Improving Problem-Solving Skills with Smart Personal Assistants: Insights from a Quasi Field Experiment

Completed Research Paper

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

Claudio Büchi
University of St. Gallen
Müller-Friedberg-Str. 8,
9000 St. Gallen
claudio.buechi@student.unisg.ch

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

Abstract

Problem-solving skills are considered one of the most important learning goals for life. Therefore, educational institutions should help learners gain these skills despite organizational and financial restrictions. Even though there exists a growing body of research about the design and use of Smart Personal Assistants, such as Google’s Assistant or Amazon’s Alexa, little is known about their ability to help learners improve their problem-solving skills. Using a mixed-method approach, we investigate the value of newly emerging Smart Personal Assistants to improve long-term problem solving skills with the help of a pre- and post-test quasi field experiment in a second grade class of a vocational business school in Switzerland. The results indicate that groups interacting with Smart Personal Assistants show significantly better problem-solving skills compared to learners using paper-based support explained by changing learning processes. Our study contributes to existing intelligent tutoring system and technology-enhanced scaffolding research.

Keywords: smart personal assistant, intelligent tutoring system, problem-solving skills, mixed-method, quasi field experiment

Introduction

In the digital age, organizations need employees who are able to deal with rapid changes and solve non-routine problems (OECD 2014). There is clear evidence of this change in the demand for skills in Germany, Japan and the United States (David et al. 2006; Ikenaga and Kambayashi 2010; Spitz-Oener 2006). One possible explanation for this shift to non-routine tasks in the workplace is that, as computers
Intelligent tutoring systems (ITS) have often tried to address the challenge of offering individual support prevalent (Oeste et al. 2015; Statista 2018). Classrooms at high schools, large-scale lectures at universities with more than 100 learners per lecturer, and massive open online courses (MOOCs) with more than 1,000 learners are becoming increasingly prevalent (Oeste et al. 2015; Statista 2018).

Intelligent tutoring systems (ITS) have often tried to address the challenge of offering individual support by leveraging the potentials of IT. There is a huge body of literature reaching about 30 years back in which IS and educational research investigated the impact of these systems on learning outcomes (Kulik and Fletcher 2016). When compared to no tutoring, intelligent tutoring systems showed that they are able to improve learning outcomes to the same extent as human tutoring (Graesser et al. 2018). Some studies also showed that ITS are able to improve learners’ problem-solving skills in selected domains (Hooshyar et al. 2016; Wang et al. 2015). Although there is general consensus about the effectiveness of ITS, they are still not integrated enough into learning environments (Nye et al. 2018). One reason for that might be that developing and introducing ITS often requires a high amount of technological knowledge and required time from the educator (Elaine 2015). Thus, they are mainly used by technology-savvy domains, such as computer science. For example, Hooshyar et al. (2016) and Wang et al. (2015) introduced an ITS in a computer science course where learners received help to conduct problem-solving activities. Furthermore, ITS applications are often complex to transfer from one context to the other and do not work on different devices and operating systems that learners use every day (Elaine 2015).

New emerging Smart Personal Assistants (SPAs) such as Google’s Assistant or Amazon’s Alexa have the potential to overcome the challenges faced by prior ITS. SPAs have become ubiquitous in modern life and are increasingly helping us perform everyday tasks in ways that they could not previously do (Davenport and Kirby, 2016). SPAs are software agents designed to support users to fulfill several daily actions (Pais et al. 2015). SPAs are running on ‘SPA-enabled devices” (endpoints) such as Apple’s iPhone, iPad and Mac, Amazon’s Echo, Google’s Home, etc. The main functionality, the “brain” of a SPA, is typically housed as a cloud service that processes voice data (converting voice-to-text, performing linguistic context analysis, and providing answers to questions, Chung et al. 2017). The popularity of SPAs, 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). 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, 59 percent of respondents from the ages of 18-24 stated that they were heavy (at least once per day) users of smart personal assistants (Statista 2019a). SPA providers such as Google, Amazon, and IBM offer large eco-systems that allow users to create their own skills without deep programming knowledge and time, thereby increasing SPA providers’ own business value. People are using SPAs on standalone devices (e.g., Googles’ Home or Amazons’ Echo), their phones (e.g., Apple’s Siri or Google’s Assistant), personal computers (e.g., Microsoft’s Cortana), and internet of things (IoT) devices (e.g., Samsung’s Smart TV). As SPAs are available on so many different devices, nearly everyone owns their own smart personal assistant. They are increasingly able to build up sophisticated interactions with their users due to rapid advancements in natural language processing. Because of this rapid developments, they are able to help users overcome more complex problems (Fast et al. 2018). For example, Vyurina and Fourny (2018) explored the interaction of a participant cooking a complex meal with the help of a SPA.

The few research that has already explored SPAs in learning environments was mainly concentrated on short-term task outcome in laboratory experiment settings or simple research designs without a control group (Arend 2018; Coronado et al. 2018; Kim et al. 2018). For example, Arend (2018) used Apple’s Siri in a secondary school class for a specific homework assignment without a control group to compare it with. Coronado et al. (2018) conducted a case study where they investigated learning satisfaction directly after the treatment. However, fostering skills such as problem-solving can hardly be achieved within a single lab experiment or homework assignment. Instead, it is long-term learning outcome that manifests after a certain period of time. Due to SPAs’ ubiquitous availability, they have the potential to change
learners’ behavior in order to develop their own problem-solving skills over time. Hence, in this paper we seek to answer the following two research questions:

**RQ1:** To what extent does using a Smart Personal Assistant help learners develop their problem-solving skills?

**RQ2:** How does using a Smart Personal Assistant affect the learning process of learners?

To answer our research questions, we employed a mixed-method approach, complementing the quantitative results of a pre- and post-test quasi field experiment (RQ1) with the qualitative data of a focus group discussion (RQ2). Mixed-method research can develop insights into new phenomena of interest that cannot be fully understood using only a quantitative and qualitative method. Currently, there is a dearth of mixed method research in IS (Venkatesh et al. 2013). We ground our work on a constructivist understanding of learning. Drawing on the concept of the ICAP-Framework and scaffolding, we design and introduce SPAs in a second grade class of a Swiss business vocational school over a period of five weeks. The experiment class used SPAs for their homework assignments whereas the control group received paper-based scaffolds. The findings suggest that the use of SPAs in everyday learning environments is an efficient way to develop learners’ long-term problem-solving skills. Thereby, our study contributes to two different research areas in IS. First, we contribute to existing ITS research showing that SPAs and their ecosystems are a promising way for future implementations of ITS, since they are widely available, and easy-to-use from an educators – non-programming expert’s – point of view. Thus, these two factors increase the likelihood of educators integrating an ITS in their learning environments, and the likelihood of learners using a certain ITS at different points in their learning process. Moreover, our work indicates that SPAs are able to change students’ learning behavior that is more aligned to a constructivist view of learning. Second, we contribute to technology-enhanced scaffolding theory by empirically proving the added value of dynamic scaffolds (SPA-enhanced) compared to static scaffolds (paper-based) for long-term skill development in an everyday learning environment.

The study also makes some practical contributions by exemplarily showing how educators can implement SPAs in their learning environment. As this kind of technology will increasingly enter learners’ lives in the next few years, educators need to know how they can use its potential as a learning resource. The remainder of this paper is structured as follows. The next section explains the theoretical background and hypothesis development. Section 3 describes the pre- and post-test quasi field experiment in more detail, including the task design, the design and functionality of the SPA, and the measurement and analysis of our constructs. Section 4 presents the results of research questions 1 and 2. Finally, we conclude with the discussion, the limitations and future research, and our conclusion.

**Theoretical Background and Hypothesis Development**

In this section, we first define and classify the term Smart Personal Assistant. Next, we elaborate on constructivism as our theoretical lens on the way how we perceive learning. Finally, we will explain the ICAP Framework and the concept of scaffolding for problem-solving activities that serves as a basis for our hypothesis development.

**Smart Personal Assistants as Learning Tutors**

SPAs are software agents that can automate and simplify many of the daily tasks of their users by mainly engaging with them via voice in- and output (Pais et al. 2015). User Assistance Systems that assist users in performing their tasks are not new (Maedche et al. 2016). However, compared to traditional user assistance systems, new emerging SPAs can be characterized with a rather high degree of interaction and intelligence (Maedche et al. 2016). SPAs have an agent program running on SPA-enabled devices (endpoints) such as Apple’s iPhone, iPad, and Mac, Amazon’s Echo or Google’s Home, etc. The main functionality, the “brain” of a SPA, is typically housed as a cloud service that use machine learning and natural language processing techniques to handle voice data (converting voice-to-text, performing linguistic context analysis, and providing answers to questions, Chung et al. 2017). SPA providers offer rich ecosystems with intuitive interfaces that allow a large amount of users to create their own skills without having deep programming knowledge, thereby increasing SPA providers’ own business value. SPAs can be divided into two types: (1) built-in SPAs that use multi-purpose devices and (2) stand-alone
SPAs that use dedicated devices. Examples of built-in SPAs include Siri (for Apple products) and Cortana (for Windows-based PCs). Examples of stand-alone SPAs include Alexa (that uses Echo, Echo Dot and Tab dedicated devices) and Google Assistant (that uses Google Home dedicated devices, Chung et al. 2017). In our study, we focus on both types of SPAs by enabling access via different devices. SPAs have the capacities to react to users’ utterances and can proactively guide them through a complex task (Pais et al. 2015). For example, Fast et al. (2018) introduced a SPA that supports users in conducting data science tasks (e.g., a predictive modeling task).

SPAs that are used to improve learning outcomes can be seen as a subcategory of intelligent tutoring systems (ITS). Researchers in the area of IS and education have often explored the effectiveness of ITS (Kulik and Fletcher 2016). Researchers pretty much agree that ITS are able to improve learning outcomes. More specific, some studies also indicate that ITS are able to improve learners’ problem-solving activities (Hooshyar et al. 2016). For example, Hooshyar et al. (2016) introduced an ITS to aid learners in learning computer programming. However, most documented instances of successful, interactive and dialogue-based tutors are implemented in technology-savvy domains, such as computer science (Hilles and Naser 2017). Most of the currently implemented ITS rely on rather complicated software architectures that require a lot of technical know-how and time effort from educators (generally 200 hours of development time for one hour teaching/instructions), especially when they want to use the tutor for different contexts (e.g., learning goals, learning content, Elaine 2015; Hilles and Naser 2017). Thus, many researchers describe the ITS development as notoriously costly (Barbhuiya et al. 2011). The issues of technological knowledge and required time from educators are two of the biggest challenges when trying to leverage the whole potential of these systems. On the contrary, SPA providers such as Amazon and Google offer powerful and easy-to-use toolkits to build up a multi-turn dialogue with their users. Few studies have specifically looked at SPAs in learning environments (Arend 2018; Kim et al. 2018). For example, Arend (2018) conducted a case-study where he introduced Apple’s Siri into an English Second Language classroom in a secondary school for a specific homework assignment. Coronado et al. (2018) developed a SPA for students learning programming language Java and showed that students were more satisfied using a SPA than a normal, static question and answer interface.

Constructivism as a Learning Paradigm

Constructivism as a paradigm or worldview states that learning is an active, constructive process. The learner is an information constructor. People actively construct or create their own subjective representations of objective reality rather than acquiring it (Glasersfeld 1987). Vygotsky’s social development theory is one of the foundations of constructivism (Vygotsky 1978). Social Development theory argues that consciousness and cognition are the end product of socialization and social behavior. Social interaction plays a fundamental role in the process of cognitive development. Within his theory, Vygotsky coins the term “more knowledgeable other”. A “more knowledgeable other” is anyone who has a better understanding or a higher ability level than the learner, with respect to a particular task, process, or concept. It is normally thought of as being a tutor or coach. Vygotsky’s theory promotes learning environments in which “more knowledgeable others” can trigger learners’ active role in learning. This means that educators should collaborate with their learners in order to help facilitate knowledge construction (Kim 2001). This works best when learners have their own individual tutor in which learning becomes a reciprocal experience for the learner and the teacher (Vygotsky 1978).

ICAP-Framework

The ICAP framework proposed by Chi and Wylie (2014) is based on a constructivist view of learning. It explains the process of effective learning by classifying observable learner behaviours into four modes: passive, active, constructive, interactive. It predicts that these modes will be ordered by effectiveness as follows: passive < active < constructive < interactive. Educators have long recognized that although learners can learn from receiving information passively, they learn much better by learning actively (Bonwell and Eison, 1991). Learning actively requires learners to engage cognitively and meaningfully with the tasks they are dealing with. They think about their learning material in depth rather than just passively receiving it (King, 1993). Each mode of the ICAP framework corresponds to different types of behaviours and knowledge-change processes predicting different learning outcomes. Table 1 outlines the
different modes (from passive to interactive) showing that for different learning scenarios (listening to a lecture, reading a text, and observing a video), learners show different modes of behavior.

| Table 1. Distinction of Learner Behavior according to ICAP Framework adapted from Chi and Wylie (2014). |
|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|
| **Listening to a lecture** | **Passive (Receiving)** | **Active (Manipulating)** | **Constructive (Generating)** | **Interactive (Dialoguing)** |
| Listening without doing anything else but oriented toward instruction | Repeating or rehearsing; Copying solution steps; Taking verbatim notes | Reflecting out loud; Drawing concept maps; Asking questions | Defending and arguing a position in dyads or small group |
| **Reading a text** | Reading entire text passages silently/aloud without doing anything else | Underlining or highlighting; Summarizing by copy-and-delete | Self-explaining; Integrating across texts; Taking notes in one's own words | Asking and answering comprehension questions with a partner |
| **Observing a video** | Watching the video without doing anything else | Manipulating the tape by pausing, playing, fast-forward, rewind | Explaining concepts in the video; Comparing and contrasting to prior knowledge or other materials | Debating with a peer about the justifications; Discussing similarities & differences |

A passive learner behavior describes learners who deal with the presented instructional information without additional physical activity (e.g., reading a text without doing anything else). An active learner behavior includes “doing something physically” (e.g., underlining or highlighting a text). A constructive learner behavior requires “producing outputs that contain ideas that go beyond the presented information” (e.g., taking notes in one’s own words). An interactive behavior requires “dialoguing extensively on the same topic, and not ignoring a partner’s contribution”. This means that both partners’ utterances are primarily constructive, and a sufficient degree of turn-taking occurs (e.g., asking and answering comprehension questions with a partner, Chi and Wylie, 2014). For example, Gobert and Clement (1999) compared learning gains among learners who studied learning materials by only reading a text (passive), writing summaries of the text (active), and drawing diagrams from the text (constructive).

As predicted by the ICAP hypothesis, the results showed that the constructive (diagram) group did better on measures of learning outcomes than the active (summary) group did, which in turn did better than the passive reading group. Properly designed SPAs as learning tutors have the capabilities to build up dialogues with students bringing them from a passive to an interactive learning mode (Chi and Wylie 2014).

**Scaffolding for Problem-Solving Activities**

Virtually everyone in their everyday as well as their professional lives solves problems on a regular basis. Gaining problem-solving skills can therefore be considered as the most important learning outcome for life (Jonassen 2000). Problem-solving skills can be defined as “... an individual’s capacity to engage in cognitive processing to understand and resolve problem situations where a method of solution is not immediately obvious (OECD 2014).” Kim and Hannafin (2011) state that learners generally pass through five problem-solving activities when solving a problem: (a) identification and engagement, (b) exploration, (c) reconstruction, (d) presentation and communication, (e) reflection and negotiation. In order to understand how learners can be supported in performing these problem-solving activities, Wood et al. (1976) first coined the term scaffolding. Scaffolding is defined as a process through which “more knowledgeable others” (e.g., teachers, peers, or tools) engage in an interaction with the learners to increase their problem-solving skills (Vygotsky 1978). This enables learners to achieve what is beyond their ability to accomplish independently. The interaction with a “more knowledgeable other” can help learners internalize the scaffolds and advance from their current knowledge to more sophisticated practices (Vygotsky 1978). While, in the beginning, more scaffolding support is provided, it is gradually decreased as learners become more capable and the locus of responsibility shifts to the learners (Azvedo and Hadwin 2005). For example, learners may have an incomplete definition of the problem and are not...
able to determine the goal of the task, the expected products, or the resources involved. Regulating the problem-solving process, determining which actions to take, when, and how, is often challenging for learners (Kim and Hannafin 2011). Interactive, technology-enhanced scaffolds hold the promise that learners are able to internalize these scaffolds, thereby gaining problem-solving skills (Ge and Er 2005). For example, Ge and Er (2005) were able to show that scaffolds helped learners to better apply problem-solving activities.

New emerging SPAs such as Google’s Assistant or Amazon’s Alexa have the capacities to become a “more knowledgeable other” for learners. They can build up a dialogue with learners by providing interactive scaffolds and react to learners’ utterances similar to a human tutor. SPAs have the capacities to bring learners into an interactive learning mode resulting in a deeper involvement with the learning material. This deeper involvement with the learning materials help to gain skills over time. Consequently, we propose:

H. Learners using Smart Personal Assistants will exhibit higher levels of problem-solving skills compared to learners using paper-based scaffolds.

Research Methodology

To answer our research questions, we chose a mixed-method approach and designed a quasi field experiment with a pre- and post-test control group design. We chose a mixed-method approach for two reasons. First, we wanted to obtain complementary views on the quantitative relations between SPA usage and problem-solving skills. This reflects Venkatesh’s first type of reasons to conduct a mixed-method approach (type 1: complementary, Venkatesh et al. 2013). Second, we compensated for the rather small sample size in the quantitative part of our study (type 6, compensation, Venkatesh et al. 2013). A pre- and post-test control group design is one of the most frequently used experimental designs and is particularly common in educational technology research (Dugard and Todman 1995). We conducted the quasi experiment with two second-grade school classes at a vocational business school in Switzerland and enriched the quantitative data with a post-experiment focus group discussion. The experiment ran over a period of five weeks, at the end of which we measured learners’ problem-solving skills in a final exam.

Background and Setting

We implemented the SPAs in a second class vocational business school in Switzerland. The choice of implementing the SPAs in these had several benefits. First, both classes were similar in terms of school grades, gender and learning goals. Second, both classes had the same teacher making the two classes suitable for a quasi experiment in an everyday learning environment. The school is located in the capital city of its region with approx. 500 learners. The relevant vocational apprenticeship lasts four years and ends with a finished apprenticeship and Abitur. The experiment class consists of 22 students and the comparison class of 23 learners. Both classes have the same teacher in the relevant subject. The relevant teacher in the two classes is very experienced, having taught for over 25 years in different school types (secondary school, high school, vocational school, etc.). He reported using technology with his learners daily (computer-based teaching, tablet-based teaching, etc.). The relevant subject was business and law. Both classes had 3 lessons a week, resulting in 15 lessons over the experiment period of 5 weeks. During the period of five weeks, the teacher dealt with the topic “introduction into law”. Both classes had exactly the same six learning goals (LG). The goals were to understand the necessity of law in everyday life (LG 1, Lesson 1), to explain differences between morality, custom and law (LG2, Lessons 2 to 4), to solve problems related to the freedom of opinion (LG3, Lessons 5 to 7), solve problems related to the freedom of religion (LG4, Lesson 8 to 10), solve problems related to the property guarantee (LG5, Lessons 11 to 13), and analyze in which cases fundamental rights have their limitations (LG6, Lessons 14 to 15). The teacher used the same methods in both classes, consisting of whole-group instruction, followed by individual and partner work. Both classes had no law-related subjects before. The experiment and control class had a final exam at the end of the five weeks, which formed a part of their overall grade in business and law. Between the end of week 1 and the end of week 4, the learners had to do four problem-based, 30-minute homework assignments for LG 3, 4, and 5. All the homework assignments had the same style as the pre- and post-test tasks and fulfilled the requirements of a problem-based task according to Jonassen (2000).
Sample

Overall, 45 learners participated in the quasi field experiment. The sample consists of two second grade classes of a vocational business school in Switzerland. The experiment class participated in the focus group discussion afterwards, which took place one week after the post-test. The average age of the experiment class was 17.4 with 12 males and 10 females. The average age of the control class was 17.2 with 12 males and 11 females. The experiment class had two persons who were not born in Switzerland and the control class had three persons (but all were able to speak German). Most of the vocational learners made their apprenticeship at large-sized banks and insurances (n=18), followed by large-sized merchandise trade companies (n=10), small and medium-sized companies (n=10) and others (n=7). We conducted ANOVA tests to make sure that the two groups are equal. The tests revealed that there was no significant difference in the background of the learners in terms of their gender (p = 0.142), apprenticeship sector (p = 0.623), school grades in business and law (p = 0.219), pre-experience with SPAs (p = 0.611), personal innovativeness (p = 0.766), and the pre-tests (p = 0.116).

Task Design

Closely connected with the learning goals and the content, both classes had to do four 30-minute homework assignments that were handed out at the end of week one (paper-based in both classes) and collected at the end of week four. The tasks of the assignments meet the requirements of problem task definitions proposed by Jonassen et al. (2000). Accordingly, problem tasks should be ill-structured, open-ended, realistic, and resonate with the executors’ experience. In Table 2, we exemplarily show homework assignment 2 and how it addresses the requirements. The tasks of the pre-test and post-test have a similar structure and are also aligned with the presented requirements.

<table>
<thead>
<tr>
<th>Task: Imagine the following scenario: The use of smartphones during the lesson is forbidden in your school. Nevertheless, Thomas K. (a classmate of yours) uses his smartphone during the lesson to tell his mother that he will be late that day. The teacher collects the smartphone and tells Thomas K. that he will keep the smartphone until the end of the week. How would you solve this problem with the help of the law?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Characteristics of a problem-task</strong></td>
</tr>
<tr>
<td>Ill-structured</td>
</tr>
<tr>
<td>Open-ended</td>
</tr>
<tr>
<td>Realistic</td>
</tr>
<tr>
<td>Resonate with the executors’ experience</td>
</tr>
</tbody>
</table>

Design and Functionality of the Smart Personal Assistant

We developed a SPA that is able to interact with the learners in the most natural way possible. Based on the ICAP-framework, building up a natural dialogue with the learners is a necessary condition to trigger learners’ interactive learning behavior. Therefore, we used Amazon’s Alexa SPA ecosystem (Amazon 2019). Specifically, we used Alexa’s Skill Development Kit 2.0 with nodeJS, since this framework seems to offer one of the most developed state-of-the-art capabilities regarding speech recognition and natural language processing. Moreover, the Alexa Skill Kit 2.0 offers an easy-to-use toolkit with a variety of blueprints and tutorials with the help of which we can use very basic coding patterns to create rich, interactive experiences (Amazon, 2019). For all the four homework assignments, we included two different logics in our interaction model. In our first logic, the SPA tutor proactively guided learners through the task by using five different problem-solving activities adapted from Kim and Hannafin’s (2011) problem-solving phases: (a) identification and engagement, (b) exploration, (c) reconstruction,
(d) presentation and communication, (e) reflection and negotiation. The SPA tutor provided its problem-solving steps in an interval of five minutes. In our second logic, the SPA was able to provide feedback on learners’ utterances and react to learners’ questions that occurred between the different problem-solving steps. Specifically, the SPA was able to react to three different types of questions. Question type 1 was the learners’ intent to skip to the next step, question type 2 was the learners’ intention to ask content-specific questions (concept clarifications, hints, etc.), and question type 3 was to ask the tutor to rephrase the current problem-solving step. This interaction model incorporated in the SPA helps learners to internalize the problem-solving steps while doing their homework assignments, thereby acquiring problem-solving skills. Figure 1 shows homework assignment 2, an exemplary excerpt of a learner dialogue and the corresponding design features.

**Task**

Imagine the following scenario: The use of smartphones during the lesson is forbidden in your school. Nevertheless, Thomas K. (a classmate of yours) uses his smartphone during the lesson to tell his mother that he will be late that day. The teacher collects the smartphone and tells Thomas K. that he will keep the smartphone until the end of the week. Thomas K. is upset about it. How would you solve this problem with the help of the law?

**Figure 1. Excerpt of an exemplary Dialogue between the Learner and SPA and corresponding Design Feature.**

Before introducing the SPA in class, we tested the SPA in the research team and asked the teacher to review our interaction model and the corresponding interaction dialogues. Moreover, we made pre-tests with the help of four learners. Finally, we collected the resulting change requests and refined the SPA (e.g., better on-boarding process, naming the goal of the SPA at the beginning, etc.). A key challenge when designing SPA materials for experimental treatments is the issue of informational equivalence. According to Larkind and Simon (1987), two representations are informationally equivalent if all the information from one representation can also be inferred from the other representation and vice versa. Therefore, we also added the scaffolds (hints to discover the right solution) to the paper-based materials of the control group.
Experimental Procedure

The experiment class (22 learners) used SPAs as learning tutors for their four homework assignments. The control class used paper-based scaffolds for their assignments. Both learning aids had the same scope of information and functions. Figure 2 depicts the study timeline. All learners completed a pre-experiment survey within a lesson before week 1 to help identify and control for pre-existing class differences (pre-experience with SPAs, personal innovativeness, apprenticeship company, gender, age). We also collected learners’ first semester school grade in business and law in the school’s learner information system for the same purpose.

Figure 2. SPA Experiment Implementation Timeline

The classroom teacher administered a 30-minute, 3-subtask pre-knowledge test to all learners in a lesson before the experiment period started. The goal of the pre-knowledge test was to identify learners’ problem-solving skills in the relevant subject. Moreover, the research team conducted a pre-experiment meeting with the teacher to discuss the experiment details with him and to ensure that learning goals, teaching methods, and tools stayed the same over the period of the experiment. At the beginning of week 1, we introduced the SPA devices (Amazon’s Alexa Echo Dot Device) in the experiment class and helped learners to install the accompanying Alexa software on their smartphones, tablets and laptops. At the end of week one, we checked if every learner had access to Alexa on one or more devices. After week 1, the learners received their four homework assignments in the form of a paper script. The experiment class was instructed to use Alexa on their preferred device (standalone device, smartphone, tablet, laptop) for completing their assignments. The comparison class was instructed to use their paper-based scaffolds with exactly the same amount of information as on Alexa. The students received the paper-based scaffolds in the form of a separate script so they could decide for themselves when to use the scaffolds. At the end of week 4, learners of both classes had to submit their homework assignments and after another week passing by, they conducted their 30-minute post-test and post-experiment survey. The teacher was instructed to not discuss the homework assignments before the post-test was conducted. After another week passing by, week 6, we conducted a 45-minute focus group discussion with the whole experiment class.

Measurement and Analysis

Quantitative Data

For measuring problem-solving skills (dependent variable), we constructed a 3-task pre- and post-experiment test with the help of the teacher. The tasks within the pre- and post-test had the same
structure as the homework assignments addressing the requirements of problem-based tasks proposed by Jonassen (Jonassen 2000). The pre- and post-test have the same total number of points and the tasks addressed learning goal 3, 4 and 5 (see Appendix A). All tasks allowed to apply Kim and Hannafin's (2011) problem-solving steps. The difficulty level of the post-test was evaluated as similar compared to the pre-test tasks by the teacher. We analyzed pre- and post-test results with three independent, experienced raters independently and blinded with a pre-defined and commonly discussed evaluation framework. One of the evaluators was the teacher of both classes. The evaluation framework allowed us to evaluate the application of the problem-solving steps by giving points to each well-applied problem-solving step within a task (e.g., 2 points for task 1 and step 1: problem identification, etc.; the points were given according to the difficulty and length of the step). The framework contained possible solutions for each of the steps and also allowed further sensemaking ideas. Overall, the learners were able to receive 12 points per task and 36 points at maximum. The evaluation framework clearly defined when a point had to be given or not. The final scores of the pre- and post-test results arose from an average of the individual appraisals. We checked for interrater correlation with the help of Pearson’s correlation coefficients (overall average of the pairwise interrater correlations = 0.9433, p < 0.05). The pre-survey consisted of pre-experience with SPA, personal innovativeness, apprenticeship company gender and age. For pre-experience with SPAs, we asked students how often they use SPAs (e.g. Apple’s Siri, Google’s Assistant or Amazon’s Alexa) in a week. For personal innovativeness, we used the four items from van Raaij and Schepers (2008). The post-survey included two open-ended questions. Question 1 was about the learners’ perceived helpfulness of the learning aid (SPA or paper-based scaffolds) and question 2 was treatment-specific, asking the experiment class about their experiences with SPAs as learning tutors. The items of the pre- and post-survey are depicted in Appendix B.

To analyze the quantitative data, we run a one-way analysis of covariance (ANCOVA), including the pre-test results as a covariate. ANCOVA helps to analyze variances between the groups and controlling for covariates. It is therefore suitable for pre- and post-test designs (Dimitrov and Rumrill 2003). We conduct a test for normality, homogeneity of variance and homogeneity of regression slopes to check if our data meet the assumptions for using an ANCOVA. Furthermore, we calculate Cohen’s d (1988) to show the effect size. He indicates that small, medium, and large effect size d values are 0.2, 0.5 and 0.8, respectively, meaning that a large effect size, for example, would indicate a difference between the means of interest of 0.8 standard deviations. Additionally, we calculate Cohens’ effect size f as input for a sample size test with G*Power 3.1.9.2 in order to check if our total sample size is large enough (Cohen 1988; Faul et al. 2007). We calculated 95% confidence intervals and used the statistic program R as a tool for analysis (Team 2013).

**Qualitative Data**

The focus group discussion was based on the answers of the open-ended questions from the post-survey. The objective of the focus group was to gain a more in-depth understanding of how SPAs affect learning processes. The focus group lasted 45 minutes with one of the researchers as facilitators. All learners from the experiment class participated. The focus group was structured as follows. First, the facilitator introduced the goal of the group discussion (learners' perceptions of using SPAs). Second, learners were asked to divide into workgroups of four to six people and were invited to discuss and negotiate opinions about how they used the SPA as learning tutor. For helping learners to structure their discussion, they received three broad areas identified in the post-survey: pros of using SPAs, cons of using SPAs and neutral observations while interacting with SPAs. Finally, a plenary discussion moderated by the facilitator encouraged further discussion and the gathering of additional content. We recorded the session, transcribed it, and used a thematic analysis to induce topics following the method of Ryan & Bernard (2003). Specifically, we used the keywords-in-context method for this study. With the help of this technique, we identified key words indicating some aspects of the learning process with SPAs and then systematically searched the corpus of text to find all instances of the word or phrase. Each time we found a word, we made a copy of it and its immediate context. We identified themes by physically sorting the examples into piles of similar meaning (Ryan and Bernard 2003). Moreover, we conducted a respondent validation by letting participants review our identified themes (Torrance 2012).
Results

RQ1: To what extent does using a Smart Personal Assistant help learners develop their problem-solving skills?

Table 3 shows a summary of the means of the pre-test and post-test results including the standard deviations.

<table>
<thead>
<tr>
<th></th>
<th>Experiment Class (SPA)</th>
<th>Control Class (paper-based scaffolds)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>pre-test</td>
<td>11.7</td>
<td>1.63</td>
</tr>
<tr>
<td>post-test</td>
<td>27.95</td>
<td>3.27</td>
</tr>
</tbody>
</table>

To check that the assumptions for the ANCOVA model are met, we conduct a test for normality, homogeneity of variance, and homogeneity of regression slopes by plotting the corresponding graphs (see Appendix C). The plots indicate that all assumptions for conducting an ANCOVA are met. Moreover, a G*Power test has shown that the sample size is large enough (minimum n = 25). We ran the ANCOVA with treatment group as independent variable and the pre-test score as a covariate. Results of the test indicate that there is a highly significant relation between SPA usage and problem-solving skills controlling for the pre-test, so we can accept H1 (F (2, 42) = 30.573, r² adjusted = 0.42, p = 1.88e-06, confidence intervals = 22.081 and 38.227, N = 45). Cohen’s d is 1.70 (confidence intervals = 1.002 and 2.401) indicating a difference between the means of 1.70 standard deviations. Since this effect is considered as high (Cohen 1988; Kotrlik and Williams 2003), we can conclude, that there are differences in using a SPA or using paper-based scaffolds in everyday learning environments. In addition, we also found significant results in all of the individual problem-solving steps between the experiment and the control group. According to Alexa’s analytics module, students used the SPA (invoked the skill) on average eight times during the experiment period and asked the SPA for help twice on average for each use (invoked the help intent).

RQ2: How does using a Smart Personal Assistant affect the learning process of learners?

Based on our quantitative findings regarding the relationship of SPA usage and problem-solving skills, we investigate how SPAs affect learning processes. The three main themes we identified from the focus group data were interaction, usage behavior, and individualization. They are presented next and discussed in the discussion section.

**Interaction.** This theme relates to the capacities of SPAs being able to build up a dialogue with the learners. Several learners mentioned that using the SPA while completing their assignments feels like having an interaction with a peer or tutor. Learners appreciated that the SPA listened to them and gave appropriate answers. Some learners mentioned that receiving challenging questions from the SPA helped them to think of the next solution steps. Some other learners perceived it as more entertaining compared to “business as usual” paper-based learning materials. For example, one learner commented: “I liked her [Alexa]. It felt like I was talking to a teacher. She responded immediately and also asked me challenging questions that helped me with the next steps.” Moreover, several learners mentioned that they liked the way how the SPA was helping them. Specifically, they said that they appreciated thinking on their own about the solution first and that they can control when they want to receive help. For example, one student commented: “It was nice that Alexa was waiting until I asked her for help. That helped me to first think on my own and only receive hints whenever I want.” Moreover, some learners also mentioned that they liked saying their solutions out loud. While hearing themselves talking, they came up with new ideas more easily. For example, one learner commented: “I liked speaking the answers out loud and not writing them down. When I am hearing myself, I get new ideas.”

**Usage Behavior.** This theme relates to the (different) way how learners use the SPAs. Most of the learners used the SPA on their smartphone rather than their standalone device and mentioned that it was really easy and funny to access Alexa similar to Apple’s Siri or Google’s Assistant. For example, one learner commented: “It was like speaking with Siri. Just like having your personal tutor always in your pocket every time and anywhere you want.” The easy access to the learning tutor on smartphones and also
other devices (e.g. Amazon’s standalone device Amazon Echo Dot) led to a great variety of different learning places. Some learners indicated that they did their tasks in other places than usual. For example, one learner commented: “When I was lying on my couch, I talked to Alexa a few times, too.” However, few learners also mentioned some areas where SPAs are not very functional. For example, when learners were in public, they seldom used Alexa, because they felt uncomfortable talking to a SPA in public. For example, one student commented: “First, I wanted to conduct the tasks in the train. But then I decided to do something else, speaking with my smartphone in front of others felt weird to me”.

**Individualization.** The third theme we identified from the focus group data relates to the capacities of SPAs to be able to recognize learners’ individual characteristics. Some learners mentioned that they liked that Alexa was able to adapt her answers to learners’ utterances and prompted them individually until learners were able to find the own solution. For example, one learner commented: “It was helpful that Alexa recognized how she can help me discovering my own solution.” Moreover, some learners mentioned that they like that the SPA remembered the current status of progress whenever learners have interrupted the task. However, some other learners wished that the SPA recognized where learners have their biggest weaknesses. For example, one learner commented: “It would be great if she [Alexa] can remember our mistakes and then concentrates on helping us with that.”

**Discussion**

Using a mixed-method approach, this study evaluate the use of new emerging SPAs in everyday learning environments to increase learners’ long term problem-solving skills. The findings support the use of SPAs to enrich everyday learning environments and also offer insights into how SPAs are able to change learning processes.

The quantitative results from our quasi field experiment proof that SPAs are able to increase learners’ problem-solving skills. This is in accordance with earlier work that was able to indicate that SPAs can support problem-solving activities (Hooshyar et al. 2016; Wang et al. 2015). One main reason for the difference between using SPAs for assignments and using paper-based scaffolds might be that SPAs are able to trigger learners’ interactive behavior. According to the ICAP-framework, an interactive learning behavior leads to better learning outcomes. This is also reflected by our qualitative findings indicating that learners perceive the interaction with SPAs as similar to human tutors (theme 1). Interestingly, the learners mentioned that they liked how the SPA scaffolded their understanding until they found the answer on their own. This is a big difference to “business-as-usual” paper-based scaffolds, where solution steps are static and right answers are often given. It seems that they enjoyed having their personal tutor on their side and only activate it when they cannot get any further. Thereby, the SPA supported them in constructing knowledge on their own. This helped them to better internalize the scaffolds and gain their own problem-solving skills. Hence, learners might perceive SPAs as what Vygotsky (1978) calls “more knowledgeable others”.

One other reason for the quantitative relationship between using SPAs and learners’ problem-solving skills might be that learners change their learning behavior (theme 2). In particular, learners change their learning places and learning times. This is also confirmed by Taylor (2006) stating that using smartphones to learn change learning contexts, for instance in terms of ergonomics (user posture, lighting, background noise), social context, and demands of users’ attention. Learners study things that interest them, at times that suit themselves, with little or no concern for consistency (Taylor 2006). As indicated by our quantitative findings, the different learning behavior might lead to higher amounts of learning time and finally better skill-development. In contrast to that, learners also mentioned that they feel weird using SPAs in public. This goes in line with the findings of Moorthy and Vu (2015). There results showed that participants preferred using SPAs in a private location (e.g. their home) explained by social acceptability. One other reason for the superiority of the SPA compared to paper-based scaffolds might be that SPAs are able to provide individual support. In specific, learners highlighted that the SPAs allowed them to learn on their own pace receiving help whenever they want it. Learning on their own pace motivates learners and gives them the feeling to work on their own academic progress (Chen 2008). These effects of individual support are also confirmed by several research papers in the area of personalized learning (Hwang et al. 2010; Song et al. 2012). Moreover, the learners mentioned that they would have wished that SPAs are able to detect individual learners’ characteristics. In specific, they mentioned that SPAs should be able to detect mistakes and use this information to repeat content in the end. This
Improving Problem-Solving Skills with Smart Personal Assistants

confirms the research effort in the area of intelligent tutoring systems research trying to capture learners’
characteristics and adapt ITS accordingly. For example, Ammar et al. (2010) designed an ITS equipped
with emotional management capabilities, which makes the capture of learner’s emotions possible during
learning and responds accordingly. Moreover, Mao et al. (2010) identified that when an intelligent tutor
system recognizes learners’ affective state and responds accordingly, the tutor is able to motivate learners
and improve the learning process.

Our work makes several theoretical contributions in the field of dialogue-based information systems.
First, we contribute to ITS research by proving that a new kind of ITS (SPAs) is able to foster learners’
long-term skill development compared to paper-based learning materials. The quasi field experiment pre-
and post-test control group design helped us to directly compare the usefulness of SPAs in an everyday
learning environment. Although past researcher has already used SPAs in learning environments, to the
best of our knowledge, there is no research of this kind. Moreover, as SPA providers offer intuitive
interfaces to build skills, we showed that these new kinds of SPAs are a promising way for future
implementations of ITS by bridging the gap between the technological know-how of software designers
and the pedagogical/didactical know-how of educators. Second, based on the ICAP-framework, we
contribute to technology-enhanced scaffolding research by empirically proofing the value of interactive
scaffolds (provided by the SPA) over static scaffolds (paper-based) for long-term skill development in an
everyday learning environment. Our work indicates that interactive scaffolds might be able to trigger
learners’ interactive learning behavior resulting in a better skill development. Moreover, our work
indicates that adapting scaffolds to individual students’ characteristics (e.g., learning pace, current
progress, knowledge gap, etc.) is beneficial for long-term skill development.

In regard to its practical implications, this study showed that effective individual tutoring provided by
SPAs can be implemented with relatively low technological knowledge and low costs. SPA developers,
such as Amazon, Google, Apple, etc., offer sophisticated development tools and interfaces where
educators can create their own SPA with basic technological expertise. As illustrated by our study,
Amazon’s Alexa ecosystem could be an effective platform to build SPAs for learning purposes. Moreover,
learners can bring their own devices (smartphone, laptop) with them where the SPA software can be
installed on. Finally, our study exemplarily showed educators how to integrate SPAs in an existing
“business as usual” learning environment.

Limitations and Future Research

There are a number of limitations to this study that should be noted. First, we used a quasi field
experiment design to examine the impact of SPAs on learners’ problem-solving skills. As in all these kind
of research designs, pre-treatment group differences may confound post-treatment outcomes (Campbell
and Cook 1979). We tried to address this point with collecting pre-experiment data and used ANOVAs to
check if both groups are equal. SPA group participants were similar on measures of demographics,
personal innovativeness, SPA experience, school grades and pre-knowledge. Moreover, both classes had
the same teacher with the same learning goals and using the same teaching methods. Nevertheless, it
would be interesting to see if experimental research is able to confirm and extend results of our identified
effects. Second, the sample size (n=45, 2 classes) can be considered as rather small for using an ANCOVA.
Nevertheless, our statistics showed a highly significant effect between these two groups. Additionally, we
conducted several tests (test for normality, test for homogeneity of variance, test for homogeneity of
regression slopes) to meet the assumptions for an ANCOVA. Given the large effect size (f = 0.8), the
G*Power test showed that the sample size is large enough. Furthermore, we applied a mixed-method
approach to partially compensate for the rather small sample size (Venkatesh et al. 2013). It would be
interesting to see if studies with larger sample sizes are able to confirm our effect. Third, we conducted the
SPA experiment over a relatively brief period under positive conditions. Such studies tend to produce
better outcomes due to novelty effects and hyperattention to experiment details (Cheung and Slavin
2013). For the current study, novelty effects may be diminished, given the high percentage of participants
reporting the high usage of SPAs on their smartphones (approx. 40% used SPAs everyday on their
smartphones, e.g. Siri). Fourth, this study also raises some ethical questions about data security and the
potential benefit of one group and not the other. Learners put their information on Amazon’s ecosystem,
which might be a threat to their personal security. While it is hard to control learner behavior, we
educated learners about the experiment and the corresponding risks at the beginning of the study. Fourth,
this study focused on problem-solving skills only. It would be very interesting to see if SPAs are also useful for other skills and knowledge types (e.g. lower and higher-order skills). Moreover, future research should take a look at different domains. SPAs might be better suited to fields, such as programming, where possible answers are more predictable. Finally, this study did not, in effect, evaluate how learning occurred inside class. For example, the teacher’s role during in-class discussions was not a study focus. For future research, it would be interesting to see how SPAs can be implemented in-class to improve students’ learning. Moreover, future research should focus on different levels of education and types of skills to further investigate in which use scenarios SPAs can offer an added value. Fifth, we compared the SPA group with paper-based scaffolds. This way, it is difficult to ascertain that the differences in learning outcomes are only because of the SPA. The design would have been better if we had used soft copies. However, we wanted to compare SPAs with the most common type of learning aids. Finally, research regarding a more fine-granular level of learner-SPA interaction would be desirable to find out why learning processes can be improved. Possible research questions include how SPA usage can trigger learners’ interactive behavior, how SPA usage influence learners’ emotions, etc.

Conclusion

Our work answered the question whether the use of SPAs could increase learners’ long-term problem-solving skills within a second grade vocational business school and further identified learners’ perceptions about the use of SPAs in an everyday learning environment. We designed and introduced SPAs over a period of five weeks and evaluated learners’ long-term problem-solving skills with the help of a pre- and post-test control group design. One group used paper-based scaffolds, whereas the other group used SPA-based scaffolds. The findings of this study demonstrate that the use of SPAs helped the learners in gaining long-term problem-solving skills and that using SPAs affect learning processes. These findings contribute to ITS and technology-enhanced scaffolding research. While further research is required, it appears that the use of new emerging SPAs are a promising enrichment of everyday learning environments in order to improve long-term skill development.

References


Appendix A

### Table 4. Pretest- and Post-test Tasks

<table>
<thead>
<tr>
<th>Learning goal</th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>LG3: solve problems related to the freedom of opinion</td>
<td>Imagine that in your country, political parties issue posters that are obviously against foreigners. Some foreigners complain against that. How would you solve this problem?</td>
<td>Imagine that in your country, Michael P, a good friend of you, issues flyers that are obviously against one of the left-wing politician in the country. Some people complain against that. How would you solve this problem?</td>
</tr>
<tr>
<td>LG4: solve problems related to the freedom of religion</td>
<td>Imagine that in your country, women are not allowed to wear burqas in public. Adem and Merve, two Islamic women, do not care about this rule and go out with their burqas. They got catched by the police and have to pay a fee now. They are complaining about it. How would you solve this problem?</td>
<td>Imagine that Muslims in your country have built a mosques in your neighborhood. After a while, a neighbor gets excited and criticizes the construction of the mosque. How would you solve the problem?</td>
</tr>
<tr>
<td>LG5: solve problems related to property guarantee</td>
<td>Imagine the state you are living in want to build a road on your land. If you do not want that, you will be expropriated, against which you are filing a complaint. How would you solve this problem?</td>
<td>Imagine that the city you are living in want to create a one-week lasting food festival and, thus, needs two of your land plots, against which you are filing a complaint. How would you solve this problem?</td>
</tr>
</tbody>
</table>

Appendix B

### Table 5. Pre-survey and Post-survey Items

<table>
<thead>
<tr>
<th>Pre-survey</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-experience with SPAs</td>
<td>Have you ever used a Smart Personal Assistant (e.g. Amazon’s Alexa, Google’s Assistant, Apple’s Siri?</td>
</tr>
<tr>
<td>If yes, how often do you use a Smart Personal Assistant per week on average?</td>
<td></td>
</tr>
<tr>
<td>Personal innovativeness</td>
<td>If I heard about a new information technology, I would look for ways to experiment with it. Among my peers, I am usually the first to try out new information technologies. In general, I am hesitant to try out new information technologies (reverse-scored). I like to experiment with new information technologies.</td>
</tr>
<tr>
<td>Demographics</td>
<td>Age, Gender, Apprenticeship company</td>
</tr>
<tr>
<td>Post-survey</td>
<td>Question 1 To what extent do you feel that the learning aid has helped you? Why?</td>
</tr>
<tr>
<td>Question 2 (treatment-specific)</td>
<td>What are your experiences when using the Smart Personal Assistant as a tutor?</td>
</tr>
</tbody>
</table>

Appendix C

<table>
<thead>
<tr>
<th>Test for normality</th>
<th>Test for homogeneity of variance</th>
<th>Test for homogeneity of regression slopes</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Normal Q-Q Plot" /></td>
<td><img src="image" alt="Residuals vs Fitted" /></td>
<td><img src="image" alt="Homogeneity of Regression Slopes" /></td>
</tr>
</tbody>
</table>

Figure 3. Check for Assumptions of ANCOVA