator within a collaborative problem-solving scenario. Similarly, Mcalister (2004) demonstrated that group members interacting with SPA-similar systems showed higher argumentation skills leading to better task outcomes. Tarouco et al. (2018) logged and analyzed student-SPA dialogs within a collaborative problem-solving scenario and figured out that students tend to perceive SPAs as their collaboration partners.

New emerging SPAs have the capacity to track the contributions of team members and the group as a whole, automatically provide feedback, and offer recommendations to team members whenever the team is stuck (Graesser et al. 2017). Some studies indicate that new emerging SPAs might also be able to compete with human facilitators to some extent or even outperform them in some specific contexts (Seeber et al. 2018; Vanlehn 2011). For example, Lucas et al. (2014) found out that group members increase their willingness to disclose when they interact with a SPA compared to a human. Interacting with a computer lowers evaluation fears and impression management (Lucas et al. 2014). One reason for that might be that group members might easily feel watched or distracted when an often unknown facilitator tries to advise them (Xolocotzin Eligio et al. 2012). Furthermore, interacting with a SPA might lead to a higher degree of agency perceived by the group members. The degree of agency is a key factor in how people experience interactions with technology and humans (Limerick et al. 2014). Additionally, recent studies in the area of trust indicate that people tend to trust automated machines more than humans in specific conditions (Yang et al. 2017). Furthermore, recent developments in voice recognition and machine learning turn SPAs into powerful interaction partners similar to human partners (Seeber et al. 2018). Nevertheless, the current state of technology is not yet ready to keep up with full-fledged expert facilitators, but if SPAs and human facilitators are limited to the same scope of information and functions, we assume that groups interacting with SPA facilitators are able to achieve better task group outcomes. Hence, we propose:

H1. Groups interacting with SPA facilitators achieve better group task outcomes compared to groups interacting with scripted human facilitators within a problem-solving context.

Additionally, SPA facilitators might also have different effects on the groups’ collaboration quality itself. Past research indicates that SPA and human facilitators have a different effect on participants’ emotions (Litman et al. 2004). For example, Vail et al. (2016) found out that human facilitators can have a significant impact on participants’ affective response, such as joy or frustration. Moreover, Pickard et al. (2016) were able to show that people tend to be more open to SPAs, feel more comfortable and are more likely to evoke negative self-admissions. Especially in relationships of authority, where an often unknown

facilitator provides advice, participants might feel controlled and are, thus, easily inhibited in their collaborative behavior. Furthermore, Gratch et al. (2007) found out that virtual humans can be more engaging than real ones, measured by the total number of words participants spoke while they had an interaction. If SPAs and human facilitators are limited to the same scope of information and functions, we assume that SPA facilitators are able to increase the collaboration quality within the group. Thus, we propose:

H2. Groups interacting with SPA facilitators achieve higher levels of collaboration quality compared to groups interacting with scripted human facilitators within a problem-solving context.

Research Methodology

To answer our research questions, we conduct a laboratory experiment in which groups of three solve a 30-minute problem task while receiving help from either SPA facilitators or scripted human facilitators. The treatment group (TG) was able to communicate with the SPA facilitator via an Alexa Echo Dot device, while the control group was able to communicate with the scripted human facilitator located in the same room. We use a scripted human facilitator as a baseline for two reasons. First, the current state of technology is not yet ready to keep up with full-fledged expert facilitators. Second, we wanted to ensure that the two groups are as comparable as possible in order to show differences between interactions with SPAs and human facilitators. The human facilitator followed a script with exactly the same scope of information and functions as incorporated into the SPA. The groups were not allowed to use the internet or any other aid.

Sample

We were able to recruit a total of 127 participants. We had to exclude seven participants from our sample because some participants did not show up, resulting in subgroups less than three. All together, we had 20 groups of three in the treatment group (SPA) and 20 groups of three in the control group (human). The participants were students from a Swiss business university lab pool with an average age of 21.59 (undergraduate = 82, graduate = 38). Most of the students studied business-related subjects (e.g. economics, business studies, law). The sample consists of 68 men and 52 women. All participants were randomly assigned to one of the subgroups. Table 1 shows the samples' characteristics between the two groups.

Table 1. Sample Characteristics		
	Treatment Group (SPA)	Control Group (Human)
Gender	m = 35, w = 25	m = 33, w = 27
Average Age	21.29	21.88
Study background	18 x economics, 30 x business studies, 10 x law, 2 x others	15 x economics, 31 x business studies, 9 x law, 5 x others
Pre-experience with SPAs	5.21 (out of 7)	5.02 (out of 7)
Personal innovativeness	4.83 (out of 7)	4.79 (out of 7)

We conducted ANOVA tests to make sure that the two groups were equal in terms of gender, age, study background, pre-experience with SPAs and personal innovativeness. The randomization was successful since all results were not significant ($p < 0.05$). Participants received a baseline monetary reward for participating with an optional additional reward depending on their group task outcome.

Task Design

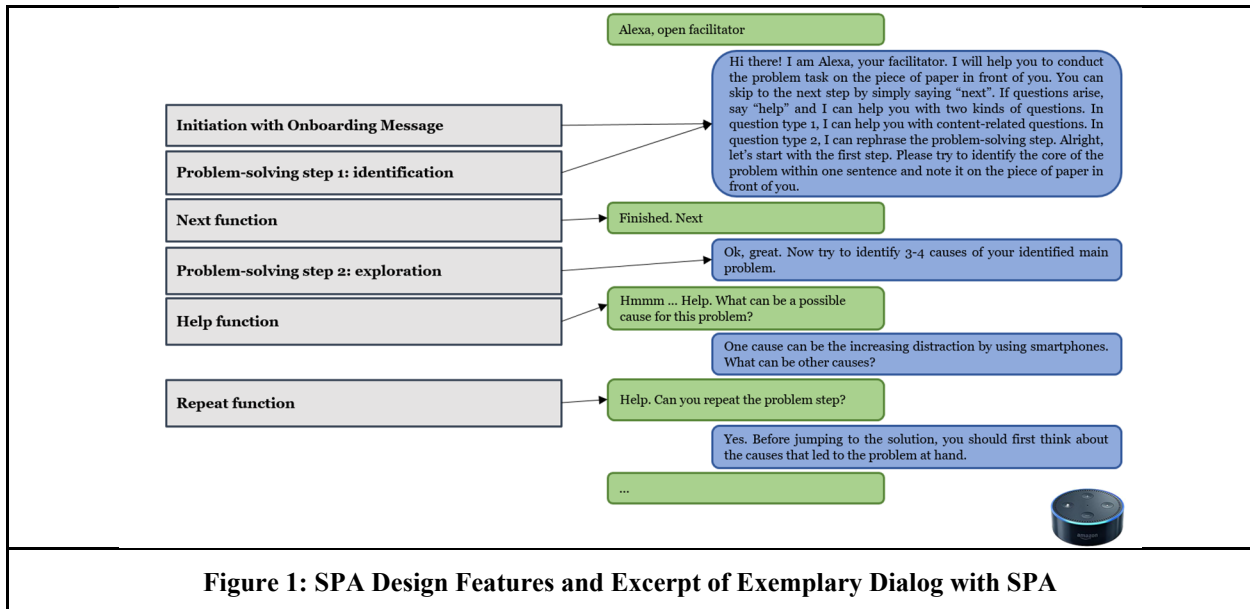
The goal of the study is to identify whether SPA facilitators are better able to increase group task outcomes and the collaboration quality compared to scripted human facilitators within a problem-solving context. Hence, it was important to choose a well-defined problem task. Problem-based learning theory describes a task as a problem task when it is ill-structured, open-ended, realistic, and resonates with the executors' experience (Jonassen 2011). A problem task definition should further promote conjecture and argumentation (Hmelo-Silver 2004). Based on the described criteria of problem task definitions, we

decided to adapt a problem task from Gallagher et al. (1992). Table 2 shows how our task addresses the criteria of a problem task definition proposed by Jonassen (2011) and Hmelo-Silver (2004).

Table 2. Group Task and Requirements	
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?	
Requirements of a problem task definition	Description of how our task addresses the requirements
R1: Ill-structured	The task can be considered as ill-structured because not every piece of information is given. Moreover, it gives no advice on how to solve the task.
R2: Open-ended	The task allows to develop different solutions. “The right solution” does not exist.
R3: Realistic	This is a problem that cities have to deal with a lot, since increasing traffic flow is a fact. It can therefore be considered as realistic.
R4: Resonate with the executors’ experience	Since every participant has been in a car before and most of them have a driver’s license, they can put themselves in the problem task’s position.
R5: Promote conjecture and argumentation	The task is missing information about the causes, the core of the problem, etc. It therefore opens room for conjecture and different kinds of argumentations.

Technological Development of the Smart Personal Assistant

To answer our research questions and show the intended effects, it was necessary to develop a SPA that is able to help the group members in the best way possible. Thus, we developed an actual SPA that is able to interact with the participants via natural language. For this purpose, we used Amazon’s Alexa Skill Development Kit 2.0 with nodeJS since this programming framework seems to be the current state-of-the-art regarding speech recognition and natural language processing. The Alexa Skill Development Kit 2.0 provides programmatic access to Alexa features as well as a lot of tutorials and blueprints, so we were able to use basic coding patterns to create rich, interactive experiences (Amazon 2019). We included two different logics in our interaction model. Figure 1 shows an excerpt of an exemplary group dialogue with the SPA facilitator and the corresponding design features (the communication between group members is not depicted).



In our first logic (proactive), the SPA facilitator proactively guided students through the task by providing five different problem-solving steps adapted from Kim and Hannafin’s (2011) problem-solving phases: (a) *identification* (define the problem), (b) *exploration* (search information), (c) *reconstruction* (select and process information), (d) *presentation and communication* (information aggregation and communicate result), (e) *reflection and negotiation* (looking back, revising ideas and theories). This way we make sure

that the SPA facilitator is able to help participants solving the problem in a structured way. The SPA facilitator provided its problem-solving steps in an interval of five minutes. The participants were also able to skip to the next step, in case they finished the step earlier. In our second logic (reactive), the SPA was able to react to participants' questions that occurred between the different problem-solving steps. The participants were either able to ask for content-specific help or ask the facilitator to rephrase the current problem-solving step.

Experimental Procedure

Participants in the treatment group (SPA facilitator) received the problem task on a piece of paper and were able to access the facilitator with the help of a standalone SPA device (Alexa's Echo Dot). Participants in the control group (scripted human facilitator) received the problem task on a piece of paper and the human facilitator was located in the same room as the participant group using a script with exactly the same scope of information and functions as incorporated in Alexa. As already mentioned in the introduction, we scripted the human facilitator, because we wanted to control for potential other factors (such as prior knowledge of the human facilitator, etc.). See Figure 2 for an exemplary set up with both treatment groups.

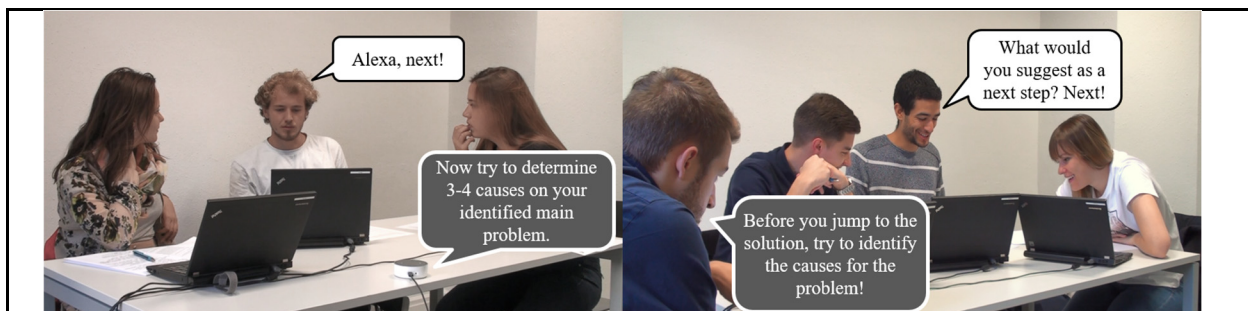
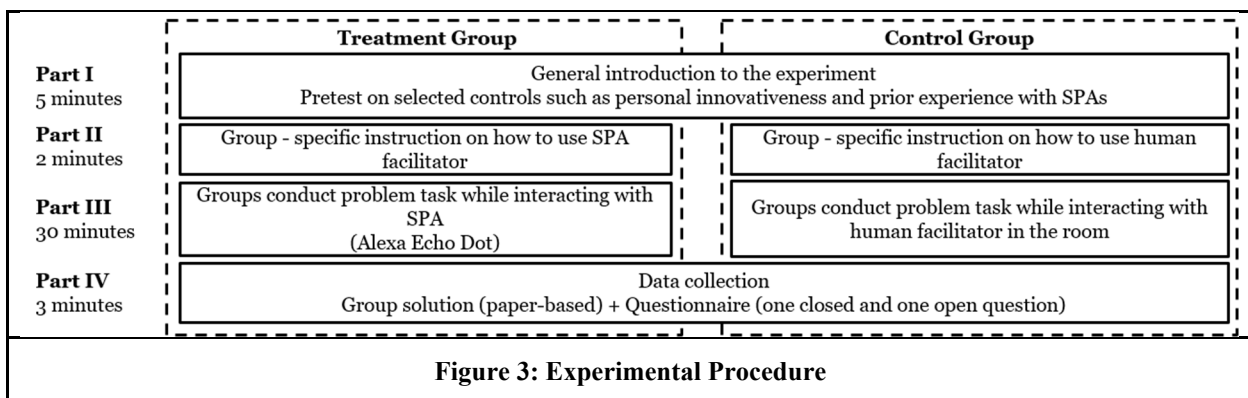


Figure 2: Example Setup with Alexa Echo Dot (left) or scripted human facilitator (right).

The experimental procedure is divided into four components. Figure 3 outlines the procedure including activities done and the time spent in each component. In part 1, participant groups entered the room (group after group), received general instructions on the experiment and had to complete a pretest on selected controls, such as gender, age, study background, personal innovativeness and prior experience with SPAs.



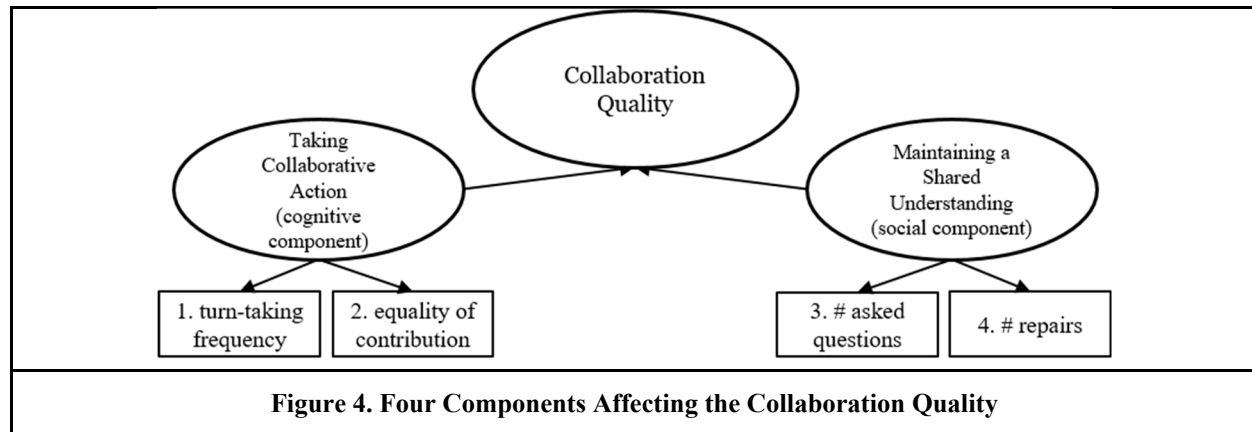
Two persons were used to give general instructions to the participant groups. These persons used a detailed script to prevent variation in the implementation. In part 2, every group received a group-specific instruction on how to use their type of facilitator. In part 3, the groups had 30 minutes to conduct the problem task while interacting with their facilitator. In part 4, we collected the group solution and participants individually filled out a short survey with two questions regarding the perceived helpfulness

of the facilitator (from 1 to 7, manipulation check) and their overall experience of the collaboration (open question). Additionally, we recorded all the sessions on video.

Measurement and Analysis

Quantitative Data

For measuring the group task outcome, two experienced raters evaluated the group results using an evaluation framework aligned with the problem-solving steps of Kim and Hannafin (2011). The evaluation framework was developed collectively in a workshop within the research team. The framework contains detailed information on 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. For example, if a group systematically derives and justifies the causes of the problem from different perspectives (problem-solving step 2: exploration), the group receives a high score near to 7. The group task outcomes were rated individually, blinded, and independently from each other. Furthermore, we checked for interrater agreement with the help of a Pearson correlation. The correlation resulted in a satisfactory result (interrater agreement = 0.9699, $p < 0.05$). For measuring the collaboration quality, we analyzed the cognitive and social component proposed by Graesser et al. (2018) and Hesse et al. (2015). The cognitive component evaluates how the participants manage the task, contribute, and build knowledge through taking appropriate action towards the solution of the problem (Hesse et al. 2015). Thus, we individually analyzed all 40 video recordings regarding (1) the frequency of applied turns and (2) the homogeneous content-specific contribution proposed by Meier et al. (2007) and Setlock et al. (2004). The frequency of turns indicates how intensively participants interact with each other. Higher turn-taking implicates that participants act more responsive to each other's utterances, i.e., they more reflect on what was stated as well as elaborate or build on arguments (Drew 2005). The equality of contribution indicates the balance of the contributions of the individual group members measured by the average length of words per utterance and the corresponding standard deviation (Hutchby and Wooffitt 2008). Hence, it is an indicator of whether the solution was developed by the joint effort of all group members or rather on the basis of a single strong participant. The two dimensions and the four components leading to our final score for the collaboration quality are depicted in Figure 3.



The social component evaluates the group members effort in building and maintaining a shared understanding, which is a prerequisite for effective collaboration. We measured the building and maintaining of a shared understanding with (3) the number of on-task related questions and (4) the number of used repairs proposed by Rummel and Spada (2005). Asking questions prevents misunderstandings and fosters a shared understanding of participants' utterances. This component was measured by the sum of question-answer pairs occurred per session. The fourth measurement counts the total number of repairs per group throughout the discourse. Repairs occur, whenever one participant detects a mismatch of intentions between the speaker and the recipient and subsequently corrects it (Dillenbourg et al. 2016). Therefore, the participants first need to detect a misunderstanding, that is, identify a mismatch between what one speaker intended to say and what was understood by another, and

try to repair it with a question (Corti and Gillespie 2016; Dillenbourg et al. 2016). Table 3 shows an overview of all measurement components of collaboration quality and describes how we measured them.

Table 3. Measurement of Collaboration Quality	
Measurement	Description
1. Turn-taking frequency	Average amount of switches between utterances of group members
2. Equality of contribution	Average amount of contribution per participant measured in average words per utterance and standard deviation within the whole task
3. # asked questions	Average amount of answered questions within the whole task incl. facilitator
4. # repairs	Average amount of successfully detected and repaired misunderstandings within the whole task incl. facilitator

The collaboration quality corresponds to the equally weighted and standardized averages of these two dimensions and four components. For pre-experience with SPAs, we asked the participants how often they use SPAs on a scale reaching from 1 to 7 where 1 is “never used” and 7 is “use it every day”. For personal innovativeness, we used the four items from van Raaij and Schepers (2008). For quantitative data analysis, we used a one-way MANOVA test to identify if there are differences between the groups in terms of task group outcome and collaboration quality. A MANOVA helps us to check if there is a relationship between task group outcome and the collaboration quality which might influence the hypotheses to be tested (French et al. 2008). Furthermore, we calculated partial eta squared for measuring the effect size (0.01 = small, 0.06 = medium, 0.14 = large, Cohen 1988). We also included a post-experiment manipulation check to see if facilitators have an overall effect on the groups by letting participants rate the perceived helpfulness of the facilitator (5.83 out of 7, no significant differences between the groups). Moreover, we used the statistic program R as a tool for analysis (Team 2013).

Qualitative Data

To gain deeper insights in how groups interact with SPAs compared to human facilitators, we used a conversation analysis approach. Conversation analysis aims to uncover hidden interaction patterns and/or give guidelines on how to improve interaction (Hoey and Kendrick 2017). The use of conversation analysis as a tool to unpacking the sequential and embedded practice of situated actions is an established technique in human-human interaction (Mackey and Gass 2011). Since computers become more interactional, conversation analysis as a method becomes more popular in HCI research (Porcheron et al. 2018). To capture interaction fragments between the group members and their facilitators, we recorded all the sessions on video. The results consist of approx. 20 hours of recorded data. We focused our analysis on groups’ help requests (help function) directed to the facilitator, since this was the main reason why participants interacted with the facilitator. Overall, we identified over 175 distinct ‘help request’ utterances, that is, talk directed at the facilitator in a seeming attempt to receive help. Often, these requests formed part of a larger sequence encompassing more than one help request. Thus, all in all, the corpus consists of 122 of these interaction fragments.

The collecting and analyzing of the data follow the conversation analysis approach of Hoey and Kendrick (2017) and is divided into seven sub-steps: (1) identify a candidate phenomenon, (2) build a collection of cases, (3) analyze each case in the collection, (4) analyze variation in the collection, (5) define the boundaries of the phenomenon, (6) analyze deviant cases and look for normative evidence, and (7) produce a formal account of the phenomenon. In step 1 (identify a candidate phenomenon), we analyzed how participants approach and react to the facilitators’ utterances. Therefore, the video recordings have been scanned for help requests of group members directed at either Alexa or the scripted human facilitator. In step 2 (build a collection of cases), we collected several fragments including multiple turns prior and after the help request and follow a set of standard transcription conventions according to the often used Jefferson’s (1984) Transcription Notation. Illustrated by a short sequence, we show how Jefferson Transcription Notation works: We note where pauses take place ((1.4) for pauses of 1.4 seconds and (.) for pauses between words shorter than 0.2 seconds), where talk is LOUD or ‘quiet’, where it is spoken >faster< or <slower> than usual, and where an utterance is elongated. A period “.” and a comma “,” indicate a falling or slightly rising intonation contour, respectively. Overlapping talk is represented using indentation and [square brackets], ((unspoken actions)) are given in double parentheses, and (unclear words or utterances) in single brackets. Utterances in bold are directed to the facilitator and all used names are replacements for the real names to ensure anonymity. Talk is represented as it is

produced, and not how it is intended to. This means that the fragments are presented and analyzed in their original language (Swiss German or German). Therefore, the result section contains interaction fragments that are in the original language and translated into English for a better comprehensibility. In step 3 (analyze each case in the collection), we started to analyze each case individually starting with the clearest cases that show differences between the treatment and control group. In step 4 (analyze variation in the collection), we tracked forms of variation across the two treatment groups and sorted the cases accordingly. In step 5 (define the boundaries of the phenomenon), we looked for counterexamples to analyze the boundaries of our identified interaction patterns. In step 6 (analyze deviant cases and look for normative evidence), the goal was to describe the interaction patterns that groups use when talking to SPAs compared to talking with humans. Finally, in step 7 (produce a formal account of the phenomena), we describe the interaction patterns in detail and explain how they operate. The interaction fragments we demonstrate in this paper are taken from different groups in the SPA facilitator or human facilitator group. We present the fragments as exemplars of identified patterns. This means that they are not the sole example of such reactions in our corpus. Rather, they demonstrate patterns that have been repeatedly found in the data.

Results

RQ1: Are Smart Personal Assistant facilitators better able to increase group task outcomes and the collaboration quality compared to scripted human facilitators?

Table 4 shows the summary of means and standard deviations for group task performance and the collaboration quality including its individual components. Groups interacting with a SPA facilitator showed both a higher collaboration quality and a higher task outcome.

Table 4. Descriptive Statistics of Treatment and Control Group				
	Treatment Group (SPA, n=20)		Control Group (scripted human, n=20)	
	M	SD	M	SD
1. Task Outcome	3.32	0.74	2.71	0.54
2. Collaboration Quality	5.25	1.06	4.30	1.29
2.1 Turn – taking frequency	220.83	68.32	171.80	59.20
2.2 Equality of contribution	20.97	9.66	17.44	11.49
2.3 # asked questions	55.83	21.11	33.0	8.28
2.4 # repairs	1.00	1.10	1.40	2.61

Before using a MANOVA, we tested assumptions. In specific, we conducted a test for normality and linearity as well as a test for homogeneity of variance. All assumptions are met. Next, we conducted the MANOVA test to see if group differences are significant. The results revealed that groups interacting with SPAs were able to significantly increase both group task outcomes and collaboration quality ($F(1,38) = 6.485$, $p = 0.0040$, $\eta^2 = 0.189$, $n = 40$). When looking at each dependent variable one at a time, we see that groups interacting with SPAs have significantly better task outcomes ($F(1,38) = 8.444$, $p = 0.0062$, $n=40$). Moreover, groups interacting with SPAs show significantly higher levels of collaboration quality ($F(1,38) = 6.880$, $p = 0.0292$, $n = 40$). Since our results are significant and the effect size can be considered as high, H1 and H2 can both be supported (Cohen 1988).

RQ2: How do groups interact with Smart Personal Assistants as facilitators compared to scripted human facilitators?

In the following, we will describe the identified interaction patterns and show exemplary interaction fragments. Again, the phenomena of interest were the help requests of participants directed at the facilitator and the corresponding interactions. We identified three interaction patterns that differed between groups interacting with SPAs compared to groups interacting with human facilitators: (1) more questions for hints and out of idle curiosity, (2) longer gaps between turns, (3) more laughter before and after SPA utterances

Interaction Pattern 1: More questions for hints and out of idle curiosity. The first interaction pattern relates to the amount and kind of questions groups asked the facilitators. Generally, we identified three reasons why the participants approached their facilitator for help: (1) to get a clearer understanding of the task (i.e., if they are unsure about what to do and what profoundness of response is asked for), (2) to get additional hint (e.g., if they are required to name three causes for traffic congestion and they only could think of two), (3) from idle curiosity what input a facilitator might provide. We observed that groups interacting with the SPA facilitator requested hints and questions for reasons of curiosity more often. Fragment 1 demonstrates an interaction where participants ask out of idle curiosity.:

... 01 Valeria: Sollen wir aus Interesse Alexa fragen? 02 (1.4) 03 Nicolas: Ich:: 04 Valeria: Weil ich also 05 Nicolas: bin dafür (.) Alexa mal auszuprobieren 06 [((lacht))] 07 Valeria: [Ja::<] weil also brauchen, die 08 Aufgaben sind einfach, brauchen tut 09 man die eigentlich nicht, (oder)? 10 Nicolas: Ja genau 11 (1.4) 12 Valeria: hmm:: 13 (1.5) 14 Nicolas: Also Alexa (1.6) hilfe. 15 (2.0) 16 Nicolas: oh ok. [((lacht))] 17 Anne: [((lacht))] 18 Valeria: [((lacht))] 19 Nicolas: Alexa (2.1) help. ...	Should we ask Alexa out of curiosity? I Because I Agree to try Alexa [((laughing))] [Yes] because need, well the tasks are straightforward, so one does not really need her, (right)? Yes exactly Hmm Alright Alexa hilfe. Oh ok. [((laughing))] [((laughing))] [((laughing))] Alexa help.
Left = Original Excerpt in German	Right = English Translation

Figure 5. Fragment 1: Asking for Idle Curiosity, Treatment Group

Interaction Pattern 2: Longer gaps between turns. The second pattern we identified is related to the gaps after the facilitators' utterances. As can be noted from fragment 2, the participants interacting with the SPA (TG) produce mutual silence period of 2 seconds or longer, (see line 12, Fragment 2) after the facilitator's input, before they jointly discuss the input received. Quite the contrary can be observed in groups interacting with a human facilitator as fragment 3, line 15, shows. The difference in the length of produced silence between the two treatment groups holds across nearly all sessions.

... 01 Sandro: Wollen wir sie fragen? 02 ((zeigt auf Echo)) 03 Jonas: Ja tuen emol probiere. 04 (0.8) 05 Sandro: Alexa, Help. Was ist eine Lösung? 06 (1.8) 07 Alexa: Okay, ich gebe euch einen Hinweis. 08 Eine mögliche Lösungsoption könnte im 09 Bereich Verkehrsregulierung sein, 10 also zum Beispiel Kreisverkehre. Was 11 könnten weitere Lösungsoptionen sein. 12 (2.2) 13 Jonas: Joh:: (2.3) (e chrüzig) ...	Do we want to ask her? ((points at the Echo device)) Yes try it. Alexa, Help. What is a solution? Okay, I will give you a hint. A possible solution might relate to traffic regulation, as for example roundabouts. What could be further solutions? Well an intersection
--	---

Figure 6. Fragment 2: Breaks after SPA Utterance, Treatment Group

... 01 Johanna: Fragen wir nach Hilfe? 02 Elisa: Ja 03 Nicole: Ok. Facilitator, help us. Wie finde 04 ich die beste Lösung? 05 (0.9) 06 Facilitator: Wenn ihr das Gefühl habt, dass die 07 Lösungsoption das Kriterium gut 08 erfüllt, dann setzt eine Vier oder 09 Fünf und sonst eine andere Zahl. 10 Wenn wir als Beispiel Verkehrs- 11 regulierung und da Kreisverkehre 12 nehmen, dann halten sich die 13 Kosten als Kriterium im Rahmen und 14 würde zum Beispiel eine Vier ergeben. 15 (0.6) 16 Johanna: Ah Kosten sind ein Kriterium zum 17 Beispiel 18 Elisa: Ah 19 (0.6)	Let's ask for help? Yes Ok. Facilitator help us. How to find the best solution? Okay, I will give you a hint. If you believe, a solution fulfills the criterion well so give it a five and otherwise a one. If we take roundabouts as an example, then costs are within reasonable limits and would account for a four. Ah costs are a criterion for example Ah
--	--

Figure 7. Fragment 3: Breaks after Human Facilitator Utterance, Control Group

Interaction Pattern 3: More laughter before and after SPA utterances. The third pattern we identified was that participants had more laughter before and after the help requests directed to the SPA, as can be seen from Fragment 4, lines 13-16. In the human SPA group, the participants nearly never responded with laughter to the humans' advices.

...		
01 Marc:	Alexa stop.	Alexa stop.
02	(2.4)	
03 Alexa:	Wie kann ich euch helfen? Sagt next	How can I help you? Say next help or
04	help oder repeat.	repeat.
05 Marc:	Help.	Help.
06	(1.4)	
07 Alexa:	Okay, ich gebe euch einen Hinweis.	Okay, I will give you a hint. A potential
08	Eine mögliche Ursache ist zum Beispiel	cause is for example the lack of ring
09	steigendes Verkehrsaufkommen. Was	roads or public transportation. What
10	könnten weitere mögliche Ursachen	could be further causes.
11	sein.	
12	(1.0)	
13 Fabian:	[[((lacht))] ja easy	[[((laughing))] well easy
14 Claudio:	[[((lacht))]]	[[((laughing))]]
15 Marc:	[[((lacht))]]	[[((laughing))]]
16	(1.2)	
...		
Figure 8. Fragment 4: Laughing after SPA Utterance, Treatment Group		

Discussion

Using a mixed-method approach, this study evaluates the use of SPAs as facilitators in a collaborative problem-solving scenario. The quantitative findings suggest that SPAs have the potential to outperform human facilitators if both have the same scope of information and functions. The qualitative findings provide more in-depth insights on possible reasons for that. Specifically, the quantitative analysis has shown that groups interacting with SPA facilitators took turns more frequently, indicating that these groups had a more vivid discussion and intense contribution. One reason for that might be that participants feel more relaxed and open when no human facilitator is in the room who can judge them (Pickard et al. 2016). This goes in line with our identified interaction pattern 1 showing that groups interacting with SPAs asked more questions for hints and out of idle curiosity. Consequently, they show a higher willingness to include the SPA facilitator into their collaborative activities.

Furthermore, the conversation floor of the SPA groups was smoother, as they had a more equal contribution. For instance, it was nearly never the move of a sole participant to address the SPA facilitator without having the consent of at least one of the other group members. The regulation of whether and who initiates the interaction with the SPA as such is therefore already a collaborative achievement and promotes the contribution of several participants. One reason for a more equal contribution might be that all participants feel more confident and comfortable when interacting with SPA facilitators compared to human facilitators. Similar effects can be discovered in the field of language learning, where students feel more open when talking to a machine compared to a human facilitator (Wik and Hjalmarsson 2009). It seems that group members feel a higher degree of agency when getting support from a computer instead of a human facilitator leading to a more efficient collaboration (Limerick et al. 2014). This is also reflected by our identified patterns 1 and 3, indicating that participants asked more questions and had more laughs during interaction with the facilitator suggesting that they are more open and unrestricted.

Moreover, the quantitative data revealed that groups interacting with SPA facilitators put greater effort into the achievement and maintenance of a shared understanding of the problem between group members. This could be shown in the total number of asked and appropriately answered questions about content and for approval of their arguments. These results can be confirmed by Roschelle and Teasley (1995), who state that successful collaboration involves a large degree of mutual engagement, joint decision making, and discussion. Conversely, the quantitative results further showed that groups interacting with the human facilitator included more repaired misunderstandings. This goes in line with the findings of Corti and Gillespie's (2016), revealing that participants put less intersubjective effort (assessed on repair activity) into interacting with an artificial agent than with a human being. Moreover, the qualitative data revealed that participants had longer gaps after they received an advice from the SPA. After more than two seconds, they reacted to the advice and continued their discussion whereas groups receiving advice from the SPA reacted immediately. One reason for these produced silences might be that group members listen to the SPA facilitator more carefully, whereas advice of the human facilitators is

directly taken over as solutions. This might come from politeness, as longer pauses in human-to-human conversation are perceived as negative (Porcheron et al. 2017). Another explanation might be that participants felt less pressure from the SPA compared to human facilitators and, thus, took time to reflect what was just said. Our findings agree with Sack et al.'s (1978) substantial auditory analyses of conversational data, revealing that human-human communication typically occurs without any silence in-between and without any overlapping speech ("no-gap-no-overlap"). One other explanation for the differences might be that participants are better able to detect turn-ends made by a human compared to a SPA facilitator and can therefore prepare their reactions faster (Heldner and Edlund 2010).

Our work has several theoretical contributions. First, it contributes to computer-supported collaboration by showing that new emerging SPAs can be successfully used for supporting collaborative problem-solving scenarios. Although we did not use a freely interacting facilitator as a baseline, the findings indicate that SPAs might be able to directly compete with human facilitators in near future. Given that the technology around SPAs is developing rapidly, they can be considered as a useful alternative to cost-intensive human facilitators. More than that, our work also suggests some possible benefits of SPA facilitators compared to human facilitators. In this light, the findings can be considered as a starting point helping to understand possible benefits of SPA facilitators compared to human facilitators. For example, group members appear to be more open and feel higher degrees of agency when interacting with SPA facilitators. These insights can serve future hypothesis development where researchers can design experiments that reveal how certain SPA characteristics facilitate group collaboration better compared to human facilitators. As SPAs increasingly enter our everyday job and private lives, it is important to understand how we can leverage their potential beyond already existing, traditional use cases. Second, by applying a conversation analysis approach, we showed that there are some hidden interaction patterns when interacting with SPAs compared to humans. These findings demonstrate that conversation analysis should be used more often in IS research to uncover hidden interaction patterns that are not visible with quantitative or often used qualitative methods (e.g., interviews). The insights resulting from this method have the potential to inform future research in the area of human computer interaction. Regarding our practical implications, we showed SPA designers how to develop SPAs for collaborative problem task support beyond already known use cases. It is worth focusing on these kinds of use cases since collaboratively solving problems will become more important in future job environments and SPAs are becoming more ubiquitous. Finally, we gave SPA designers first impressions about differences between human-SPA and human-human communication, which helps to inform future designs of SPAs.

Limitations and Future Research

Our study has several limitations. The primary limitation relates to our control group. We compared the value of SPAs with scripted human facilitators. It can be expected that freely interacting human facilitators would have led to different results. However, we tried to ensure high quality support by aligning the used script to the well-known problem-solving steps of Kim and Hannafin (2011) and also included a manipulation check indicating that participants perceived the support as helpful. SPA technology is not yet ready to build up a free interaction with the user. Nevertheless, rapid developments in machine learning and natural language processing are promising and indicate that SPAs can soon turn into full-fledged collaboration partners or facilitators (Seeber et al. 2018). The second limitation is the small sample size of 40 groups. However, with using a mixed-method approach, we tried to partially compensate for that (Venkatesh et al. 2013). Despite the small sample size, the statistical tests were able to show that there are significant differences between groups interacting with SPA and human facilitators. It would be interesting to see if further research with larger sample sizes can confirm these results. A third limitation is that there might be within-group dependencies which increases the likelihood that the analysis will incorrectly suggest a significant effect (Baldwin et al. 2008). However, the goal of the study is to show that SPAs have an effect on social interaction. Hence, we accounted for intra-group dependencies with collaboration quality as one of our dependent variables. A fourth limitation is that some interactions with the SPA are driven by curiosity. This means that we cannot be fully sure that the effects are equal when curiosity levels are the same in both groups. However, we asked participants about their pre-experience with SPAs, which resulted in a rather high experience level. A fifth limitation is that we investigated our effects in a rather narrow computer-supported collaboration context (face-to-face, collaborative problem-solving). Further research should try to replicate our results with other kinds of collaboration settings (other tasks, group sizes, etc.). A sixth limitation is that our research design and

analysis does not allow to fully investigate *why* SPA facilitators perform better than scripted human facilitators. Future research should consider finding answers to that questions. Finally, it would be very interesting to see if SPAs can also keep pace with full-fledged expert human facilitators and whether other researchers can detect different interaction patterns with the help of conversation analysis methods.

Conclusion

This paper investigated the value of SPAs in a collaborative problem-solving context. Using a mixed-method approach, we first identified relationships quantitatively and enriched the data with a conversation analysis approach. The findings suggest that SPAs are able to improve participants' task performance and also their collaboration quality. Moreover, we were able to uncover some specific interaction patterns that differ between human-SPA and human-human communication. Our work contributes to computer-supported collaboration and emphasizes the use of conversation analysis as a method to uncover often hidden interaction patterns in human computer interaction.

References

- Amazon 2019. *New Amazon Alexa Skill Kit 2.0*.
<https://developer.amazon.com/blogs/alexa/post/1a4e8b01-663d-4680-8efd-c28e96e31655/now-available-version-2-of-the-ask-software-development-kit-for-java>. Accessed 20 November 2018.
- Atkinson, J. M., and Heritage, J. 1984. *Structures of social action*: Cambridge University Press.
- Baldwin, S. A., Stice, E., and Rohde, P. 2008. "Statistical analysis of group-administered intervention data: reanalysis of two randomized trials," *Psychotherapy research : journal of the Society for Psychotherapy Research* (18:4), pp. 365–376.
- Briggs, R. O., Kolfshoten, G. L., Vreede, G.-J. de, Lukosch, S., and Albrecht, C. C. 2013. "Facilitator-in-a-box: process support applications to help practitioners realize the potential of collaboration technology," *Journal of Management Information Systems* (29:4), pp. 159–194.
- Briggs, R. O., Vreede, G.-J. de, and Nunamaker Jr, J. F. 2003. "Collaboration engineering with ThinkLets to pursue sustained success with group support systems," *Journal of Management Information Systems* (19:4), pp. 31–64.
- Chung, H., Iorga, M., Voas, J., and Lee, S. 2017. "Alexa, Can I Trust You?" *Computer* (50:9), pp. 100–104.
- Cohen, J. 1988. *Statistical power analysis for the behavioral sciences 2nd ed.*: Erlbaum Associates, Hillsdale.
- Corti, K., and Gillespie, A. 2016. "Co-constructing intersubjectivity with artificial conversational agents: People are more likely to initiate repairs of misunderstandings with agents represented as human," *Computers in Human Behavior* (58), pp. 431–442.
- Dillenbourg, P., Lemaignan, S., Sangin, M., Nova, N., and Molinari, G. 2016. "The symmetry of partner modelling," *International Journal of Computer-Supported Collaborative Learning* (11:2), pp. 227–253.
- Drew, P. 2005. "Conversation analysis," *Handbook of language and social interaction* (2:3), pp. 71–102.
- eMarketer 2018. *Alexa, Say What?! Voice-Enabled Speaker Usage to Grow Nearly 130% This Year*.
<https://www.emarketer.com/Article/Alexa-Say-What-Voice-Enabled-Speaker-Usage-Grow-Nearly-130-This-Year/1015812>. Accessed 23 November 2018.
- Fast, E., Chen, B., Mendelsohn, J., Bassen, J., and Bernstein, M. 2018. "Iris: A Conversational Agent for Complex Tasks," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal, Canada: ACM, p. 473.
- French, A., Macedo, M., Poulsen, J., Waterson, T., and Yu, A. 2008. "Multivariate analysis of variance (MANOVA)," *San Francisco State University*.
- Gallagher, S. A., Stepien, W. J., and Rosenthal, H. 1992. "The effects of problem-based learning on problem solving," *Gifted Child Quarterly* (36:4), pp. 195–200.
- Goda, Y., Yamada, M., Matsukawa, H., Hata, K., and Yasunami, S. 2014. "Conversation with a chatbot before an online EFL group discussion and the effects on critical thinking," *The Journal of Information and Systems in Education* (13:1), pp. 1–7.

- Graesser, A. C., Cai, Z., Hu, X., Foltz, P. W., Greiff, S., and Shaffer, D. W. 2017. "Assessment of collaborative problem solving," *Design Recommendations for Intelligent Tutoring Systems* (5:2), pp. 275–285.
- Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., and Hesse, F. W. 2018. "Advancing the science of collaborative problem solving," *Psychological Science in the Public Interest* (19:2), pp. 59–92.
- Gratch, J., Wang, N., Okhmatovskaia, A., Lamothe, F., Morales, M., van der Werf, R. J., and Morency, L.-P. 2007. "Can Virtual Humans Be More Engaging Than Real Ones?" in *Human-computer interaction: 12th international conference, HCI International 2007, Beijing, China*, J. A. Jacko and Marie-Claire Jenkins, Richard Churchill, Stephen Cox, and Dan Smith (eds.), Berlin: Springer, pp. 286–297.
- Grudin, J. 1994. "Computer-supported cooperative work: History and focus," *Computer* (27:5), pp. 19–26.
- Hao, J., Zapata-Rivera, D., Graesser, A. C., Cai, Z., Hu, X., and Goldberg, B. 2018. "Towards An Intelligent Tutor For Teamwork: Responding to Human Sentiments," *Design Recommendations for Intelligent Tutoring Systems*, p. 151.
- Harsley, R., Di Eugenio, B., Green, N., Fossati, D., and Acharya, S. 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*, A. Micarelli, J. Stamper and K. Panourgia (eds.), Cham, s.l.: Springer International Publishing, pp. 227–233.
- Heldner, M., and Edlund, J. 2010. "Pauses, gaps and overlaps in conversations," *Journal of Phonetics* (38:4), pp. 555–568.
- Hesse, F., Care, E., Buder, J., Sassenberg, K., and Griffin, P. 2015. "A framework for teachable collaborative problem solving skills," in *Assessment and teaching of 21st century skills*: Springer, pp. 37–56.
- Hmelo-Silver, C. E. 2004. "Problem-based learning: What and how do students learn?" *Educational Psychology Review* (16:3), pp. 235–266.
- Hoey, E. M., and Kendrick, K. H. 2017. "Conversation analysis," *Research methods in psycholinguistics: A practical guide*, pp. 151–173.
- Hutchby, I., and Wooffitt, R. 2008. *Conversation analysis*: Polity.
- Jeong, H., and Hmelo-Silver, C. E. 2016. "Seven Affordances of Computer-Supported Collaborative Learning: How to Support Collaborative Learning? How Can Technologies Help?" *Educational Psychologist* (51:2), pp. 247–265.
- Jonassen, D. 2011. "Supporting Problem Solving in PBL," *Interdisciplinary Journal of Problem-Based Learning* (5:2), pp. 95–119.
- Kim, M. C., and Hannafin, M. J. 2011. "Scaffolding problem solving in technology-enhanced learning environments (TELEs): Bridging research and theory with practice," *Computers & Education* (56:2), pp. 403–417.
- Kulik, J. A., and Fletcher, J. D. 2016. "Effectiveness of Intelligent Tutoring Systems," *Review of educational research* (86:1), pp. 42–78.
- Limerick, H., Coyle, D., and Moore, J. W. 2014. "The experience of agency in human-computer interactions: a review," *Frontiers in human neuroscience* (8), p. 643.
- Litman, D. J., Rosé, C. P., Forbes-Riley, K., Vanlehn, K., Bhembé, D., and Silliman, S. 2004. "Spoken Versus Typed Human and Computer Dialogue Tutoring," in *Intelligent Tutoring Systems: 7th International Conference, ITS 2004, Maceió, Alagoas, Brazil, August 30 - September 3, 2004. Proceedings*, J. C. Lester, R. M. Vicari and F. Paraguaçu (eds.), Berlin, Heidelberg: Springer, pp. 368–379.
- Lucas, G. M., Gratch, J., King, A., and Morency, L.-P. 2014. "It's only a computer: Virtual humans increase willingness to disclose," *Computers in Human Behavior* (37:5), pp. 94–100.
- Mackey, A., and Gass, S. M. 2011. *Research methods in second language acquisition: A practical guide*: John Wiley & Sons.

- Maedche, A., Morana, S., Schacht, S., Werth, D., and Krumeich, J. 2016. "Advanced user assistance systems," *Business & Information Systems Engineering* (58:5), pp. 367–370.
- Mcalister, S., Ravenscroft, A., and Scanlon, E. 2004. "Combining interaction and context design to support collaborative argumentation using a tool for synchronous CMC," *Journal of Computer Assisted Learning* (20:3), pp. 194–204.
- Meier, A., Spada, H., and Rummel, N. 2007. "A rating scheme for assessing the quality of computer-supported collaboration processes," *International Journal of Computer-Supported Collaborative Learning* (2:1), pp. 63–86.
- Merrill, M. D., Drake, L., Lacy, M. J., Pratt, J., and Group, I. R. 1996. "Reclaiming instructional design," *Educational Technology* (36:5), pp. 5–7.
- Mills, K. L. (ed.) 2003. *Computer-supported cooperative work*, Citeseer.
- Nelson, L. M. 1999. "Collaborative problem solving," *Instructional design theories and models: A new paradigm of instructional theory* (2), pp. 241–267.
- Oxley, N. L., Dzindolet, M. T., and Paulus, P. B. 1996. "The effects of facilitators on the performance of brainstorming groups," *Journal of social behavior and personality* (11:4), pp. 633–646.
- Pais, S., Casal, J., Ponciano, R., and & Lourenço, S. 2015. "Unsupervised assistive and adaptive intelligent agent in smart environment," in *Intelligent Environments and System Conferences ICIES*, Paris, France.
- Pickard, M. D., Roster, C. A., and Chen, Y. 2016. "Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions?" *Computers in Human Behavior* (65), pp. 23–30.
- PISA 2017. *How does PISA measure students' ability to collaborate?*
- Porcheron, M., Fischer, J. E., McGregor, M., Brown, B., Luger, E., Candello, H., and O'Hara, K. 2017. "Talking with Conversational Agents in Collaborative Action," in *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17 Companion*, C. P. Lee, S. Poltrock, L. Barkhuus, M. Borges and W. Kellogg (eds.), Portland, Oregon, USA. 2/25/2017 - 3/1/2017, New York, New York, USA: ACM Press, pp. 431–436.
- Porcheron, M., Fischer, J. E., Reeves, S., and Sharples, S. 2018. "Voice Interfaces in Everyday Life," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, Montreal QC, Canada. 4/21/2018 - 4/26/2018, New York, New York, USA: ACM Press, pp. 1–12.
- Roschelle, J. 2013. "Special issue on CSCL: discussion," *Educational Psychologist* (48:1), pp. 67–70.
- Roschelle, J., and Teasley, S. D. 1995. "The Construction of Shared Knowledge in Collaborative Problem Solving," in *Computer Supported Collaborative Learning*, C. O'Malley (ed.), Berlin, Heidelberg: Springer, pp. 69–97.
- Rummel, N., and Spada, H. 2005. "Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings," *Journal of the Learning Sciences* (14:2), pp. 201–241.
- Sacks, H., Schegloff, E. A., and Jefferson, G. 1978. "A simplest systematics for the organization of turn taking for conversation," in *Studies in the organization of conversational interaction*: Elsevier, pp. 7–55.
- Sarker, S., Xiao, X., and Beaulieu, T. 2013. "Guest editorial: qualitative studies in information systems: a critical review and some guiding principles," *MIS quarterly* (37:4), pp. iii–xviii.
- Seeber, I., Bittner, E., Briggs, R. O., Vreede, G.-J. de, and Vreede, T. 2018. "Machines as Teammates: A Collaboration Research Agenda," in *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, USA.
- Setlock, L. D., Fussell, S. R., and Neuwirth, C. 2004. "Taking it out of context," in *Proceedings of the 2004 ACM conference on Computer supported cooperative work - CSCW '04*, J. Herbsleb and G. Olson (eds.), Chicago, Illinois, USA. 06.11.2004 - 10.11.2004, New York, New York, USA: ACM Press, p. 604.

- Statista 2019a. *US Voice Technology Usage by Age*. <https://www.statista.com/statistics/879907/us-voice-technology-usage-age/>. Accessed 1 May 2019.
- Statista 2019b. *Virtual Digital Assistant Market Worldwide*. <https://www.statista.com/statistics/589079/worldwide-virtual-digital-assistants-consumer-market/>. Accessed 1 May 2019.
- Tarouco, L. M. R., Silveira, C., and Krassmann, A. L. 2018. "Collaborative Learning with Virtual Entities," in *Learning and Collaboration Technologies. Design, Development and Technological Innovation: 5th International Conference, LCT 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings, Part I*, P. Zaphiris and A. Ioannou (eds.), Cham: Springer International Publishing, pp. 480–493.
- Team, R. C. 2013. *R: A language and environment for statistical computing*. <http://www.R-project.org/>. Accessed 1 May 2019.
- Tegos, S., Demetriadis, S., and Tsiatsos, T. 2014. "A Configurable Conversational Agent to Trigger Students' Productive Dialogue: A Pilot Study in the CALL Domain," *International Journal of Artificial Intelligence in Education* (24:1), pp. 62–91.
- Vail, A. K., Wiggins, J. B., Grafsgaard, J. F., Boyer, K. E., Wiebe, E. N., and Lester, J. C. 2016. "The Affective Impact of Tutor Questions: Predicting Frustration and Engagement," *International Educational Data Mining Society* (2:9), pp. 247–254.
- van Raaij, E. M., and Schepers, J. J. L. 2008. "The acceptance and use of a virtual learning environment in China," *Computers & Education* (50:3), pp. 838–852.
- Vanlehn, K. 2011. "The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems," *Educational Psychologist* (46:4), pp. 197–221.
- Venkatesh, V., Brown, S. A., and Bala, H. 2013. "Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems," *MIS quarterly*, pp. 21–54.
- Vtyurina, A., and Fourney, A. 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*, Montreal QC, Canada, pp. 1–7.
- Wik, P., and Hjalmarsson, A. 2009. "Embodied conversational agents in computer assisted language learning," *Speech Communication* (51:10), pp. 1024–1037.
- Winkler, R., Söllner, M., Neuweiler Maya Lisa, Rossini, F. C., and Leimeister, J. M. 2019. "Alexa, can you help us solve this problem? How conversations with smart personal assistant tutors increase task group outcomes," in *CHI'19 Conference on Human Factors in Computing Systems Extended Abstract*, CHI'19 Conference on Human Factors in Computing Systems Extended Abstract (ed.), Glasgow, Scotland, SIGCHI, pp. 1–6.
- Witte, E. H. 2007. "Toward a group facilitation technique for project teams," *Group Processes & Intergroup Relations* (10:3), pp. 299–309.
- Xolocotzin Eligio, U., Ainsworth, S. E., and Crook, C. K. 2012. "Emotion understanding and performance during computer-supported collaboration," *Computers in Human Behavior* (28:6), pp. 2046–2054.
- Yang, J. X., Unhelkar, V. V., Li, K., and Shah, J. A. 2017. "Evaluating effects of user experience and system transparency on trust in automation," in *Evaluating effects of user experience and system transparency on trust in automation*, Chicago, USA: IEEE.