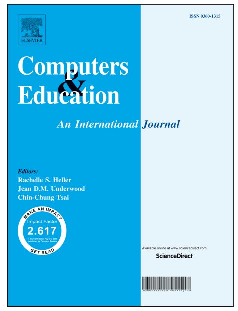
Please quote as: Winkler, Rainer; Söllner, Matthias & Leimeister, Jan Marco (2021) Enhancing Problem-Solving Skills with Smart Personal Assistant Technology. Computers & Education, 165 104148. ISSN 0360-1315

Journal Pre-proof



Enhancing Problem-Solving Skills with Smart Personal Assistant Technology

Rainer Winkler, Matthias Söllner, Jan Marco Leimeister

PII: S0360-1315(21)00025-7

DOI: https://doi.org/10.1016/j.compedu.2021.104148

Reference: CAE 104148

To appear in: *Computers & Education*

Received Date: 18 March 2020

Revised Date: 24 August 2020

Accepted Date: 22 January 2021

Please cite this article as: Winkler R., Söllner M. & Leimeister J.M., Enhancing Problem-Solving Skills with Smart Personal Assistant Technology, *Computers & Education*, https://doi.org/10.1016/j.compedu.2021.104148.

Title: Enhancing Problem-Solving Skills with Smart Personal Assistant Technology

Author names and credit statement (Given name and family name) and affiliations:

Author 1:

Rainer Winkler (corresponding author)

Institute of Information Management

University of St. Gallen

Müller-Friedberg-Str. 8

9000 St. Gallen

rainer.winkler@unisg.ch

Credit:

Conceptualization, Methodology, Software, Validation, Writing – Original Draft, Writing-Review & Editing, Visualization

Author 2:

Matthias Söllner

Information Systems and Systems Engineering

University of Kassel

Henschelstr. 4

34127 Kassel

soellner@uni-kassel.de

Credit:

Conceptualization, Methodology, Data Curation, Writing – Review & Editing, Supervision

Author 3:

Jan Marco Leimeister

Institute of Information Management

University of St. Gallen

Müller-Friedberg-Str. 8

9000 St. Gallen

janmarco.leimeister@unisg.ch

Credit:

Conceptualization, Methodology, Writing – Review & Editing, Supervision

Enhancing Problem-Solving Skills with Smart Personal Assistant Technology

Research Paper

Abstract

Smart Personal Assistants (SPAs; such as Amazon's Alexa or Google's Assistant) let users interact with computers in a more natural and sophisticated way that was not possible before. Although there exists an increasing amount of research of SPA technology in education, empirical evidence of its ability to offer dynamic scaffolding to enhance students problem-solving skills is still scarce. To fill this gap, the aim of this paper is to find out whether interactions with scaffolding-based SPA technology enable students to internalize and apply problem-solving steps on their own in a 10th grade high school and a vocational business school class. Students in the experiment classes completed their assignments using Smart Personal Assistants, whereas students in the control classes completed the same assignments using traditional methods. This study used a mixed-method approach consisting of two field quasi-experiments and one post-experiment focus group discussion. The empirical results revealed that students in the experiment classes acquired significantly more problem-solving skills than those in the control classes (Study 1: $p = 0.0396$, study 2: $p < 0.001$), and also uncovered several changes in students' learning processes. The findings provide first empirical evidence for the value of using SPA technology on skill development in general, and on problem-solving skill development in particular.

Keywords: Smart Personal Assistant Technology, Problem-solving skills, Scaffolding, Technology-mediated learning, Mixed-method, Field quasi-experiment.

1 Introduction

Today's organizations increasingly need employees who are able to deal with fast-changing environments and solve non-routine problems (OECD 2014). There is clear evidence of this shift in the skills requested in several countries, such as Germany and the United States (David et al. 2006; Spitz-Oener 2006). According to predominant constructivist learning theories, students need individual interaction with a personal tutor to best learn these skills (Vygotsky 1978). However, educational institutions such as high schools, vocational schools, and universities struggle to offer this kind of individual support due to financial and organizational constraints (Rietsche et al. 2018). The growing number of classroom sizes in high schools and vocational schools, mass lectures at universities with more than 100 students per lecturer, and massive open online courses (MOOCs) with more than 1,000 participants make individual interaction with a teacher or tutor even more difficult (Oeste et al. 2015). The tension between increasing student-educator ratios and the need for individual interaction raises the question of how to offer individual support to students to enable them to gain the necessary problem-solving skills.

Research in the area of technology-mediated learning has tried to address this challenge by leveraging the potentials of IT. In specific, there is a huge body of literature reaching back over 40 years in which educational researchers have examined the impact of computer tutoring systems on learning outcomes (Kulik and Fletcher 2016). In the last five to ten years, a lot of researchers were able to show that step-based computer tutoring systems were nearly as effective as average human tutors and more effective compared to no tutoring condition and teacher-led classroom instruction (Ma et al. 2014; Nesbit et al. 2014; Vanlehn 2011). “Step-based” means that the tutor evaluates and reacts to each step that the student makes in a problem-solving session as opposed to evaluating only the final answer given by the student. Many of these systems use some kind of scaffolding strategy to support students’ learning processes. Scaffolding strategies describe the way learning activities are broken up into smaller pieces and how tools and structures help students gain more knowledge (Kim and Hannafin 2011). Saye and Brush (2002) distinguish between static and dynamic scaffolds. Static scaffolds are static supports that can be anticipated and planned in advance based on typical student difficulties with a task. In contrast to static scaffolds, computer tutoring systems try to use dynamic scaffolds which are situational and requires the system to continuously diagnose the understandings of students and provide timely support based on student responses.

Despite the proven effectiveness of scaffolding-based computer tutoring systems in the past (Graesser et al. 2018), these systems have struggled to reach widescale adoption in learning environments (Nye 2014). Many of the existing computer tutoring implementations rely on rather complicated software architectures that require a lot of technical know-how, time, and effort to build and maintain these systems (Afzal et al. 2019). This is why they are mostly used in technology-savvy domains, such as computer science courses (Hooshyar et al. 2016). A new emerging consumer technology called Smart Personal Assistants (SPAs, e.g., Amazon’s Alexa, Google’s

Assistant or Apple's Siri), can be an important next step for providing dynamic scaffolds to support students learning process across different domains, because these systems have a very high degree of interactivity and intelligence and can be developed without much technical know-how, time and effort. By 2022, it's predicted that SPAs will run on 870 million devices, which will significantly increase the exposure of many people to this technology (Perez 2017). Experts anticipate that artificial intelligence (AI) in education will grow by 43% in the period 2018-2022 (Educause 2019). An SPA is an AI application that uses inputs – such as the user's voice, vision (images), and contextual information – to provide assistance by answering questions in natural language, making recommendations, and performing actions (Hauswald et al. 2015). SPAs include an agent program that runs on SPA-enabled devices (endpoints) – such as Apple's iPhone, iPad, and Mac, or Amazon's Echo or Google's Home. In contrast to often-used computer tutoring systems, the main functionality of SPAs is typically "black-boxed" and is housed as a cloud service of big tech companies to handle incoming voice or text data and produce outputs (Chung et al. 2017).

One of the AI software application categories in education that Luckin et al. (2016) describe is intelligent tutoring systems. Intelligent tutoring systems (ITS) can be used to simulate one-to-one personal tutoring. Based on learner models, algorithms, and neural networks, they can make decisions about the learning path of an individual student, select the content, provide scaffolding, and help to engage the student in a dialogue to improve students' knowledge and skill development. SPAs can be seen as a new subtype of ITS that calls for fresh research for studying their influence (Terzopoulos and Satratzemi 2019).

Compared to other subtypes of ITS, the novelty of SPAs lies in two major aspects which potentially fundamentally affect the way how they can offer scaffolds and support students: the underlying technology of SPAs as well as the creation and easy accessibility of SPAs for students and educators. In the past, ITS typically relied on rigid behavioral patterns, where action-conditioned, rule-based, bayesian networks and data mining were the most frequent AI techniques applied (Mousavinasab et al. 2018). Most of the ITS were only able to respond to simple requests by matching the user input against a set of stored patterns (McTear et al. 2016). There was no or little machine learning techniques incorporated in the system. However, based on advances in natural language understanding, natural language processing, and machine learning techniques such as deep learning, as well as higher computing power located on clouds, modern systems are now better able at detecting and classifying user intents. Nowadays, SPAs use mainly deep learning algorithms which help to model a large number of domains. These algorithms help SPAs to better recognize students' utterances without the need of self-generated training data, which other types of ITS often still need (Kloos et al. 2020). One example is Amazon's Alexa, which supports users to carry out everyday tasks via an advanced voice user interface, ultimately acting as their personal assistant. The natural language understanding and processing happens on a cloud hosted by Amazon and all students need is an App or standalone device such as Amazon's Alexa Echo Dot. These technical advancements enable the systems to offer dynamic scaffoldings to students in a more natural and sophisticated way. SPAs can answer students' questions via voice and immediately, so they can act as a private coach.

The second major aspect that differentiates SPAs from other types of ITS is their accessibility. SPAs are increasingly integrated in devices that students use every day (such as Google's assistant, Apple's Siri on smartphones, and Microsoft's Cortana on desktop PCs) and accessible mostly via one click or voice command – which is why they have the potential to become daily companions in both their private and school lives. Furthermore, graphical user interfaces enable students and educators to easily create their own SPAs without having deep programming knowledge or using authoring tools specially developed for this purpose, in contrast to other ITS subtypes. SPAs can be differentiated into two types: (1) built-in SPAs that are included in multi-purpose devices (e.g., Siri for Apple products and Cortana for Windows-based PCs); and (2) stand-alone SPAs that are included in dedicated devices (e.g., Alexa that uses Echo, Echo Dot, and Tab dedicated devices). In our study, we implement both types of SPAs by enabling access via different devices.

Given that the design and use of SPAs in education is an emerging field, most extant studies are explorative and only a few empirical studies exist that rigorously measure the effect of SPAs on learning performance. Canbek and Mutlu (2016) examined the potential effect of SPAs in learning environments by conceptually investigating potential use cases of different SPAs, such as Amazon's Alexa and Apple's Siri. They found out that one major benefit of SPAs is improving students' listening and speaking skills without needing human tutors. Dousay and Hall (2018) observed how teachers and administrators perceive the implementation of 90 Amazon Alexa Echo Dots in four school districts in the US with approximately 900 students. Teachers and administrators used Alexa to set reminders for activities, events, student dismissals, student medications, etc. They found out that one of the biggest challenges of implementing SPAs was the training of teachers and administrators. Jean-Charles (2018) focused on the perception of teachers regarding using SPAs in their learning environments. In specific, he asked pre-service teachers about their perspective on using Google's Assistant in the classroom. One of the biggest concerns of the teachers was to keep the information secure while allowing the SPA to gather the data to perform its functions. Moreover, in agreement with Dousay and Hall (2018), Jean-Charles (2018) found out that most of the teachers did not feel prepared to use this technology in a meaningful way. Arend (2018) implemented an SPA for a specific learning environment. She implemented Apple's Siri during a literacy activity in a third-year college English class. She used a conversation analysis approach and found out that one of the big advantages of interacting with Siri was that Siri triggered students to make their thoughts explicit, which helped them to structure their thinking processes. There are also a few studies that used SPA-like systems to support problem-solving skills. For example, Mavrikis et al. (2014) used a speech-based ITS, which helped students to think aloud significantly more than in past interactions and helped them to rephrase their language to employ mathematical language. Pai et al. (2020) created a text-based SPA that helped students solve mathematical problems using the method of remedial instruction and found increased motivation levels. Moreover, Damacharla et al. (2019) used Amazon's Alexa for emergency care provider training to help avoid medical errors; they found that it resulted in more efficient and effective training.

In summary, we see that the majority of existing empirical research in the area of SPAs and education is rather explorative (e.g., observations, qualitative interviews with students and educators) and few empirical studies that rigorously measure the effectiveness of SPAs in learning environments exist (Dousay and Hall 2018; Jean-Charles 2018). In particular, empirical studies are lacking when it comes to longitudinal effects, e.g., whether SPAs can help to train certain skills over the course of a longer period of time. This is crucial, since we know from prior research that the development of skills takes time and is difficult to achieve within a single short experiment (Soderstrom and Bjork 2015). To contribute to a better understanding of the usefulness of SPA technology in education, this paper seeks to answer the following two research questions:

RQ1: Does using a Smart Personal Assistant help students develop their problem-solving skills?

RQ2: How does using Smart Personal Assistant technology influence the learning process of students?

The paper is grounded within the constructivist learning paradigm (Glasersfeld 1987), which describes learning as a change in meaning constructed from experience. In constructivist learning environments, educators shift from the role of the knowledge provider to the role of the personal coach for learning. With the SPAs developed in this paper, we too adopt this perspective of the individual coach. Two characteristics seem to be central for the constructivist learning process. First, constructivist instruction asks students to use their knowledge to solve problems that are meaningful and realistically complex. Second, it only offers scaffolding when students require it rather than providing one-size-fits-all solutions (Ertmer and Newby 1993). Additionally, the ICAP Framework proposed by Chi and Wylie (2014) states that an interactive learning behavior of a the students is most effective. The SPAs developed in this study build on these insights by helping students to apply their knowledge (application of problem-solving steps) and by offering help through dynamic scaffolding only when students require it in order to bring students into an interactive learning behavior.

2 Method

To answer our research questions, we created a quasi-experimental design, where we complement the quantitative results of two field quasi-experiments (RQ1) with the qualitative insights of a post-experiment focus-group discussion (RQ2). Quasi-experimental designs in education are adopted in cases where random formation of groups is not possible and therefore experimental and control groups are formed with already-existing classes (Fraenkel et al. 2011). In two different schools, we conducted pre-tests to check whether classes within each school are similar in terms of initial problem-solving skills, gender, age, school grades in the relevant subject, personal innovativeness, and pre-experience with SPAs. The pre-test scores showed that within each school, the two classes were similar. Thus, we were able to randomly choose one to be the experimental class that follows an intervention program in each school. Detailed pre-test results and student characteristics can be seen in Appendix G.

2.1 Background and Setting

We selected classes from a second-year high school (October/November 2018) and a second-year vocational business school (January/February 2019) for the experimental and control groups. The choice of implementing SPAs in these learning environments has two main benefits. First, in each school, the experiment class and control class were similar. Moreover, in each school, the two classes had the same teacher and the same learning content, making the two classes suitable for a field quasi-experiment approach. Second, choosing two different school types allowed us to widen the scope of our study and strengthen our results. In a high school, school performances are usually better as students aim to prepare themselves for university. In a vocational business school, school performances are generally a bit lower as students tend to start to work after school. Both schools we chose are located in the capital city of its region. The high school had approximately 1,300 students in total and eleven second-year classes and the vocational business school had 500 students in total and five second-year classes. Both experiment and control classes took courses on the "Introduction into law", which is a unit in the business and law curriculum. We chose this unit for two reasons. First of all, it is the first occasion that students come into contact with law, therefore students have no pre-knowledge and the classes are comparable. Second, the unit includes the solving of legal problems and is therefore very suitable for the acquisition of problem-solving skills (the focus of the study).

The experimental groups completed the module with SPA technology, while the control groups completed the same module with traditional paper-based learning materials. Within each school, the same teacher taught the experiment class and the control class. The relevant teacher in both schools was very experienced, having taught for over 25 years in different types of schools (elementary school, high school, vocational school, etc.). They reported using technology with their students daily (computer-based teaching, tablet-based teaching, etc.). Over the experiment period of five weeks, both classes in both schools had to complete the same module with the same six learning goals (LG). These learning goals are depicted in Appendix A. The teachers used the same teaching methods in both classes, consisting of frontal teaching, followed by individual and partner work. None of the students in the four classes had had law-related subjects before.

2.2 Sample

Following Sahin and Yilmaz (2020), we used convenience sampling in this study, meaning that we chose the schools from two different school types that had two comparable classes and were willing to participate in the study.

2.2.1 Experiment 1 in High School

The sample in experiment 1 consisted of students in two second-year classes of a high school. The experiment class had 9 males and 13 females, with an average age of 16.9 years. The control class had 11 males and 12 females, with an average age of 17.1 years. We conducted ANOVA tests to make sure that the experiment and control class were similar. The tests revealed no significant differences regarding gender ($p = 0.650$), age ($p = 0.788$), school grades in business and law ($p = 0.558$), pre-experience with SPAs ($p = 0.941$), personal innovativeness ($p = 0.191$), and the pre-test problem-solving skill results ($p = 0.774$). This experiment class did not participate in the post-experiment focus group discussion.

2.2.2 Experiment 2 in Vocational Business School

The sample in experiment 2 consisted of students in two second-year classes of a vocational business school. The experiment class had 12 males and 10 females, with an average age of 17.4 years. The control class had 12 males and 11 females, with an average age of 17.2 years. We conducted ANOVA tests to make sure that the two classes were similar. The tests revealed no significant differences in the background of the students in the two classes in terms of their gender ($p = 0.248$), age ($p = 0.743$), school grades ($p = 0.100$), pre-experience with SPAs ($p = 0.552$), personal innovativeness ($p = 0.332$), and the pre-test problem-solving skill results ($p = 0.376$). All students of the experiment class of this school participated in the post-experiment focus group discussion.

2.3 Task and Smart Personal Assistant Design

Between the end of week 1 and the end of week 4, the students had to do four 30-minute, problem-based, homework assignments for learning goals (LGs) 2, 3, 4 and 5. LG 1 was part of the in-class introduction of the related subject. The homework assignments were designed and developed together with the two class teachers. The homework assignments contained problem tasks related to the difference between morality, custom, and law as well as the fundamental rights “freedom of opinion”, “freedom of religion”, and “property guarantee”. All four classes received the task catalog as a paper-based script. The students from the experiment classes used Amazon’s Alexa for the homework assignments. The students from the control classes received an additional paper-based script with steps and hints that are also included in the Alexa Skill. All the homework assignments were in the same style as the pre- and post-test tasks. Furthermore, they fulfilled the requirements of a problem-based task according to Jonassen (2000). Accordingly, the tasks were open-ended, unstructured, resonated with the students' experience, and realistic. As an example, in Appendix B, we show homework assignment 2 and how it addresses these requirements.

We used the Amazon's Alexa Voice Services (AVS) and the Amazon Web Services (AWS) for two reasons. First of all, AVS offers an intuitive graphical interface to design the SPAs (Amazon 2019). Second, Alexa's Skill Development Kit 2.0 seems to offer one of the most developed state-of-the-art capabilities regarding natural language understanding and processing. The database (data related to the problem-solving steps, Alexa's answers for each problem-solving step with each assignment, etc.) and corresponding functions are developed and stored in AWS Lambda.

The implemented dialog flow was derived by scaffolding theory and included static and dynamic scaffolds with a proactive and reactive logic. In the proactive logic, the SPA provided static scaffoldings that guided students through the task by using five different problem-solving activities adapted from Kim and Hannafin's (2011) problem-solving phases (see grey rectangles on the left side in Figure 1): *(a) problem identification and engagement, (b) problem exploration, (c) problem reconstruction, (d) solution presentation and communication,* and *(e) reflection and negotiation.* For example, in a first step (related to problem identification and engagement), the SPA asked the students to identify the core of the problem. These scaffolds were the same for each student and each task. Alexa has started with the first step as soon as the students have chosen one of the four homework assignments.

In the reactive logic, the SPA provided dynamic scaffolds by providing feedback on students' partial solutions and answering students' questions while working on the problem steps. With the help of AWS Lambda functions, we tracked the current progress of the students. The SPA was therefore able to recognize whether an answer was wrong or right and responded accordingly. For example, in Figure 1, the SPA gave a “Are you sure …” response. Moreover, the SPA was able to provide feedback to three different types of questions or queries. Type 1 was the students’ query to skip to the next step. Type 2 was content-specific questions (concept clarifications, asking for hints, etc.). In Type 3, the students requested to repeat the current problem-solving step (see grey rectangles on the left side). For example, in Figure 1, the student asked a content-specific question whether it is public law and the SPA provided feedback. Whenever Alexa was not able to understand the students, it invoked a fallback intent that allowed students to rephrase the question or jump to the next problem-solving step.

The proactive and reactive logic helps students to interact with the SPA in a conversational manner, thereby internalizing the problem-solving steps, which should result in a gain of their problem-solving skills. As an example, Figure 1 shows an excerpt of a student dialogue and the corresponding proactive and reactive functions.

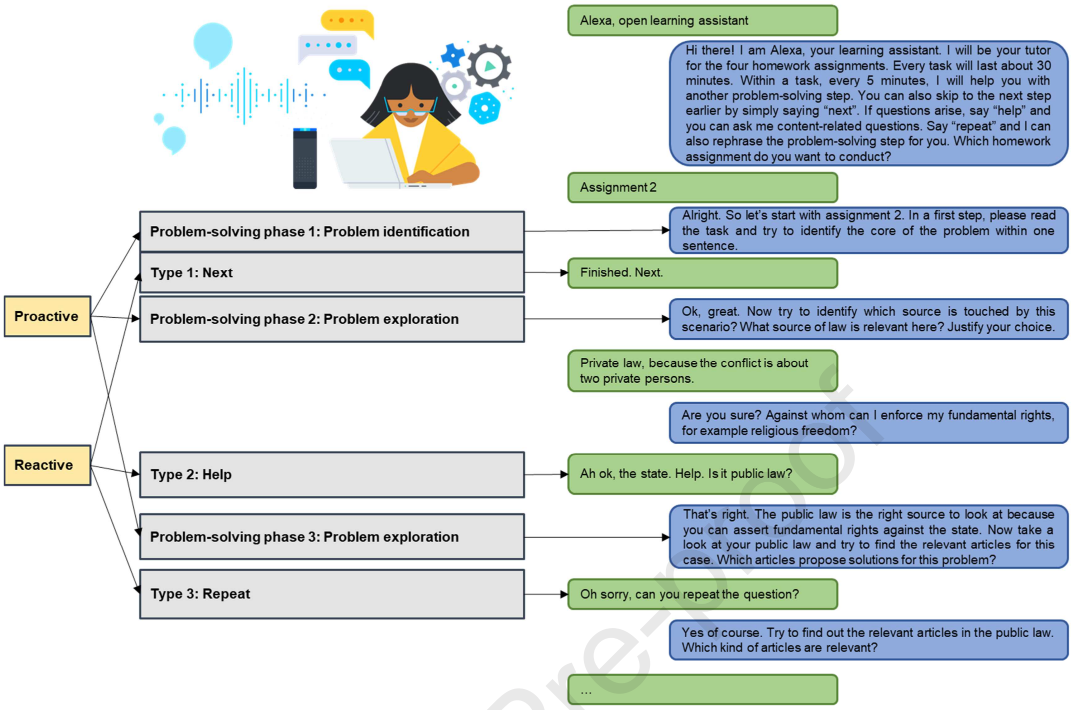


Figure 1. Example dialogue between the student and SPA, and corresponding proactive and reactive logic.

A key challenge when designing SPA-based learning materials for experimental treatments is information equivalence. Larkind and Simon (1987) argue that two representations are informationally equivalent if all the information from one representation can also be inferred from the other representation and vice versa. This ensures that differences stem from the technology itself, rather than the content of the treatment. Therefore, we included all the information incorporated in the SPA also on the paper-based learning materials provided for the control class. Students in the control class received an additional script (separate from the task script) containing all the steps and hints that are also included in the Alexa Skill. Figure 2 shows parts of the additional script, which is the same excerpt as in Figure 1.

Assignment 2:

Step 1: Let's start with assignment 2. In a first step, please read the task and try to identify the core of the problem within one sentence.

Step 2: Try to identify which source is touched by this scenario? What source of law is relevant here? Justify your choice.

Step 3: Take a look at your public law and try to find the relevant articles for this case. Which articles propose solutions for this problem?

…

Hints:

….

Step 3: The public law is the right source to look at because you can assert fundamental rights against the state.

Figure 2. Excerpt of additional script for homework assignment 2 provided for control group.

2.4 Experimental Procedure

Figure 3 depicts the study timeline for experiments 1 and 2.

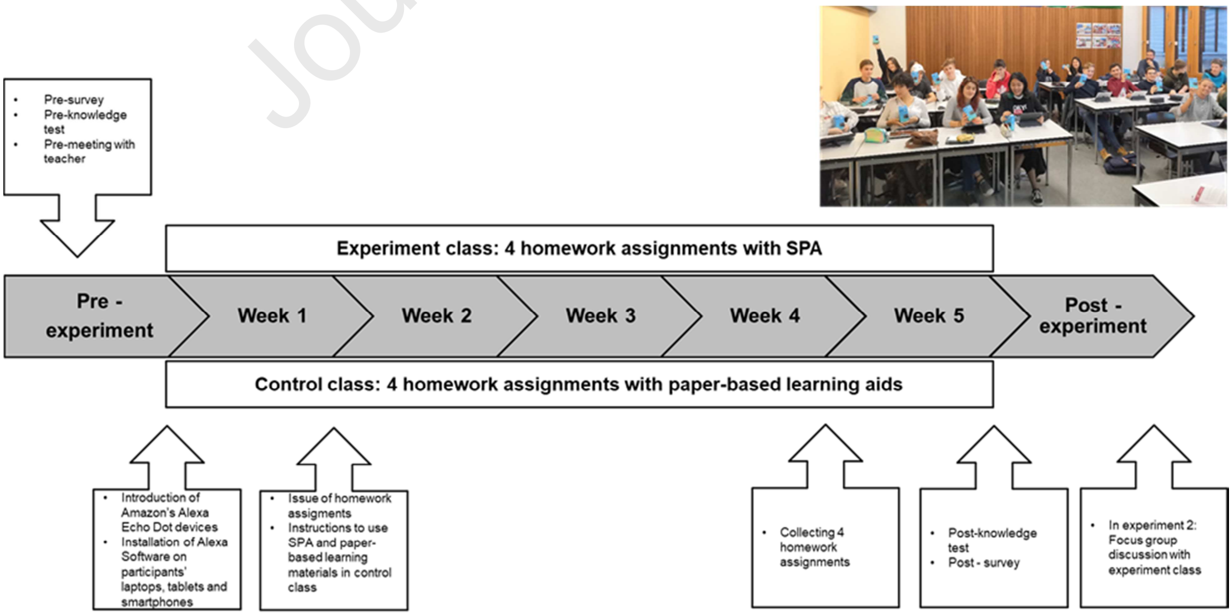


Figure 3. SPA experiment implementation timeline

Several tasks were conducted before the experiment started, i.e., during the pre-experiment state. During a lesson, all students completed a survey to help identify and control for pre-existing class differences (see Appendix C). Also, the classroom teachers administered a 30-minute, 3-subtask pre-knowledge test to all students to compare the knowledge level of each class. Moreover, the research team organized a pre-experiment meeting with the teachers to discuss the experiment details with them and to ensure that both classes are taught the same way over the period of the experiment.

In week 1, we installed the Amazon Alexa Echo Dot devices in the experiment classes and helped students to install the accompanying Alexa software on their smartphones, tablets, and laptops. At the end of week one, we ensured that students in the experiment classes had access to Alexa on one or more devices. After week 1, all four classes received their four homework assignments in the form of a paper script (non-digital). The experiment classes were instructed to use Alexa on their preferred devices (standalone device, smartphone, tablet and laptop) while conducting their homework assignments. The experiment group students did not use SPAs during class. The control classes were instructed to use the paper-based learning materials that were sent to them in form of an additional script (with exactly the same amount of information as on Alexa). At the end of week 4, all students submitted their homework assignments. One week later, they took their 30-minute post-knowledge test and post-experiment survey. The teacher was instructed to not discuss the homework assignments before the students conducted the post- knowledge test to ensure that the final results were not corrupted. In the post-experiment phase, the experiment class in the vocational business school participated in a 45-minute focus group discussion.

2.5 Measurement and Analysis

2.5.1 Quantitative Data

For measuring problem-solving skills (dependent variable), we constructed a 3-task pre- and post-experiment tests. The tasks were similar to the homework assignments, all addressing the requirements of problem-based tasks proposed by Jonassen (2000). The pre-test and post-test had the same number of tasks and possible top score and the tasks addressed learning goals 2, 3, 4 and 5 (see Appendix D). All tasks require the application of Kim and Hannafin's (2011) problem-solving steps. Two teachers compared the pre-test and the post-test in terms of difficulty and found them to be similar. We analyzed pre-test and post-test results with three experienced raters independently and blinded with a pre-defined and commonly-discussed rating framework. One of the raters was the class teacher. The rating framework allowed us to evaluate the application of the problem-solving steps proposed by Kim and Hannafin (2011) giving points to each well-applied problem-solving step within a task (see Appendix E). The final scores of the pre- and post-test results arose from an average of the individual appraisals. The maximum number of points to be achieved was 36: i.e., 12 points per task. Moreover, we calculated gain scores as the difference between a students' posttest and pretest scores. We checked for inter-rater agreement on the pre- and posttest scores of the three raters with the help of Light's Kappa for multiple raters for both classes in both schools. The values for Study 1 and 2 are depicted in Table 1. The strength of agreement refers to the recommendation of Cohen (1960).

Table 1. Interrater – Agreement Scores

	Study 1: High school			Study 2: Vocational school		
	Kappa	P-value	Strength of Agreement	Kappa	P-value	Strength of Agreement
Pre-knowledge test	0.843	< 0.01	Very Good	0.742	< 0.01	Good
Post-knowledge test	0.784	< 0.01	Good	0.610	< 0.01	Good

The pre-experiment survey consisted of items regarding the students' pre-experience with SPA, personal innovativeness, gender, and age. We asked students how often they use SPAs (e.g., Apple's Siri) in a week to measure their pre-experience with SPAs. For personal innovativeness, we used the four items from van Raaij and Schepers' (2008) scale. The detailed pretest-results and the detailed information about the similarity of the two classes can be found in Appendix G. The post-experiment survey included two open-ended questions. Question 1 was about the students' perceived helpfulness of the learning aid (SPA or paper-based learning materials). Question 2 was treatment-specific, asking the experiment classes about their experiences with the SPAs used. The items of the post-survey are shown in Appendix F.

To analyze the quantitative data, we conducted a one-way ANCOVA (analysis of covariance), including the pre-test scores as a covariate. ANCOVA helps to analyze variances between the groups and control for covariates. It is therefore suitable for pre- and post-test designs (Dimitrov and Rumrill 2003). We calculated Cohen's d (1988) to show the effect size. Moreover, we calculated 95% confidence intervals and used the statistic program R as a tool for analysis (Team 2013).

2.5.2 Qualitative Data

The focus group discussion has built on the open questions in the post-survey and addressed RQ2. It aimed to gain a more in-depth understanding of how SPA technology affects students' learning processes. The discussion lasted 45 minutes and had one of the researchers as the facilitator. All students from the experiment class in the vocational business school participated. The focus group included the following steps. First, the facilitator introduced the goal of the group discussion (to gather students' perceptions of using SPAs while learning). Second, students were asked to divide into workgroups of four to six and were invited to discuss and negotiate opinions about how they used the SPAs during their homework assignments. To help students structure their discussion, they received three broad subject areas identified from the post-survey responses: 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. 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 keywords indicating some aspects of the learning process with SPAs and then systematically searched the corpus of the transcribed text to find all instances of the word or phrase. Each time we came across an instance of the word or phrase, we made a note of it and its immediate context. We identified themes by physically sorting the instances into piles of similar meaning (Ryan and Bernard 2003). Then, we conducted a respondent validation by getting participants to verify our identified themes.

3 Quantitative Results

First, we present the results related to our first research question: *Does using a Smart Personal Assistant help students develop their problem-solving skills?*

3.1 Experiment 1 in High School

Table 2 shows the means and standard deviations of the pre-knowledge and post-knowledge test results and the gain scores.

Table 2. Summary of Means

	Experiment Class (SPA)			Control Class (traditional learning materials)		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Pre-knowledge test	10.46	1.67	23	10.27	2.44	22
Post-knowledge test	23.15	4.33	23	20.63	5.60	22
Gain scores	12.69	3.74	23	10.36	3.51	22

To check that the assumptions for the ANCOVA model are met, we conducted statistical and visuals test for normality, homogeneity of variance, and homogeneity of regression slopes (see more details in Appendix H). In particular, we conducted a Shapiro-Wilk test to test the normality assumption ($p = 0.8927$), a Levene's test to check for the homogeneity of variance assumption (*p = 0.7706*), and an interaction analysis for the homogeneity of regression slopes assumption (*p = 0.601*). The test results (all are not significant) as well as the plots indicate that the assumptions for conducting an ANCOVA are met. We ran the ANCOVA with post-test scores as the dependent variable, the treatment group as the independent variable, and the pre-test score as a covariate. The tests show that, controlling for the pre-test, there is a significant relation between SPA usage and problem-solving skills: $F(2, 42) = 4.514$, r^2 *adjusted* $= 0.5255$, $p = 0.0396$, *confidence intervals for mean difference* = *-0.5929 and 5.6329*, $N = 45$. Cohen's d is 0.5178, indicating a difference between the means of 0.5178 standard deviations. Since this effect is considered as medium (Cohen 1988), we can assume that student interactions with SPAs have a positive effect on acquiring problem-solving skills compared to paper-based learning materials. Moreover, a comparison of the gain scores reveals that participants in the SPA group learned significantly more than participants in the paper-based learning materials group ($p = 0.0400$).

3.2 Experiment 2 in Vocational Business School

Table 3 shows the means of the pre-knowledge test and post-knowledge test results, including the standard deviations and gain scores.

Table 3. Summary of Means

	Experiment Class (SPA)			Control Class (traditional learning materials)		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Pre-knowledge test	11.7	1.63	22	10.67	2.16	23
Post-test	27.95	3.27	22	21.88	3.67	23

Gain scores	16.25	3.50	22	11.21	3.74	23

The statistical test results (Shapiro-Wilk test for normality assumption, $p = 0.7847$, Levene's test for homogeneity of variance assumption, $p = 0.7203$, test for homogeneity of regression slopes, $p = 0.260$) and plots given in Appendix H indicate that all assumptions for conducting an ANCOVA are met. We ran the ANCOVA with the post-knowledge test scores as the dependent variable, the treatment group as the independent variable, and the pre-knowledge test score as a covariate. Results of the ANCOVA indicate that, controlling for the pre-test, the relation between SPA usage and problem-solving skills is highly significant: $F(2, 42) = 30.573$, r^2 *adjusted* $= 0.42$, $p = 1.88e\text{-}06$, *confidence interval for mean difference* = *3.9172* and *8.2228*, *Cohen's d* = *1.70, N* = *45*. Since the Cohen's d for this result is considered as high (Cohen 1988), we assume that students' interactions with SPAs have a positive effect on acquiring problem-solving skills compared to the use of traditional learning materials. Moreover, a comparison of the gain scores reveals that participants from the SPA group learned significantly more than participants from the paper-based learning materials group ($p = 8.17e\text{-}06$).

3.3 Focus Group Discussion

Since our quantitative findings lead us to believe that interactions with an SPA have a positive effect on problem-solving skills, we now investigate our second research question about *how SPAs affect students' learning processes* to gather deeper insights into the observed effect. The three main themes we identified from the focus group data are *interaction, usage behavior,* and *individualization.* The main topics and subtopics and the corresponding frequencies of the statements are presented next (see Table 4) and further elaborated in section 4 (Discussion).

Table 4. Subtopics identified from the focus group data

Identified topics and subtopics (out of a student's perspective)	Frequency of Statements
I. Interaction	
- I felt like having an interaction with a real tutor.	22
- Receiving challenging questions and follow-up questions triggered my thinking processes.	15
- SPA responded immediately when I asked a question.	9
II. Usage Behavior	
- Very convenient to use it on my smartphone similar to Google's Assistant or Apple's Siri.	11
- I also used the SPA on other learning places where I would normally study.	7
III. Individualization	
- Felt like the SPA nearly always detected what my questions were about.	13
- It was great that the SPA remembered the current status when I continued to work on a task.	7
- SPA only helped when I needed support.	4

Interaction. This theme relates to SPA technology being able to build up a dialogue with the students. Several students mentioned that using the SPA while completing their assignments felt like having an interaction with a peer or a tutor. Students appreciated that the SPA listened to them and adapted its answers accordingly. Some students mentioned that they like to receive challenging questions and hints from the SPA; it helped them to think of the next solution steps. Other students perceived it as more exciting compared to "business as usual" paper-based learning materials. For example, one student 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 students mentioned that they liked the way the SPA was helping them. Specifically, they said that they appreciated thinking on their own about the solution first and that they could control when they received 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 when I want.” Moreover, some students also mentioned that they liked saying their solutions out loud. While hearing themselves talk, they came up with new solution ideas more easily. For example, one student commented: “I liked speaking the answers out loud and not writing them down. When I hear myself, I get new ideas.”

Usage Behavior. This theme relates to the different ways students used the SPAs compared to paper-based learning materials. Most of the students used the SPA on their smartphone rather than their standalone device and mentioned that it was really easy and fun to access Alexa similar to Google’s Assistant or Apple’s Siri. For example, one student commented: “It was like speaking with Siri. Just like having your personal tutor always in your pocket anytime 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 locations. A handful of students indicated that they did their tasks in other places than usual. For example, one student commented: “When I was lying on my couch, I talked to Alexa a few times, too.” However, a few students also mentioned some areas where SPAs were not very functional. For example, when students were in public, they seldomly used Alexa because they felt uncomfortable talking to an SPA in public. 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 topic relates to the capacity of SPAs to recognize students' individual characteristics. Some students mentioned that they appreciated that Alexa was able to adapt her answers to their utterances and prompted them individually until they were able to find their own solution. For example, one student commented: "It was helpful that Alexa recognized how she can help me discover the solution." Moreover, some students mentioned that they liked that the SPA remembered the status of a partly-finished task so they could continue working from there next time. However, some other students wished that the SPA could recognize where students have their biggest weaknesses. For example, one student commented: "It would be great if she [Alexa] can remember our mistakes and then concentrate on helping us with that."

4 Discussion

The aims of the current study were to investigate the effect of interactions with SPA technology on high school and vocational school students' problem-solving skills and how these interactions impact their learning processes. We observed that the students in the experimental groups had higher increases in their levels of problem-solving skills and we discovered changes in how they learned. With the help of SPAs, students were able to build up individual interactions and receive individual support on devices they use every day. These quantitative and qualitative data indicate a positive effect of SPA technology. Interestingly, this effect was even stronger in vocational business schools where students' characteristics are different. This study expands prior research around SPAs that used mainly explorative approaches (Dousay and Hall 2018).

Although the quantitative data of the rather small sample do not allow more detailed analyses, the qualitative data indicates reasons for the positive effect of interactions with SPA technology on problem-solving skills. First of all, SPAs are able to offer dynamic scaffolds that help to trigger students' interactive learning behavior. According to the constructivist learning paradigm and the ICAP-framework (Chi and Wylie 2014), interactive learning behavior occurs when students discuss the learning content with others where both make a similar contribution. This behavior can be considered as the gold standard leading to better learning outcomes. Our findings suggest that SPAs might replace these discussion partners to some extent. This can also be shown by our qualitative findings indicating that students perceive the interaction with SPAs as similar to interactions with human tutors (theme 1 in the focus group discussion). This reawakens the discussion about the role of SPAs and other AI systems in education. Should AI systems be an *educator-replacing* or an *educator-assisting* technology? The role of AI used only for automation in education systems is questioned (Baker 2016). Cukurova (2019) argues that AI systems should be considered as on a continuum with regard to the extend they are decoupled from humans, rather than only an approach to provide full automation. Indeed, we argue that SPAs can be positioned on both sides of the continuum. In our setting, SPAs have served mainly as educator-replacing to aid educators in settings where they cannot provide individual support due to financial and organizational reasons. SPAs could therefore be used mainly in isolation from educators. In addition, SPAs could also be used to support the decision-making processes of educators. When students interact with SPAs, a great amount of student data can be collected, such as the current level of knowledge, frequency of exercises, motivation at individual, and collective levels. This student data can help educators to introduce new measures (e.g., to explain ununderstood contents again, to offer low-achievers individual help). In our view, the question about the role of AI systems in education is less about an either-or but more about both: i.e., replacing educators where they cannot offer help and informing educators' decision processes. Moreover, the qualitative results indicate that students liked how the SPA fostered their understanding until they found the answer on their own. This indicates the positive effect of dynamic scaffolds and is a big difference to "business-as-usual" paper-based learning materials, where there is no interaction possible and scaffolds are static. It seems that students enjoyed having their personal tutor on their side and knowing that they can receive individual support when they cannot get any further on their own. The SPA then supported them in constructing knowledge on their own. This helped students to better internalize the learning materials and gain their problem-solving skills.

Second, what could also contribute to the positive relationship between using SPAs and students' problem-solving skills might be that students can change their learning behavior (theme 2). In particular, students can change their learning places and learning times. Similar results can be confirmed by Taylor (2006), who stated that using smartphones as learning assistants changed learning contexts, for instance in terms of ergonomics (user posture, lighting, background noise), social context, and demands of users' attention. The different learning places and times might lead to higher amounts of learning time and, finally, better skill-development. On the other hand, students also mentioned that they felt weird using SPAs in public, which is also confirmed by findings from Moorthy and Vu (2015). Their results showed that participants preferred using SPAs in a private location (e.g., their home) due to social acceptability.

Third, one other reason for the superiority of the SPA compared to paper-based learning materials might be that SPAs are able to provide more dynamic support. In specific, students highlighted that the SPAs allowed them to learn at their own pace, receiving help whenever they wanted it. Learning at their own pace motivates students and gives them the feeling of working on their own academic progress (Chen 2008). These effects of receiving individual help whenever they want can also be confirmed by several other research papers in the area of personalized learning (Ammar et al. 2010; Song et al. 2012). Students did however mention that they wished that SPAs could detect individual students' characteristics and in particular track their weaknesses. This confirms the research effort in the area of computer tutoring systems trying to capture students' characteristics and adapt the systems accordingly. For example, Ammar et al. (2010) designed a computer tutoring system equipped with emotional management capabilities, which can capture student's emotions during learning and respond accordingly.

Our work contributes to two different research areas in educational technology. First of all, we contribute to computer tutoring research by providing much-needed empirical evidence on the effect of interactions with a new kind of ITS (SPAs) on students' long-term skill development compared to paper-based learning materials, especially in the field of business and law. To the best of our knowledge, empirical evidence for this technology in this area is still missing. Second, we contribute to technology-mediated learning by arguing that interactions with a new kind of information technology have the potential to change students' learning processes, resulting in increased levels of skill development. Our qualitative results assume that this new technology might be able to offer dynamic scaffolds in a more natural and sophisticated way. Since student-teacher ratios are increasing in today's learning scenarios such as university mass lectures and online courses, SPAs could be very useful. In regard to its practical implications, this study exemplarily showed educators how to build and integrate SPAs in an existing "business as usual" learning environment.

5 Limitations and Future Research

The present study is not without limitations. First, we used a field quasi-experiment design to examine the impact of SPAs on students' problem-solving skills. With this treatment design, pre-treatment group differences between the experimental group and the control group may confound post-treatment outcomes. We tried to address this point by collecting pre-experiment data and using ANOVAs to check if experimental and control groups are similar. Moreover, both classes in each school had the same teacher with the same learning goals and using the same teaching methods. Nevertheless, future experimental research may be able to confirm and extend our results.

Second, the sample size of our two field quasi-experiments (n=45, 2 classes per school) can be considered as rather small for using an ANCOVA. However, our statistics showed a medium effect (in the high school) and a large effect (in the vocational business school) between these two groups, which indicates that there is a relationship between SPA usage and the increase of problem-solving skills. However, this small sample size does not allow further investigations regarding subsets of the sample. For example, it would be particularly interesting to analyze low achiever's engagement with SPAs compared to high achievers or those with low and high metacognitive skills to self-regulate their learning process. In addition, it may be revealing to observe students using these SPAs and derive interaction patterns from these data. These research questions cannot be answered with this study but provide exciting avenues for future research. 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, the focus group consisted only of participants of the second experiment in the vocational school. We cannot completely exclude the possibility that the statements made by participants of the first experiment (high school) would differ.

Fourth, 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 small, given the large percentage of participants reporting already high usage of SPAs on their smartphones (approximately 40% used SPAs every day on their smartphones and another 40% used it weekly; e.g., Siri). Nevertheless, using SPAs in the context of homework was something completely new for many students, which raises the question of the extent to which these systems can be used in the long term. It would be interesting to see in long-term studies whether these students also show increased problem-solving skills after 3-4 months or even longer. Another exciting question would be to investigate whether the time of use of these devices increases or decreases over the course of a whole semester. This study cannot provide answers to these questions, but they offer interesting avenues for future research.

Finally, this study also raises some ethical questions about data security and the potential benefit of one group and not the other. Students put their information on Amazon's ecosystem, which might be a risk for some students. While it is hard to control student behavior, we ethically approved the study and educated students about the experiment and the corresponding risks at the beginning of the study. For future research, it would be interesting to see how SPA technology can be implemented in-class to improve students' skill development. Moreover, future research should focus on individual differences in students' characteristics that might influence learning processes with SPAs.

6 Conclusion

Our study answered two research questions. First, we investigated whether SPA technology could increase students' problem-solving skills. Within two experiments in two different school types (high school and vocational school), we were able to show that the use of SPAs over a period of five weeks has a positive effect on skill development, more precisely the development of problem-solving skills. Second, we examined how the use of SPA technology changes students' learning processes. Students showed more interactive learning behavior and used the SPA in different ways compared to traditional learning aids. Moreover, students appreciated that they received individualized support from the SPA. The study provides empirical evidence for the usefulness of SPA technology and offers fresh insights into how this technology can change students learning processes. The findings of this study contribute to computer tutoring and technology-mediated learning research.

Funding

No funding.

Ethical approval

In this study, all procedures involving human participants were in accordance with the ethical standards of the institutional and/or national research committee.

Informed consent

Informed consent was obtained from all individual participants included in the study.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Referenes

References

Afzal S, Dhamecha T, Mukhi N, Sindhgatta R, Marvaniya S, Ventura M, Yarbro J (2019) Development and Deployment of a Large-Scale Dialog-based Intelligent Tutoring System. In: Human Language Technologies (ed), vol 2, pp 114–121

Amazon (2019) New Amazon Alexa Skill Kit 2.0. https://developer.amazon.com/blogs/alexa/post/1a4e8b01-663d-4680-8efdc28e96e31655/now-available-version-2-of-the-ask-software-development-kit-for-java. Accessed 20 November 2018

Ammar MB, Neji M, Alimi AM, Gouardères G (2010) The affective tutoring system. Expert Systems with Applications 37:3013–3023

Arend B (2018) Investigating Siri as a virtual assistant in a learning context. In: Proceedings of 12th annual International Technology, Education and Development Conference, pp 1–10

Baker RS (2016) Stupid tutoring systems, intelligent humans. International Journal of Artificial Intelligence in Education 26:600–614

Brush TA, Saye JW (2002) A summary of research exploring hard and soft scaffolding for teachers and students using a multimedia supported learning environment. The Journal of Interactive Online Learning 1:1–12

Chen C-M (2008) Intelligent web-based learning system with personalized learning path guidance. Computers & Education 51:787–814. https://doi.org/10.1016/j.compedu.2007.08.004

Cheung ACK, Slavin RE (2013) The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. Educational Research Review 9:88–113

Chi MTH, Wylie R (2014) The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. Educational Psychologist 49:219–243. https://doi.org/10.1080/00461520.2014.965823

Chung H, Iorga M, Voas J, Lee S (2017) Alexa, Can I Trust You? Computer 50:100–104. https://doi.org/10.1109/MC.2017.3571053

Cohen J (1960) A coefficient of agreement for nominal scales. Educational and psychological measurement 20:37–46

Cohen J (1988) Statistical power analysis for the behavioral sciences 2nd ed. Erlbaum Associates, Hillsdale

Cukurova M (2019) Learning analytics as AI extenders in education: Multimodal machine learning versus multimodal learning analytics. In: Proceedings of the Artificial Intelligence and Adaptive Education Conference (ed) Learning analytics as AI extenders in education: Multimodal machine learning versus multimodal learning analytics

Damacharla P, Dhakal P, Stumbo S, Javaid AY, Ganapathy S, Malek DA, Hodge DC, Devabhaktuni V (2019) Effects of Voice-Based Synthetic Assistant on Performance of Emergency Care Provider in Training. International Journal of Artificial Intelligence in Education 29:122–143. https://doi.org/10.1007/s40593-018-0166-3

David H, Katz LF, Kearney MS (2006) The polarization of the US labor market. American economic review 96:189–194

Dimitrov DM, Rumrill PD (2003) Pretest-posttest designs and measurement of change. Work 20:159–165

Dousay TA, Hall C (2018) Alexa, tell me about using a virtual assistant in the classroom. In: Proceedings of EdMedia: World Conference on Educational Media and

Technology. Association for the Advancement of Computing in Education (AACE), Amsterdam, Netherlands

Easwara Moorthy A, Vu K-PL (2015) Privacy Concerns for Use of Voice Activated Personal Assistant in the Public Space. International Journal of Human-Computer Interaction 31:307–335. https://doi.org/10.1080/10447318.2014.986642

Educause (2019) 2019 Horizon Report. https://library.educause.edu/resources/2019/4/2019-horizon-report

Ertmer PA, Newby TJ (1993) Behaviorism, Cognitivism, Constructivism: Comparing Critical Features from an Instructional Design Perspective. Performance Improvement Quarterly 6:50–72. https://doi.org/10.1111/j.1937-8327.1993.tb00605.x

Fraenkel JR, Wallen NE, Hyun HH (2011) How to design and evaluate research in education. New York: McGraw-Hill Humanities/Social Sciences/Languages

Glasersfeld Ev (1987) Constructivism. The concise Corsini encyclopedia of psychology and behavioral science 6:19–21

Goksel-Canbek N, Mutlu ME (2016) On the track of Artificial Intelligence: Learning with Intelligent Personal Assistants. HumanSciences 13:592. https://doi.org/10.14687/ijhs.v13i1.3549

Graesser AC, Hu X, Sottilare R (2018) Intelligent tutoring systems. In: International handbook of the learning sciences. Routledge, pp 246–255

Hauswald J, Laurenzano MA, Zhang Y, Li C, Rovinski A, Khurana A, Dreslinski RG, Mudge T, Petrucci V, Tang L, Mars J (2015) Sirius: An open end-to-end voice and vision personal assistant and its implications for future warehouse scale computers. In: ACM SIGPLAN Notices (ed) ACM SIGPLAN Notices, 50th edn. ACM, pp 223–228

Hooshyar D, Ahmad RB, Yousefi M, Fathi M, Horng S-J, Lim H (2016) Applying an online game-based formative assessment in a flowchart-based intelligent tutoring system for improving problem-solving skills. Computers & Education 94:18–36

Jean-Charles A (2018) Internet of Things in Education: Artificial Intelligence Voice Assistant in the Classroom. In: Proceedings of Society for Information Technology & Teacher Education International Conference (ed) Proceedings of Society for Information Technology & Teacher Education International Conference. Association for the Advancement of Computing in Education (AACE)

Jonassen DH (2000) Toward a design theory of problem solving. ETR&D 48:63–85. https://doi.org/10.1007/BF02300500

Kim MC, Hannafin MJ (2011) Scaffolding problem solving in technology-enhanced learning environments (TELEs): Bridging research and theory with practice. Computers & Education 56:403–417

Kloos CD, Alario-Hoyos C, Munoz-Merino PJ, Ibanez MB, Estevez-Ayres I, Fernandez-Panadero C (2020) Educational Technology in the Age of Natural Interfaces and Deep Learning. IEEE R. Iberoamericana Tecnologias Aprendizaje 15:26–33. https://doi.org/10.1109/RITA.2020.2979165

Kulik JA, Fletcher JD (2016) Effectiveness of Intelligent Tutoring Systems. Review of educational research 86:42–78. https://doi.org/10.3102/0034654315581420

Larkin, J., Simon, H. (1987) Why a Diagram is (Sometimes) Worth Ten Thousand Words. Cognitive science 11:65–100. https://doi.org/10.1016/S0364-0213(87)80026-5

Luckin R, Holmes W, Griffiths M, Forcier LB (2016) Intelligence unleashed: An argument for AI in education. Pearson Education, London

Ma W, Adesope OO, Nesbit JC, Liu Q (2014) Intelligent tutoring systems and learning outcomes: A meta-analysis. Journal of educational psychology 106:901–918. https://doi.org/10.1037/a0037123

Mavrikis M, Grawemeyer B, Hansen A, Gutierrez-Santos S (2014) Exploring the Potential of Speech Recognition to Support Problem Solving and Reflection. In: Rensing C, Freitas S de, Ley T, Muñoz-Merino PJ (eds) Open Learning and Teaching in Educational Communities, vol 8719. Springer International Publishing, Cham, pp 263–276

McTear M, Callejas Z, Griol D (2016) The conversational interface. Springer 6:102

Mousavinasab E, Zarifsanaiey N, R. Niakan Kalhori S, Rakhshan M, Keikha L, Ghazi Saeedi M (2018) Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods. Interactive Learning Environments 11:1–22. https://doi.org/10.1080/10494820.2018.1558257

Nesbit JC, Adesope OO, Liu Q, Ma W (2014) How Effective are Intelligent Tutoring Systems in Computer Science Education? In: IEEE 14th International Conference on Advanced Learning Technologies. IEEE, pp 99–103

Nye BD (2014) Barriers to ITS Adoption: A Systematic Mapping Study. In: Hutchison D, Kanade T, Kittler J, Kleinberg JM, Kobsa A, Mattern F, Mitchell JC, Naor M, Nierstrasz O, Pandu Rangan C, Steffen B, Terzopoulos D, Tygar D, Weikum G, Trausan-Matu S, Boyer KE, Crosby M, Panourgia K (eds) Intelligent Tutoring Systems, vol 8474. Springer International Publishing, Cham, pp 583–590

OECD (2014) Creative problem solving: Students' skills in tackling real-life problems. PISA 2012 results, / Organisation for Economic Co-operation and Development, Programme for International Student Assessment ; Vol. V. OECD, Paris

Oeste S, Lehmann K, Janson A, Söllner M, Leimeister JM (2015) Redesigning University Large Scale Lectures: How To Activate The Learner. In: 75th Academy of Management Annual Meeting, Vancouver, British Columbia

Pai K-C, Kuo B-C, Liao C-H, Liu Y-M (2020) An application of Chinese dialogue-based intelligent tutoring system in remedial instruction for mathematics learning. Educational Psychology 35:1–16. https://doi.org/10.1080/01443410.2020.1731427

Perez S (2017) Voice-enabled smart speakers to reach 55% of U.S. households by 2022. https://techcrunch.com/2017/11/08/voice-enabled-smart-speakers-to-reach-55-of-u-s-households-by-2022-says-report/

Rietsche R, Duss K, Persch JM, Soellner M (2018) Design and Evaluation of an IT-based Formative Feedback Tool to Foster Student Performance. In: 39th International Conference on Information Systems (ICIS), San Francisco, CA, USA, pp 1–17

Ryan GW, Bernard HR (2003) Techniques to identify themes. Field methods 15:85–109

Sahin D, Yilmaz RM (2020) The effect of Augmented Reality Technology on middle school students' achievements and attitudes towards science education. Computers & Education 144:103710. https://doi.org/10.1016/j.compedu.2019.103710

Soderstrom NC, Bjork RA (2015) Learning versus performance: An integrative review. Perspectives on Psychological Science 10:176–199

Song Y, Wong L-H, Looi C-K (2012) Fostering personalized learning in science inquiry supported by mobile technologies. Education Tech Research Dev 60:679–701

Spitz-Oener A (2006) Technical change, job tasks, and rising educational demands: Looking outside the wage structure. Journal of labor economics 24:235–270

Taylor J (2006) Evaluating mobile learning: What are appropriate methods for evaluating learning in mobile enviroments. Big issues in mobile learning 1:25–27

Team RC (2013) R: A language and environment for statistical computing. http://www.R-project.org/. Accessed 1 May 2019

Terzopoulos G, Satratzemi M (2019) Voice Assistants and Artificial Intelligence in Education. In: Eleftherakis G, Lazarova M, Aleksieva-Petrova A, Tasheva A (eds) Proceedings of the 9th Balkan Conference on Informatics. ACM, New York, NY, USA, pp 1–6

van Raaij EM, Schepers JJL (2008) The acceptance and use of a virtual learning environment in China. Computers & Education 50:838–852

Vanlehn K (2011) The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. Educational Psychologist 46:197–221. https://doi.org/10.1080/00461520.2011.611369

Venkatesh V, Brown SA, Bala H (2013) Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. MIS quarterly:21–54

Vygotsky LS (1978) Mind in society: The development of higher mental process. Cambridge, MA: Harvard University Press

APPENDICES

Appendix A: Lessons and learning goals

Lesson (15 lessons in total per class)	Learning Goals
Lesson 1	LG1: Students should understand the necessity of law in everyday life.
Lesson 2 to 4	LG2: Students should explain differences between morality, custom and law.
Lesson 5 to 7	LG3: Students should solve problems related to the freedom of opinion.
Lesson 8 to 10	LG4: Students should solve problems related to the freedom of religion.
Lesson 11 to 13	LG5: Students should solve problems related to the property guarantee.
Lesson 14 to 15	LG6: Students should analyze in which cases fundamental right have their limitations.

Appendix B: Homework assignment 2 and characteristics of a problem-based task

Homework Assignment 2

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 text his mother that he will be late that day. The teacher collects the smartphone and tells Thomas K. that she will keep the smartphone until the end of the week. How would you solve this emerging problem with the help of the law?	
Characteristics of a problem-task (Jonassen 2000)	**How we addressed them**
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.
Open-ended	The task leaves room for interpretation. There is not a single right answer. The students have to interpret the legal articles in a correct way.
Realistic	The task relates to a topic that is currently highly discussed in the relevant country. It can therefore be considered as realistic.
Resonate with the executors' experience	Since every participant owns a smartphone and has experienced similar situations in their own class, they can put themselves in Thomas K.'s position.

Appendix C: Pre-survey items

Pre-survey	

Variable	Item	Scale
Pre-experience with SPAs	1. Have you ever used a Smart Personal Assistant (e.g. Amazon's Alexa, Google's Assistant, Apple's Siri? 2. If yes, how often do you use a Smart Personal Assistant per week on average?	Yes/No open
Personal innovativeness	1. If I heard about a new information technology, I would look for ways to experiment with it. 2. Among my peers, I am usually the first to try out new information technologies. 3. In general, I am hesitant to try out new information technologies (reverse-scored). 4. I like to experiment with new information technologies.	1 to 7 (7 = highest) 1 to 7 1 to 7 1 to 7
Demographics	1. Age 2. Gender 3. Apprenticeship company (for vocational business school only)	open open open

Appendix D: Pretest- and post-test tasks

Learning goal	Pre-test	Post-test

LG3: solve problems related to the freedom of opinion	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?	Imagine that in your country, Michael P, a good friend of yours, issues flyers that are obviously against one of the left-wing politicians in the country. Some people complain against that. How would you solve this problem?
LG4: solve problems related to the freedom of religion	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 are caught by the police and have to pay a fee now. They are complaining about it. How would you solve this problem?	Imagine that Muslims in your country have built a mosque in your neighborhood. After a while, a neighbor gets excited and criticizes the construction of the mosque. How would you solve the problem?
LG5: solve problems related to property guarantee	Imagine the state you are living in wants to build a road on your land. If you do not allow this, your land will be expropriated. You are filling a complaint against the expropriation of your	Imagine that the city you are living in want to create a one-week food festival and, thus, needs two of your land plots, against which you are filing a complaint. How would you solve

	land. How would you solve this problem?	this problem?

Appendix E: General rating framework per task

Problem-solving step	Key Questions (1 Point per question)	Points
1. Problem identification and engagement	What is the main problem that is touched here? Who are the involved parties?	2
2. Problem exploration	Which area of the law is affected? What are the interests of the individual parties?	2
3. Problem reconstruction	What is a possible solution for the problem? Which article could apply here? What does the given law article say?	3
4. Solution presentation and communication	How can the law article be applied in the context? Are other articles also applicable?	3

	Who can raise which claims?	
5. Reflection and Negotiation	How can you justify the solution? How can the related fundamental rights be restricted?	2
Total		**12**

Appendix F: **Post-survey items**

Post-survey		
Variable	Item	Scale
Question 1	To what extent do you feel that the learning materials has helped you? Why?	Open
Question 2 (treatment-specific)	What are your experiences when using the Smart Personal Assistant as a tutor?	Open

Appendix G: Detailed pre-test results and student characteristics

	Experiment Class (SPA)	Control Class (traditional learning materials)
Study 1: High School		

	Points achieved	Std. Dev.	N	Mean	Std. Dev.	N
Sub-task 1 (total: 12)	2.65	1.32	23	2.60	1.94	22
Sub-task 2 (total: 12)	3.31	1.59	23	3.21	2.53	22
Sub-task 3 (total: 12)	4.50	2.11	23	4.46	2.83	22
Total	**10.46**	**1.67**	**23**	**10.27**	**2.44**	**23**
Study 2: Vocational Business School						
	Points achieved	Std. Dev.	N	Mean	Std. Dev.	N
Sub-task 1 (total: 12)	2.85	1.25	23	2.55	1.54	22
Sub-task 2 (total: 12)	4.00	1.65	23	3.55	2.34	22
Sub-task 3 (total: 12)	4.85	2.02	23	4.60	2.12	22
Total	**11.70**	**1.63**	**23**	**10.67**	**2.16**	**23**

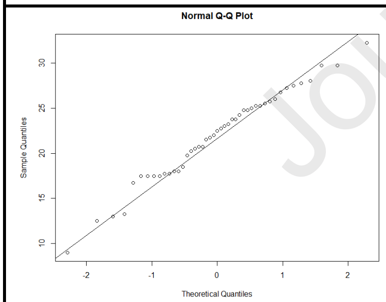
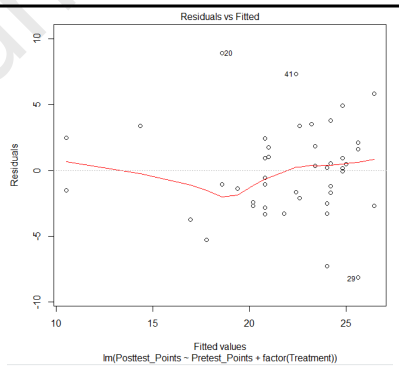
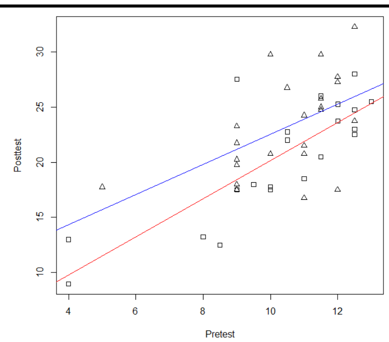
	Experiment Class (SPA)	Control Class (traditional learning	p-value

		materials)	
Study 1: High school			
Gender	Male = 9, Female = 13	Male = 11, Female = 12	0.650
Age	16.9	17.1	0.788
School grades (6 is best)	4.69	4.61	0.558
Pre-experience with SPAs (7 is best)	3.17	3.14	0.941
Personal innovativeness (7 is best)	4.55	4.80	0.191
Study 2: Vocational Business School			
Gender	Male = 12, Female = 10	Male = 12, Female = 11	0.248
Age	17.4	17.2	0.743
School grades (6 is best)	4.47	4.20	0.100
Pre-experience with	2.77	2.39	0.552

SPAs (7 is best)			
Personal innovativeness (7 is best)	4.20	4.57	0.332

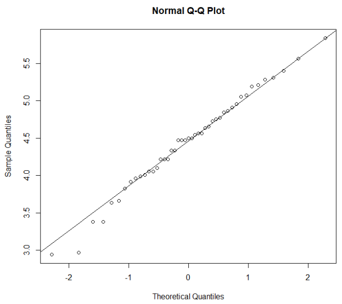
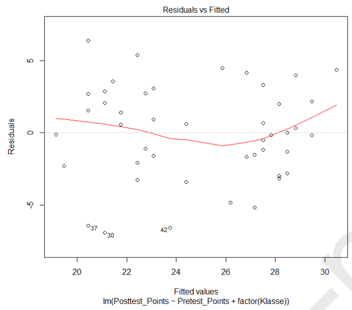
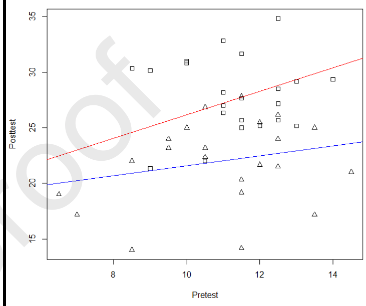
Appendix H: Graphical check for assumptions for ANCOVA

Experiment 1 in high school

Statistical and visual test for normality	**Test for homogeneity of variance**	**Test for homogeneity of regression slopes**
Shapiro-Wilk normality test *W = 0.98714* *p-value = 0.8927*	Levene's test *F = 0.0861* *p-value = 0.7706*	Interaction Analysis *F = 0.277* *p-value = 0.601*
		

Experiment 2 in vocational business school

Tests for normality	**Tests for homogeneity of**	**Tests for homogeneity of**

	variance	regression slopes
Shapiro-Wilk normality test $W = 0.98408$ $p\text{-value} = 0.7847$	Levene's test $F = 0.1299$ $p\text{-value} = 0.7203$	Interaction Analysis $F = 1.305$ $p\text{-value} = 0.260$
		

- Smart Personal Assistant technology foster students' problem-solving skills
- Smart Personal Assistant interactions change students' learning behavior
- Two field quasi-experiments and focus group provides evidence for this effect
- Findings contribute to computer tutoring and technology-mediated learning research