Please quote as: Winkler, R., Hobert, S., Fischer, T., Salovaara, A., Söllner, M., and Leimeister, J. M. (2020) Engaging Learners in Online Video Lectures with Dynamically Scaffolding Conversational Agents. In European Conference on Information Systems (ECIS) 2020, ECIS.

ENGAGING LEARNERS IN ONLINE VIDEO LECTURES WITH DYNAMICALLY SCAFFOLDING CONVERSATIONAL AGENTS

Research paper

Abstract

Online education creates new opportunities for learners, which has led to sharply increasing enrollment in the last few years. Despite these benefits, past research shows that the lack of individual interaction with educators creates low learner engagement that leads to high attrition rates, which remains a major challenge in the field. Dynamically scaffolding conversational agents built into online video lectures promise to address this problem by individually interacting with learners, similar to educators' scaffolding behavior. These agents are equipped with recent natural language processing capabilities, creating human-like conversations that help learners to be more engaged in the learning process. To test our hypothesis, we built a dynamically scaffolding conversational agent named Sara and compared it with an often-implemented static conversational agent built into two online video lectures. We deployed a lab experiment with 182 learners. The neurophysiological measurements revealed that Sara significantly improved learner engagement partly explained by differences in learners' perceptions in the way they experienced the interaction. This study connects to already existing conversational agent studies in online education and highlights the importance of including dynamically scaffolding conversational agent in online video lectures to address the problem of low learner engagement in online education.

Keywords: Conversational agent, learner engagement, dynamic scaffolding, NeuroIS, experiment.

1 Introduction

In recent decades, the number of learners enrolled in online learning courses has risen rapidly (Song et al., 2017). In 2017, 33.1% of students worldwide took at least one course online, compared to 24.8% in 2012 (Lederman, 2018). Massive open online courses (MOOCs) are increasingly becoming a standard learning scenario. This development brings a lot of benefits. In specific, online courses can be provided to an unlimited amount of people independent of time and place, resulting in a more varied learner population compared to the traditional university population in terms of age, gender, socio-economic status, culturally and linguistically diverse backgrounds, geographical locations and life experiences (Richardson et al., 2017). However, this new shift from offline to online learning is faced with a challenge: Learners are mostly forced to watch online video lectures in a passive and one-size-fits-all way without being able to directly interact with the educators. This is problematic, since we know from constructivist learning theories that we learn better when we actively engage in the learning process (Glasersfeld, 1987). There is consensus in literature that learner engagement is one of the key factors of learning success (Chi and Wylie, 2014). Nevertheless, past research shows that learners lack engagement in online learning environments (Lim, 2004). Meaningful interactions, such as scaffolding dialogs between educators and learners, help learners to become more engaged in the learning process (Ferguson and Clow, 2015).

Trying to address the problem of low learner engagement in online learning, conversational agents (CA) have aroused interest in the field. A CA is a computer system intended to converse with a human (Rubin et al., 2010). In education, a CA uses text, speech, graphics, haptics, gestures, and other modes in various combinations for helping learners to conduct tasks and gain knowledge (Kerry et al., 2009). New emerging CAs with natural language processing (NLP) capabilities are able to interact with learners in a free manner (Luger and Sellen, 2016). In specific, CAs can analyze the individual learner in detail in order to offer dynamic scaffolds when needed, similar to the scaffolding behavior of human educators (Graesser et al., 2017). This is different from most of the currently implemented CAs in online education which are rather static with one-size-fits-all hints for all learners or allowing only short replies or buttons to click (Ruan et al., 2019). For example, Song et al. (2017) created a CA in which learners were triggered to reflect about their learning. The CA was not really able to individually react to the learners' utterances and offer corresponding questions. These kinds of CAs are often not able to build up meaningful interactions with learners that are necessary during online video lecture. Until now, studies that investigate the effect of CAs on learner engagement are scarce. For example, Graesser et al. (2001) found mixed results about the CA's effect on learning engagement in a summarizing strategy task. It remains unclear whether interactions with dynamically scaffolding CAs built into online video lectures are able to increase learner engagement and how learners perceive those interactions. In the wake of increasing opportunities in online learning on the one hand and new emerging CA technology on the other, our study answers the following question:

RQ1: Do dynamically scaffolding conversational agents increase learner engagement during online video lectures?

RQ2: How do learners perceive interactions with dynamically scaffolding conversational agents during online video lectures?

To answer our research questions, we employed a mixed method approach, complementing the neurophysiological data of a comparison group design (RQ1) with the qualitative data of an open-ended question in the post-survey (RQ2). This resembles type 1 of Venkatesh et al's (2013) purposes of mixed method research (complementarity). Mixed-method research can develop insights into new phenomena of interest that cannot be fully understood using only one method (Johnson and Onwuegbuzie, 2004). We use a mixed method approach because our aim is to provide a holistic understanding of how dynamically scaffolding CAs built into online video lectures influence learner engagement. Based on scaffolding and learner engagement theory, we argue that dynamically scaffolding CAs improve learner engagement and that this effect is partially explained by differences in how learners perceive the CA interactions. To test this hypothesis, we employed a repeated-measure experiment design with 182 learners from a European university where we compared the dynamically scaffolding CA Sara with a statically

scaffolding CA and a control group within an online video lecture setting. Learners watched two video lectures on programming where Sara was layered on top of an existing online video lecture. She interrupted the online video lecture at fixed times to interact with the learners. The facial expression and skin detection scores suggest that the dynamically scaffolding CA is able to significantly increase learner engagement while learners are watching the video compared to the statically scaffolding CAs and the control group. The qualitative data of the open question in the post-survey reveals that learners from the dynamically scaffolding CA group more often reported that they felt emotionally involved and that the interaction felt a bit like a real dialog with an educator. Moreover, the dynamic scaffolding dialog helped them to structure their own thinking processes. However, some statements also indicate that especially voice recognition technology is not ready to compete with human interaction yet. Our results build on past research regarding CA design in online learning environments. To the best of our knowledge, there is no study that investigated the effect of CAs on learner engagement during online video lectures. Our findings emphasize the importance of including dynamically scaffolding mechanisms in CAs and to implement CAs during online video lectures in order to address the problem of low learner engagement in online education.

2 Theoretical Background and Hypotheses Development

2.1 Conversational Agents in Education

A CA is a computer system intended to converse with a human (Rubin et al., 2010). CAs in education have been developed for a wide range of applications, including tutoring (e.g., Graesser et al., 2001), answering questions (e.g., Kerly et al., 2007), conversation practice for language learners (e.g., Ayedoun et al., 2015), educators and learning guides (e.g., Johnson et al., 2000), and dialogs to promote reflection and metacognitive skills (e.g., Song et al., 2017). The design, implementation and strategies of CAs used in education vary widely, reflecting the diversity of emerging CA technologies. Conversations with CAs are generally mediated through simple text-based forms (e.g., Song et al., 2017), where users click buttons or type answers and questions using the keyboard. Some systems use embodied CAs (e.g., Cassell, 2000) that can display emotions and gestures, while others use a simpler avatar (e.g., Kerly et al., 2007). Some systems operate via speech output using text-to-speech synthesis (e.g., Griol et al., 2017), and speech input systems are becoming increasingly feasible (e.g., Winkler et al., 2019). With the latest technological developments in natural language processing, interactions with a CA are slowly approaching a real human-human interaction, where users can freely interact with CAs and where they can adapt their responses to users' utterances (Lu et al., 2018). These technological improvements have great potential to help CAs to imitate individual educator-learner interactions, which is considered as the gold standard of learning. More specific, CAs with natural language processing capabilities might be able to build up dynamic scaffolding dialogs with learners, where they are able to ask learners questions and can adapt their answers according to learners' utterances. Past researchers mainly implemented CAs that are not fully able to build up a free dialog with learners (Vanlehn, 2011). For example, Ruan et al.'s (2019) implemented a CA called QuizBot that allowed learners to provide simple answers and click on buttons to get a further explanation or go to the next question. There is little research regarding the relationship of CAs and learner engagement (Gulz et al., 2011). For example, Panzoli et al. (2010) proposed three levels of interaction that should help CA designers to build scaffolds that support learner engagement in a computer environment. Li and Graesser (2017) investigated the impact of agent formality on learner engagement in an authentic reading and writing environment and revealed that formal CA discourse causes higher levels of engagement. Until now, studies that investigate the effect of dynamically scaffolding CAs on learner engagement are missing. However, past research let us assume that dynamically scaffolding CAs can be beneficial, especially when they are built into existing online video lectures. This calls for the question whether dynamically scaffolding CAs are able to engage learners more deeply while watching a video. Especially in the field of online learning, where a lack of learner engagement is one of the main causes of learners' attrition and failure rates, this would be an important finding to inform future research regarding the design of CAs in online learning.

2.2 Scaffolding Theory

Bruner et al.'s (1976) theory of scaffolding emerged around 1976 and was particularly influenced by the work of Russian psychologist Lev Vygotsky (Vygotsky, 1978). Vygotsky argued that we learn best in a social environment, where we construct meaning through interactions with others. His Zone of Proximal Development Theory, which states that we can learn more in the presence of a knowledgeable other person, became the basis for the theory of scaffolding. The main goal of the educator is to offer scaffolds, such as questions and hints, within the individual zone of proximal development of the learner. Once the learner is able to work independently on a particular subtask, the scaffolds are gradually removed. It is a process by which a beginner can achieve a goal or objective that would otherwise not be achievable without support (Bull et al., 1999). Scaffolding conversational agents can be further divided into dynamically and statically scaffolding CAs (Molenaar et al., 2012). Static scaffolds are exactly the same for all learners and do not change over time (e.g., a preparated list of hints). The support provided by static scaffolding cannot be adapted to individual persons or situations. In contrast, dynamic scaffolding CAs allow more flexibility, as they individually adapt to the learner and the situation. With dynamic scaffolding, the actual goal of scaffolding can be achieved, i.e. the adaptation as required and the reduction of support over time (Raad, 2000). Azevedo et al. (2004) examined the role of different scaffolding lectural interventions (dynamic, static and no scaffolding) and revealed that the dynamic scaffolding condition facilitated a shift in learners' mental models significantly more than the static and no scaffolding condition did. Similarly, Molenaar et al. (2012) tested the effects of dynamic scaffolding on self-regulation of middle school learners working in a computer-based learning environment and found that scaffolding had a positive effect on the dyad's learning performance compared to the non-scaffolding control group. This kind of research let us believe that dynamic scaffolding incorporated in CAs might also have a positive influence on learner engagement in online video lectures.

2.3 Learner Engagement and NeuroIS

Learner engagement could be described as the holy grail of learning (Halverson and Graham, 2019). Research shows that multifarious benefits occur when learners are engaged in their own learning, including motivation and achievement (Sinatra et al., 2015). Learner engagement is a measure that reflects the quantity and quality of a learner's involvement with its learning material (Halverson and Graham, 2019). A learner is engaged when he or she is active in their learning, eager to participate, willing to expend effort, motivated and inspired (Sinatra et al., 2015). The '**I**nteractive, **C**onstructive, **A**ctive and **P**assive (ICAP)' Framework proposed by Chi and Wylie (2014) explains the different engagement modes while learning by classifying observable learner behavior into four modes: interactive, constructive, active, and passive. It predicts that these modes will be ordered by effectiveness as follows: interactive > constructive > active > passive. Educators have long recognized that although learners can learn by receiving information passively, they learn much better by learning actively (Benware and Deci, 1984). Learning actively requires learners to engage cognitively and meaningfully with the tasks that they are dealing with. When learning actively, learners think about their learning material in depth rather than just passively receiving it. When learners are in a passive engagement mode, they consume information passively without carrying out any physical activity. An example of this is to watch a tutorial without doing anything else than watching. In contrast, an active engagement mode means that a physical activity is carried out. An example of this is the marking of text passages while reading. A constructive engagement mode goes one step further. The learner creates new output that goes beyond the subject matter, such as writing notes in his or her own words during the lecture. An interactive engagement mode can be considered as the gold standard of learning where both partners have similar contributions to the dialog. An example of this is a scaffolding dialog between an educator and a learner (Chi and Wylie, 2014).

The measurement of learner engagement is a highly discussed topic in educational psychology (Shernoff et al., 2014). The types of measurements can be roughly divided into macro-level (e.g., learner's self reportings, observations) and micro-level measurements (e.g., neuroIS measurements). Research in the area of educational psychology claims that learner engagement should be better measured on a micro-

level (Siadaty et al., 2016). Macro-level measurements, such as self-reportings, have the inherent problem of retrospection or biased results. For example, in self-reportings, learners were asked to reflect on to the lesson they just experienced as they consider their responses. Another type of macro-level measurement, observations, eliminates the need for retrospection, because the observations can be made in real time but involves a bias from the observer. Micro-level measurements such as physiological sensors (e.g., skin detection and facial expression) can address both of these challenges. Physiological sensors provide an unfiltered window into learners' cognitive and emotional activity (Figner and Murphy, 2011). The nascent field of NeuroIS is drawing upon the theories, methods and tools offered by cognitive neuroscience and psychophysiology (Dimoka et al., 2012). The methodology used ranges from functional magnetic resonance tomography to the measurement of electrical brain activity (electroencephalography) to the recording of the activity of the autonomous nervous system (e.g., by measuring skin resistance, Vom Brocke and Liang, 2014). Vom Brocke and Liang (2014) derived three specific strategies of NeuroIS: (a) Strategy 1: inform the building and evaluation of IT artifacts; (b) Strategy 2: use of neuroscience tools to evaluate IT artifacts; (c) Strategy 3: use neuroscience tools as built-in functions of IT artifacts. In our paper, we focus on strategy 1 to inform the building and evaluation of CAs in online video lectures. Past research already used NeuroIS measurement methods for learner engagement. For example, Boucheix and Lowe (2010) used eye-tracking technology to determine the impact of different types of animation on learner engagement. Shen et al. (2009) used skin conductance, blood pressure and EEG sensors to measure learner's emotional engagement state while learning from interactive electronic lectures. In our research, we used facial expression analysis and galvanic skin response to measure learners' engagement level during the online video lectures.

Already implemented CAs in online learning may be able to put learners into a constructive engagement mode by helping them to reflect on the learning material (e.g., summarizing the content of a video). In contrast to that, we argue that dynamically scaffolding CAs might be able to trigger learners' interactive learning engagement mode by building up a scaffolding dialog with learners and not ignoring a partner's contribution (Chi and Wylie, 2014). These kinds of CAs are able to adapt to the contributions of the learner where the dialog is primarily constructive and a sufficient degree of turn-taking occurs. Thus, our hypothesis is as follows:

Dynamically scaffolding CAs built into online video lectures positively influence levels of learner engagement compared to statically scaffolding CAs.

3 Research Methodology

To test our hypothesis, we employed a repeated-measurement experiment design, where learners were able to interact with a dynamically or statically scaffolding CA while watching an online video lecture. All learners watched two online video lectures in programming.

3.1 Sample

We recruited undergraduate and graduate students (*n*=182, 74 female, 108 male, aged 18 to 35) from the lab pool of a major European university. Table 1 shows the sample's characteristics between the two treatment groups and the control group. We chose a smaller sample size (*n*=35) for the control group because our main interest lies in the differentiation between dynamically and statically scaffolding CAs.

	TG1 statically scaffolding CA	**TG2 dynamically scaffolding CA**	**TG3 Control Group**
Sample size	74	73	35
Gender	M 45, F 29	M 42, F 31	M 20, F 15
Average Age	22.75	22.90	22.20
Pre-experience with CAs in years	3.75	3.70	3.50

Personal Innovativeness (out of 7)	4.7	4.8	4.8
Language Level (Proficiency) (out of 6)	5.3	5.45	5.5
Nationality	BLINDED for REVIEW		

Table 1. Characteristics of the Sample

All learners were randomly assigned to one of the treatment groups (TG). Learners received 20 US dollars as a baseline incentive and 10 US dollars depending on their performance. The randomization was successful since treatment groups were similar in terms of *gender, age, previous experience with CAs in years, personal innovativeness, language level* and *nationality* ($p > 0.05$).

3.2 Experimental Design

We compared the dynamically scaffolding CA Sara with a statically scaffolding CA and included a control group. Learners watched two different video lectures where an CA interrupted the video after four segments. This results in four interaction points per video in which the learners interacted with one of the two types of conversational agents (except for the control group, where learners simply watched the video lectures). The students received the treatments and videos in random order. The statically scaffolding CA offered standardized hints in a way that learners could click to continue or answer a voice command. The dynamically scaffolding CA offered questions and allowed free student' answers. Sara was able to open a scaffolding sub-dialog when learners gave a *wrong* or *don't know* answer. Our dependent variable (DVs) is learner engagement, including pre-knowledge test scores as covariates, and learning channel preference and cognitive load as control variables.

3.3 Videos and Design of Conversational Agents

Every learner watched two online videos in random order, both dealing with the introduction to programming with Python. Both videos are from a famous python course taught by Charles R. Severance from the University of Michigan School of Information publicly available under py4me.com. We decided to choose this learning content for two reasons. First, we wanted to make sure that the learners had little prior knowledge of the topic. Since the university in question is a business school without a programming focus, the topic was suitable for this. Second, the reason why we specifically chose the Python theme and the learning video is that we wanted to ensure that the videos were representative in terms of quality, typicity and video format. The Python programming theme is representative for online learning materials because the amount of technically oriented learning content has increased enormously in recent years (Shah, 2019). The videos were in English, web-based and picture-in-picture, a format that is very common in today's MOOCs. Video 1 was about constants and variables and lasted 4 minutes and 52 seconds (without CA interaction). Learners learned the meaning and building of constants and variables including variable name rules. Video 2 was about conditional execution and lasted 4 minutes and 48 seconds (without CA interaction). The learners were taught the importance of conditional execution and how to create it. Every video starts with a 10-second resting period (black screen with white text "Resting Period"). The videos did not build on each other. The learner interacted with a CA four times per video. After one video segment was over, the video stopped and the CA automatically popped up in the same browser window where the CA started to interact with the learner. Figure 1 shows an exemplary excerpt of a learner dialog in TG1 (statically scaffolding) and TG2 (dynamically scaffolding). In the statically scaffolding condition, the CA provides one-size-fits-all standardized hints. Learners can decide to click or say next when they want to continue.

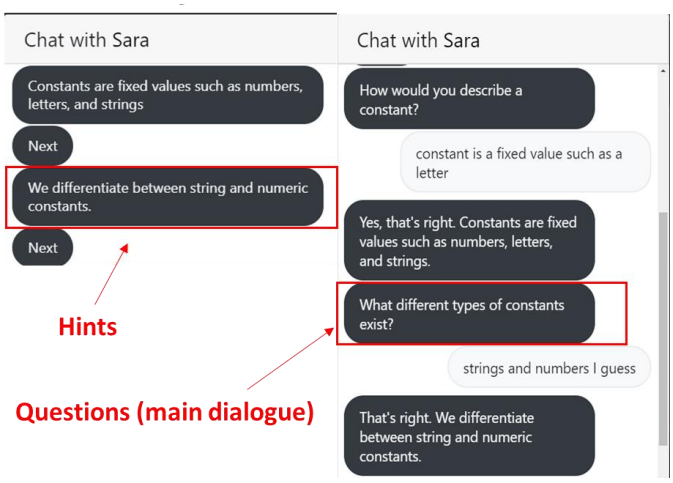

Figure 1. Difference between statically scaffolding CA (left) and dynamically scaffolding CA (right).

These hints can be considered as static scaffolds, because the CA is not able to adapt the answers to learners' reactions. In the dynamic scaffolding condition, the CA offers open questions that allow learners to provide a freely uttered answer by voice or text. The interaction logic is shown in Figure 2. The dynamically scaffolding CA Sara is able to detect a *correct*, a *wrong* or a *don't know* answer. Following the principle of dynamic scaffolding and knowledge construction dialogs (Rosé et al., 2003), we implemented a main and sub dialog interaction logic. In the main dialog, Sara asked the learner questions that should help to revise the content of the previous video segment (Text in green rectangles at the bottom). For example, "What does the following python code mean? If x<10: print('smaller')".

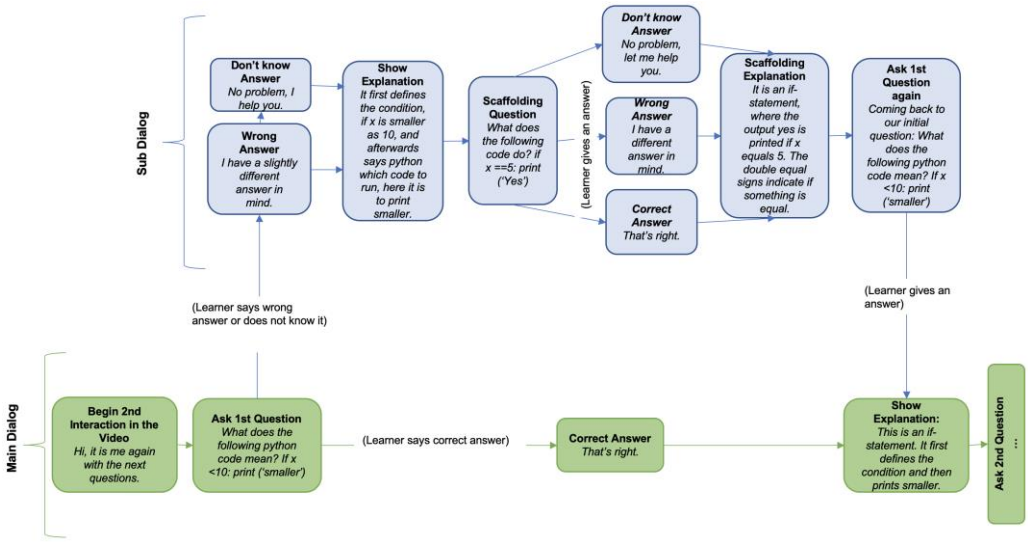

Figure 2. Interaction logic including main- and sub dialog.

Every time the learner did not know the answer or provided a wrong answer, Sara reacted to it and opened a sub-dialog, where she offered scaffolds that helped the learners to discover the right solution until she entered the main dialog again. For example, a sub-dialog scaffold can be an explanation followed by a similar example of the main question (see blue rectangles at the top). The hints in the static scaffolding condition and the questions of the main dialog in the dynamic scaffolding condition are information-equivalent. This means that hints and questions provide the same amount of information. The only difference is the dynamic scaffoldings in the sub dialog condition. The statically scaffolding CA was not able to react to different learner answers. From the technical side, we used pre-trained neural networks. We have used the publicly available NLP.js framework for intent classification, which has

proven effective in recent benchmarks (Git Hub, 2019). The NLP.js framework is available at https://github.com/axa-group/nlp.js. We seeded our NLP module with possible correct, wrong and don't know answers. For this purpose, we conducted 6 learner pretests with Sara and asked the learners to write down as many answer possibilities to Sara's question as they know. In addition, the research team reviewed the possible answers and added additional answers. All in all, we had a training set of 800 single statements. We calculated the intervention selection accuracy score for the dynamically scaffolding group. The intervention selection accuracy was defined by the number of correctly assigned scaffolding sub-dialogs (i.e., everytime the classifier tagged a student answer correctly as "wrong" or "don't know" and thus opened a scaffolding sub-dialog, see right side of Figure 1) divided by the total number of all sub-dialogs given. The research team tagged the interaction logs and found that in total, 51 out of 73 students entered the sub dialog part of the dialogue at least one time with 199 sub-dialogs in total (out of 1168 potentially possible sub-dialogs). 190 interventions out of 199 sub-dialog interventions were correct. This leads to an intervention selection accuracy of 95.48%. For voice recognition, we used the Web Speech API of HTML 5 and sent it to our NLP server. To ensure that the two videos were equally difficult for the learners and that possible effects on learner engagement were not due to differences in the difficulties of the two videos, we asked learners in the post-survey to rate the difficulty for both videos on a scale of 1 to 7. The learner ratings on the difficulties perceived of the videos were at the same level (video 1: 5.43 of 7, video 2: 5.63 of 7, $p > 0.05$).

3.4 Experimental Procedure

The experimental procedure is depicted in Figure 3.

Figure 3. The experimental procedure

After approval, the experimenter introduced the experiment to the learner and the learner started in room 1 with the pre-survey in absence of the experimenter. After completing the pretest, the experimenter returned and accompanied the learner into room 2, which contained a chair facing a computer monitor. The experimenter put on skin sensors on the fingers and made sure that face recognition was correctly calibrated. The experimental set-up is depicted in Figure 4. In the next step, the experimenter exited the room and instructed the learner to pay attention to the two lectures as there would be a short quiz on the material afterwards. The learner watched the stimulus without pauses and was not allowed to take notes. Once the lecture was over, the experimenter escorted the learner to room 1 again for conducting the post-survey. After completion, the learner was asked to guess the purpose of the study. She or he was then debriefed on the experiment. Most of the learners originally thought that the study was conducted to investigate how to effectively teach programming skills.

Figure 4. Experimental set-up

4 Measurement and Analysis

4.1 Neurophysiological Data

To measure our dependent variable learner engagement, we used two common neuroIS measurements, galvanic skin response and facial expression score.

4.1.1 Galvanic Skin Response Score

The Galvanic Skin Response is a short-term decrease of the electrical resistance of the skin caused by the typical increase of the sympathetic tone in emotional-affective reactions (Francis and Oliver, 2018). This leads to an increased secretion of sweat, corresponding to an increase in skin conductance (Lykken and Venables, 1971). Since any physiological arousal associated with emotion alters skin conductivity, measurements of electrodermal activity can be used to objectify psychophysiological relationships. There are a lot of studies that have used Galvanic Skin Response scores to measure learner engagement. For example, McNeal et. al's (2014) research measured learner engagement during a geology course using the Galvanic Skin Response Score. Before the learners were allowed to watch the videos, we attached shimmer™ finger sensors to their index, middle, and ring fingers, which allowed the skin level to be measured throughout the video. Once the sensor made skin contact, it automatically began collecting data. The data was then downloaded to the software iMotion™ to measure and analyze the skin conductance levels. Data were collected every 0.125s. The unit of measure for skin conductivity in the International System of Units is Siemens (Boucsein, 2013). We plotted the large dataset to get a first impression of the data and formed an average of the whole galvanic skin response score for both videos. We omitted the scores for the phases when learners were interacting with the CA because we aimed to measure the effect of CAs on learner engagement during the online videos. We standardized the scores between 0 and 1 to make it comparable with the facial expression score and compute a final engagement score.

4.1.2 Face Expression Score

In face coding, human emotions are measured by facial expressions. With facial expression analysis, it is possible to test the impact of content, products or services that are intended to evoke emotional excitement and facial reactions (Brackett and Katulak, 2007). One of the strongest indicators of emotions is the face expression. Different facial expressions such as laughter are associated with certain changes in important facial features. Computer-based facial expression analysis attempts to mimic human coding skills by capturing unfiltered emotional responses to any kind of emotionally appealing content. The facial expressions are captured by a camera and the resulting emotional states are analyzed by automated computer algorithms (iMotions, 2019). Once captured, face videos can easily be imported into a software for editing the facial expression analysis. We used the software iMotions™ to gather facial expression measurements. IMotion offers a cumulated face expression score that is calculated from the weighted sum of the following facial movements: brow raise, brow furrow, nose wrinkle, lip corner depressor, chin raise, lip pucker, lip press, mouth open, lip suck and smile (iMotions, 2019). The more often the system detects one of these movements, the higher the facial expression score is. We omitted the data for the phases when learners were interacting with the CA to ensure that these kinds of data are not falsifying the scores. We standardized the engagement score between 0 and 1 and computed a final engagement score, where galvanic skin response and face expression are weighted the same. Figure 4 shows an exemplary plot of a learner from the dynamically scaffolding treatment group with time (in minutes) on the x-axis, and galvanic skin response score (GSR) and face expression score on the y-axis. Furthermore, we added the corresponding CA interaction phases.

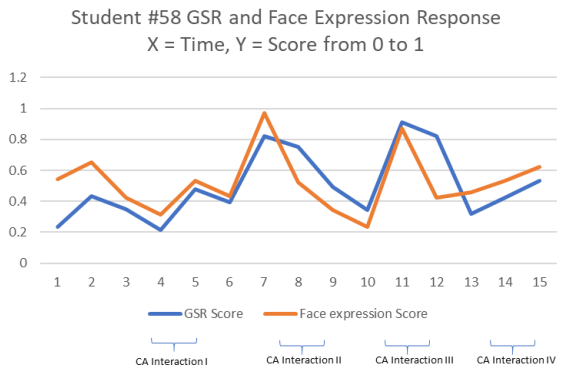

Figure 5. Learner #58 GSR and facial expression response for the dynamically scaffolding treatment group.

For data analysis, we first conducted an ANCOVA to identify differences between the groups (including control group). After that, we conducted a simple linear regression to test the difference between dynamic and static scaffolding. Since we only have to test one hypothesis, we do not need more sophisticated post-hoc tests. We conducted a test for normality, homogeneity of variance and homogeneity of regression slopes to check if our data meets the assumptions for using an ANCOVA. We used the statistic program R as a tool for analysis.

4.2 Qualitative Data

We used the qualitative data from the open question of the post-survey to gather a better understanding of the relationship between dynamically scaffolding CAs and learner engagement. We asked the learners the following question: How did you experience the interaction with the conversational agent? We 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 analyzed the data in a two-step process. First, we identified key words indicating learners' perceptions on the interaction with the CAs. We 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. Second, 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).

5 Results

RQ1: Do dynamically scaffolding conversational agents increase learners' engagement during online video lectures?

The main goal of our study is to investigate whether our dynamically scaffolding CA Sara is able to increase learner engagement during online video lectures compared to often implemented statically scaffolding CA types. To test our hypothesis, we conducted an ANCOVA with learner engagement as the dependent variable and the pretest score as a covariate. Table 2 shows the summarized engagement score including the galvanic skin response and the face expression score. All scores were standardized between 0 and 1. The engagement score is calculated from the average of galvanic skin response and facial expression.

Measurement	TG1 statically scaffolding CA	TG2 dynamically scaffolding CA	Control Group
I. Engagement Score	**0.31**	**0.42**	**0.31**
a. Galvanic Skin Response	0.40	0.46	0.33
b. Face Expression	0.21	0.38	0.28

Table 2. Average measurement of galvanic skin response and facial expression score

The galvanic skin response shows a higher value for dynamically scaffolding CAs (0.46) compared to statically scaffolding CAs (0.40) and the control group (0.33). The facial expression score shows a similar picture. Again, the group with the dynamically scaffolding CA shows the highest value (0.38) followed by the control group (0.28) and the statically scaffolding CA (0.21).

To check whether there are significant differences between the three groups, we conducted an ANCOVA and continued with a simple linear regression to see whether there are significant differences between statically and dynamically scaffolding CAs. All assumptions for ANCOVA and simple linear regression are met. We ran the ANCOVA with the treatment group as the independent variable and the pre-test score as a covariate. Results of the test indicate that there are significant differences between the groups ($F(2,360)$=46.327, p=4.5e-11, N=182). Moreover, the simple regression identifies that there is a significant difference between statically and dynamically scaffolding CAs (T=7.238, p=4.57e-12, *adjusted r2*=0.157, N=137). Thus, we can accept our hypothesis.

RQ2: How do learners perceive interactions with dynamically scaffolding Conversational Agents during online video lectures?

Based on our neurophysiological findings regarding the positive relationship between dynamically scaffolding CAs and learner engagement, we investigated how learners perceived the interactions with dynamically scaffolding CAs. The three main themes we identified from the open question in the post-survey were *emotional involvement*, *structured thinking*, and *voice usage*. The subtopics and the corresponding frequencies of the responses are presented in Table 3. Below, we shortly discuss the different themes.

	Total	**Statically scaffolding CA**	**Dynamically Scaffolding CA**
	N	Frequency	Frequency
I. Emotional Involvement			
- Was like talking with a real tutor	19	5	14
- Improves involvement a lot	10	2	8
- Did not feel under pressure compared to a teacher	4	1	3
II. Structured Thinking			
- Helped me to express the content in my own words	12	1	11
- Helped me to reflect about what was said in the video	11	4	7
- Helped me to improve my understanding of the concepts	5	1	4
III. Voice Usage			
- Surprisingly accurate	9	4	5
- Exciting to use voice	6	2	4
- Voice was cold and robotic	6	2	4

Table 3. Themes identified for the open-ended question asking learners about their experience when interacting with the CA.

5.1.1 Emotional Involvement

This theme relates to the capacities of dynamically scaffolding CAs being able to involve learners in a human-like interaction. Learners in the dynamically scaffolding SPA group appreciated that they were able to use free utterances when interacting with Sara and that the CA was able to adapt its answers accordingly. As one learner from the dynamically scaffolding CA group put it: “Useful, makes me think about what I have just seen. It improves involvement a lot. It was actually better than the video as listening to the guy and reading the slides at the same time can be annoying.” Other learners mentioned that it almost felt like having a conversation with a real tutor. For example, one learner mentioned: “Great experience, sometimes Sara was not able to understand me but all in all it felt like chatting with a real tutor.” And another learner said: “Chatting with Sara helped me to rethink the different video parts, almost like a human tutor.” For other learners, it felt like a kind of relationship. For example, one student said: “It was a human-machine relationship. Not so many emotions were transmitted but she

[CA Sara] did a good job." The coding of the themes showed that these kinds of statements about emotional involvement were mentioned much less in the static CA group.

5.1.2 Structured Thinking

This theme relates to the ability of learners to organize their own thinking structures. Many learners mentioned that the interaction with the CA helped them to revise the just learned content from the video segment and use their own words to repeat the content. For example, one learner stated: "Both videos helped me to think about what was said by the instructor. Voice-based was sometimes not able to understand my utterances. Tough, I preferred voice over text because it is more convenient and helped me to structure my thinking processes." Some learners also mentioned that they preferred the dynamically scaffolding CA compared to the statically scaffolding CA related to their structuring of thinking processes. For example, one learner commented: "I liked writing text instead of clicking next, because this way I can express the content in my own words."

5.1.3 Voice Usage

This theme relates to learners' experiences while interacting with the CA by voice. While some of the learners were excited about using voice to interact with the CA, others were a bit disappointed about the current state of the technology. The positive aspects learners mentioned were that they were surprised about the accurateness of the voice recognition and that talking things out loud helped them to learn more deeply. For example, one learner mentioned: "It was fascinating how much it understood and could say if I was correct. It helped me to express my own thoughts." Another learner mentioned: "The video with the voice was better because it forces me to think about the content rather than simply clicking next". Others felt a bit annoyed because technology seems to not be able to approach real human-human communication yet. For example, one learner commented: "When I could communicate in text it was easier for me to form the answers since I could make corrections, check my answers, and felt like I have less pressure while answering." Another learner mentioned: "Sometimes it didn't understand me correctly, so I had to adjust the way I was talking."

6 Discussion

This study aimed at investigating the effect of dynamically scaffolding CAs on learner engagement in an online video lecture setting. Online education creates new opportunities for learners to gain knowledge independent of place and time. However, the lack of learner engagement and the corresponding high attrition rates in online courses remain challenges in the field (Hew and Cheung, 2014). Based on scaffolding theory, we propose that the use of dynamically scaffolding CAs during online video lectures might be able to increase learner engagement compared to statically scaffolding CAs or no CAs. The neurophysiological measurements confirmed that dynamically scaffolding CAs are able to significantly increase learner engagement compared to statically scaffolding CAs and no CAs. Interestingly, the use of statically scaffolding CAs showed no differences to the control group.

One main reason for the positive relationship between dynamically scaffolding CAs and learner engagement might be that learners are more emotionally involved in the interaction (Theme 1). Dynamically scaffolding CAs are able to react to learners' utterances individually, similar to human-human communication. This mechanism might help learners to get more deeply involved in a conversation, resulting in higher learner engagement. This goes in line with the ICAP Framework proposed by Chi and Wylie (2014), which states that an interactive, dialoging learning mode is the best way of learning. This is also reflected by our qualitative findings indicating that the interaction with dynamically scaffolding CAs almost felt like having a conversation with an educator. When an educator interacts with a learner individually, he or she is able to adapt the answers to the learner. Compared to a statically scaffolding CA, the dynamically scaffolding CA might be able to behave more like the gold standard of the scaffolding behavior of an educator. This might be the reason for a deeper involvement in the course materials. The

finding is also supported by research in the area of computers-as-social-actors' paradigm, which explains that the more social cues an agent has, the more it feels like talking to a real human (Nass et al., 1994).

One other reason for the positive, quantitative relationship between dynamically scaffolding CAs and learner engagement might be that the individual interaction during video lectures helps learners to directly reflect about what they have seen and to express the received information in their own words. This goes in line with findings from Azevedo et al. (2004), who was able to show that learners conducted more self-regulating learning activities when receiving dynamic scaffolds. Most of the currently implemented CAs in online learning usually rely on rather standardized, one-size-fits-all hints to improve learning processes, which cannot capture learners' individual utterances (Adamson et al., 2014). For example, Song et al. (2017) developed a CA that helps learners reflect about their learning on a weekly basis, which was not able to fully adapt to the answers of the learner. Past research in the area of self-talking highlights the important role that articulation plays in the learning process (Lepadatu, 2012). For example, Lepadatu et al. (2012) confirmed that learners who talked out loud while doing a ball throwing exercise performed better. This can also be shown by our qualitative findings, where learners mentioned that voice input helped them to bring structure into their thinking (theme 2). While some learners were surprised at how well the CA understood them, others clearly showed that voice recognition systems are still reaching their limits when it comes to capturing free expressions from users (theme 3). This insight can also be confirmed by past research that deals with voice recognition systems (Grant et al., 2019).

Our work makes three main theoretical contributions and also has practical implications. First, we contribute to research in the area of online learning by empirically proving that dynamically scaffolding CAs built into video lectures can get learners into a more engaged learning mode. A more engaged learning mode is crucial for learning success in online learning since educators are not able to interact with each learner individually. Second, we contribute to NeuroIS research by showing how neurophysiological measurements can be used to understand the impact of conversational systems on learner engagement in online education. To better understand the problem of low engagement and corresponding high attrition rates in online courses, it is very important to use different measures of learner engagement. Third, we contribute to scaffolding theory by showing that scaffolding during the instruction phase is only beneficial when it is dynamic since our findings show that static scaffolding indicates no different effects compared to no scaffolding. In regard to its practical implications, this study showed CA developers and online course providers how to design and use CAs on top of already existing online video lectures. As illustrated by our study, publicly available NLP frameworks such as NLP.js can be powerful tools to create CAs that are able to classify intents and to offer dynamic scaffolds.

7 Limitation and Future research

There are a number of limitations to this study that should be noted. First, we measured learner engagement during the relative short duration of the video. It is very difficult to deduce from these results how CAs would have affected the overall learner engagement over the whole period of an online course. Nevertheless, we tried to measure learner engagement continuously with neurophysiological measures rather than a single-point self-reporting of the students. Future research should try to implement dynamically scaffolding CAs in online video lectures in real online learning environments in order to measure long-term effects on learner engagement. Second, learners might have been curious about the novel CA, which could have led to falsified results. 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. 50% used SPAs every day on their smartphones, e.g., Siri). Also, here, field experiments in real online learning environments might help to measure learner engagement over a longer period of time. This would help to see whether novelty plays a significant role. Last but not least, the lab experiment investigated the effect of dynamically scaffolding SPAs in a rather narrow context (programming in python). Future research should try to deploy similar experiments that show the importance of dynamically scaffolding CAs built into video lecture in similar contexts. For example, they could investigate the shown effect for different types of learning materials for the entire duration of an online course.

References

Adamson, D., G. Dyke, H. Jang and C. P. Rosé (2014). "Towards an agile approach to adapting dynamic collaboration support to student needs" *International Journal of Artificial Intelligence in Education* 24 (1), 92–124.

Ayedoun, E., Y. Hayashi and K. Seta (2015). "A Conversational Agent to Encourage Willingness to Communicate in the Context of English as a Foreign Language" *Procedia Computer Science* 60 (1), 1433–1442.

Azevedo, R., J. G. Cromley and D. Seibert (2004). "Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia?" *Contemporary Educational Psychology* 29 (3), 344–370.

Benware, C. A. and E. L. Deci (1984). "Quality of learning with an active versus passive motivational set" *American Educational Research Journal* 21 (4), 755–765.

Boucheix, J.-M. and R. K. Lowe (2010). "An eye tracking comparison of external pointing cues and internal continuous cues in learning with complex animations" *Learning and Instruction* 20 (2), 123–135.

Boucsein, W. (2013). *Elektrodermale Aktivität: Grundlagen, Methoden und Anwendungen:* Springer-Verlag.

Brackett, M. A. and N. A. Katulak (2007). "Emotional intelligence in the classroom: Skill-based training for teachers and students" *Applying emotional intelligence: A practitioner's guide*, 1–27.

Bull, K. S., P. Shuler, R. Overton, S. Kimball, C. Boykin and J. Griffin (1999). "Processes for Developing Scaffolding in a Computer Mediated Learning Environment".

Cassell, J. (2000). "EMBODIED CONVERSATIONAL INTERFACE AGENTS" *Communications of the ACM* 43 (4), 70–78.

Chi, M. T. H. and R. Wylie (2014). "The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes" *Educational Psychologist* 49 (4), 219–243.

Dimoka, A., F. D. Davis, A. Gupta, P. A. Pavlou, R. D. Banker, A. R. Dennis, A. Ischebeck, G. Müller-Putz, I. Benbasat and D. Gefen (2012). "On the use of neurophysiological tools in IS research: Developing a research agenda for NeuroIS" *MIS quarterly*, 679–702.

Ferguson, R. and D. Clow (eds.) (2015). *Examining engagement. Analysing learner subpopulations in massive open online courses (MOOCs):* ACM.

Figner, B. and R. O. Murphy (2011). "Using skin conductance in judgment and decision making research" *A handbook of process tracing methods for decision research*, 163–184.

Francis, A. L. and J. Oliver (2018). "Psychophysiological measurement of affective responses during speech perception" *Hearing research* 369, 103–119.

Git Hub (2019). *NLP Framework BenchmarKING*. URL: https://github.com/axa-group/nlp.js/blob/master/docs/benchmarking.md.

Glasersfeld, E. v. (1987). "Constructivism" *The concise Corsini encyclopedia of psychology and behavioral science* 6, 19–21.

Graesser, A. C., Z. Cai, B. Morgan and L. Wang (2017). "Assessment with computer agents that engage in conversational dialogues and trialogues with learners" *Computers in Human Behavior* 76, 607–616.

Graesser, A. C., K. Vanlehn, C. P. Rosé, P. W. Jordan and D. Harter (2001). "Intelligent tutoring systems with conversational dialogue" *AI Magazine* 22 (4), 39.

Grant, R. H., T. L. Hewitt, M. J. Mason, R. J. Moore and K. A. Winburn (2019). *Cognitive intervention for voice recognition failure:* Google Patents.

Griol, D., J. M. Molina and Z. Callejas (2017). "Combining speech-based and linguistic classifiers to recognize emotion in user spoken utterances" *Neurocomputing*.

Gulz, A., M. Haake, A. Silvervarg, B. Sjödén and G. Veletsianos (2011). "Building a Social Conversational Pedagogical Agent". In D. Perez-Marin and I. Pascual-Nieto (eds.) *Conversational Agents and Natural Language Interaction*, pp. 128–155: IGI Global.

Halverson, L. R. and C. R. Graham (2019). "Learner Engagement in Blended Learning Environments: A Conceptual Framework" *Online Learning* 23 (2), 145–178.

Hew, K. F. and W. S. Cheung (2014). "Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges" *Educational Research Review* 12, 45–58.

iMotions (2019). *Facial Expression Analysis*. URL: https://imotions.com/biosensor/fea-facial-expression-analysis/.

Johnson, R. B. and A. J. Onwuegbuzie (2004). "Mixed methods research: A research paradigm whose time has come" *Educational researcher* 33 (7), 14–26.

Johnson, W. L., J. W. Rickel and J. C. Lester (2000). "Animated pedagogical agents: Face-to-face interaction in interactive learning environments" *International Journal of Artificial Intelligence in Education* 11 (1), 47–78.

Kerly, A., P. Hall and S. Bull (2007). "Bringing chatbots into education. Towards natural language negotiation of open learner models" *Knowledge-Based Systems* 20 (2), 177–185.

Kerry, A., R. Ellis and S. Bull (2009). "Conversational Agents in E-Learning". In T. Allen, R. Ellis and M. Petridis (eds.) *Applications and Innovations in Intelligent Systems XVI*, pp. 169–182. London: Springer London.

Lederman, D. (2018). *Online Education Ascends*. URL: https://www.insidehighered.com/digital-learning/article/2018/11/07/new-data-online-enrollments-grow-and-share-overall-enrollment.

Lepadatu, I. (2012). "Use self-talking for learning progress" *Procedia - Social and Behavioral Sciences* 33, 283–287.

Li, H. and A. Graesser (2017). "Impact of Pedagogical Agents' Conversational Formality on Learning and Engagement". In E. André, R. Baker, X. Hu, M. M. T. Rodrigo and B. Du Boulay (eds.) *Artificial intelligence in education. 18th International Conference, AIED 2017, Wuhan, China, June 28-July 1, 2017 : proceedings*, pp. 188–200. Cham: Springer (visited on 11/11/2019).

Lim, C. P. (2004). "Engaging learners in online learning environments" *TechTrends* 48 (4), 16–23.

Lu, Y., C. Chen, P. Chen, X. Chen and Z. Zhuang (2018). "Smart Learning Partner: An Interactive Robot for Education". In C. Penstein Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren and B. Du Boulay (eds.) *Artificial Intelligence in Education*, pp. 447–451. Cham: Springer International Publishing.

Luger, E. and A. Sellen (2016). "Like Having a Really Bad PA: The Gulf between User Expectation and Experience of Conversational Agents" *In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, Ca, USA: ACM*, 5286–5297.

Lykken, D. T. and P. H. Venables (1971). "Direct measurement of skin conductance: A proposal for standardization" *Psychophysiology* 8 (5), 656–672.

McNeal, K. S., J. M. Spry, R. Mitra and J. L. Tipton (2014). "Measuring student engagement, knowledge, and perceptions of climate change in an introductory environmental geology course" *Journal of Geoscience Education* 62 (4), 655–667.

Molenaar, I., C. Roda, C. van Boxtel and P. Sleegers (2012). "Dynamic scaffolding of socially regulated learning in a computer-based learning environment" *Computers & Education* 59 (2), 515–523.

Nass, C., J. Steuer and E. Tauber (eds.) (1994). *Computers are social actors:* ACM.

Panzoli, D., A. Qureshi, I. Dunwell, P. Petridis, S. de Freitas and G. Rebolledo-Mendez (2010). "Levels of Interaction (LoI): A Model for Scaffolding Learner Engagement in an Immersive Environment". In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, V. Aleven, J. Kay and J. Mostow (eds.) *Intelligent Tutoring Systems*, pp. 393–395. Berlin, Heidelberg: Springer Berlin Heidelberg.

Raad, B. de (2000). *The Big Five Personality Factors: The psycholexical approach to personality*. Ashland, USA: Hogrefe & Huber Publishers.

Richardson, J. C., Y. Maeda, J. Lv and S. Caskurlu (2017). "Social presence in relation to students' satisfaction and learning in the online environment: A meta-analysis" *Computers in Human Behavior* 71, 402–417.

Rosé, C. P., D. Bhembe, S. Siler, R. Srivastava and K. Vanlehn (2003). "Exploring the effectiveness of knowledge construction dialogues" *Artificial intelligence in education: Shaping the future of learning through intelligent technologies*, 497–499.

Ruan, S., L. Jian, J. Xu, B. Joe-Kun Tham, Z. Qiu, Y. Zhu, E. L. Murnane, E. Brunskill and J. A. Landay (2019). "QuizBot: A Dialogue-based Adaptive Learning System for Factual Knowledge". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Ed. by Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems: ACM.

Rubin, V. L., Y. Chen and L. M. Thorimbert (2010). "Artificially intelligent conversational agents in libraries" *Library Hi Tech* 28 (4), 496–522.

Ryan, G. W. and H. R. Bernard (2003). "Techniques to identify themes" *Field methods* 15 (1), 85–109.

Shah, D. (2019). *By the Numbers: MOOCs in 2018*. URL: https://www.classcentral.com/report/mooc-stats-2018/.

Shen, L., M. Wang and R. Shen (2009). "Affective e-learning: Using "emotional" data to improve learning in pervasive learning environment" *Journal of Educational Technology & Society* 12 (2), 176–189.

Shernoff, D. J., M. Csikszentmihalyi, B. Schneider and E. S. Shernoff (2014). "Student engagement in high school classrooms from the perspective of flow theory". In *Applications of flow in human development and education*, pp. 475–494: Springer.

Siadaty, M., D. Gašević and M. Hatala (2016). "Measuring the impact of technological scaffolding interventions on micro-level processes of self-regulated workplace learning" *Computers in Human Behavior* 59, 469–482.

Sinatra, G. M., B. C. Heddy and D. Lombardi (2015). *The challenges of defining and measuring student engagement in science:* Taylor & Francis.

Song, D., E. Y. Oh and M. Rice (2017). "Interacting with a conversational agent system for educational purposes in online courses". In: *2017 10th International Conference on Human System Interactions (HSI). Proceedings : International Hall, University of Ulsan, Ulsan, Republic of Korea, July 17-19, 2017*. Piscataway, NJ: IEEE, pp. 78–82.

Torrance, H. (2012). "Triangulation, respondent validation, and democratic participation in mixed methods research" *Journal of mixed methods research* 6 (2), 111–123.

Vanlehn, K. (2011). "The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems" *Educational Psychologist* 46 (4), 197–221.

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

Vom Brocke, J. and T.-P. Liang (2014). "Guidelines for neuroscience studies in information systems research" *Journal of Management Information Systems* 30 (4), 211–234.

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

Winkler, R., M. Söllner, Neuweiler Maya Lisa, F. C. Rossini and J. M. Leimeister (2019). "Alexa, can you help us solve this problem? How conversations with smart personal assistant tutors increase task group outcomes". In: *CHI'19 Conference on Human Factors in Computing Systems Extended Abstract*. Ed. by CHI'19 Conference on Human Factors in Computing Systems Extended Abstract: SIGCHI, pp. 1–6.

Wood, D., J. S. Bruner and G. Ross (1976). "The role of tutoring in problem solving" *Journal of child psychology and psychiatry* 17 (2), 89–100.