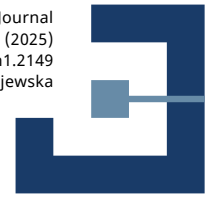
The JALT CALL Journal
ISSN 1832-4215, 21(1), 2149 (2025)
https://doi.org/10.29140/jaltcall.v21n1.2149



Castledown

Using Alexa in technology-supported task-based learning to mediate EFL texts: a pilot study

Klaudia Gajewska
Maria Curie-Sklodowska University in Lublin, Poland
klaudia.gajewska@mail.umcs.pl

Given the novelty of human–machine communication (HMC) in task-based language learning (TBLL), we conducted a multimethodological pilot study to explore the use of Alexa in out-of-class task-based learning to mediate texts in English as a foreign language. Fourteen Polish intermediate learners participated in the quasi-experiment and programme evaluation. The treatment was investigated with reference to its impact on the participants' mediating skill, their self-perceived acceptance of the technology, and the HMC process. Even though the statistical analysis of the pre- and post-tests revealed that the intervention had no effect on the students' ability to mediate texts, the findings from quantitative and qualitative analyses of the learners' logs, the Alexa app, post-study questionnaires, and interviews were positive. Despite HMC breakdowns, the learners kept interacting with the technology by producing modified output through rephrasing and had very high opinions of the use of the tool in learning to mediate. The results enhance the understanding of the new phenomenon of voice assistance and its application in TBLL; however, longer-term research on the implementation of Alexa (and the like) is required in view of the limitations of the current research related to the short duration of the treatment and small sample size.

Keywords: Alexa, voice assistant (VA), human–machine communication, mediation, task-based language learning

Introduction

The phenomenon of speech-based human–machine communication (HMC), standing here for "the creation of meaning between humans and machines, with technology theorized as a communicator" (Guzman, 2018, p. 2), attracted public attention in the 2010s when Siri, Google Assistant, Alexa, Cortana, and Celia, defined as "cloud-based virtual assistants" (Dizon, 2021, p. 1), entered

the mainstream mobile industry in the form of voice-controlled software and apps. Programmed to understand voice commands (Adamopoulu & Moussiades, 2020) and "vocally respond to human questions and requests" (Guzman & Lewis, 2020, p. 72), these voice assistants (VAs) have been enthusiastically applied in a variety of settings (e.g., home, car, outdoors, work)(Social Lens Research, 2018) for a variety of purposes (e.g., communication, home and media control)(Gillett, 2020), with convenience, enjoyment, the possibility to automate time-consuming activities through voice, and inclusiveness for people with disabilities being the prime motives accounting for the technology's uptake (Adobe Inc., 2019).

Speech-based HMC has quickly generated research interest in dialogue-based computer-assisted language learning (CALL), with child, teenage, and adult language learners' interaction with VAs having been subjected to preliminary examination in different English as a second/foreign language (ESL/EFL) educational settings (e.g., Canada, Japan, Spain, Taiwan) and types of instruction (formal or informal) over the last few years (e.g., Chen et al., 2020; Dizon, 2017; Moussalli & Cardoso, 2016; Tai & Chen, 2022). Despite the questionable reliability of the information they retrieve from the internet (Underwood, 2021), their failure to efficiently process nonstandard language input (Kukulska-Hulme, 2019), and concerns around security and privacy breaches (Terzopoulos & Satratzemi, 2019), VAs have captured researchers' attention, given the technology's role in enriching the quantity and quality of language input through multimodal responses (Moussalli & Cardoso, 2019) and extending traditional classroom language practice to personalised and stress-free contexts (Kukulska-Hulme & Lee, 2020) by promoting learners' interactive language use off-site through their negotiation of meaning (Dizon, 2017), exposure to authentic language input (Kim, 2018), and reception of immediate feedback on accuracy (Underwood, 2021).

Notwithstanding its clear educational advantages, the implementation of VA poses basic problems in language learning. The relationship between the technology's use and the development of learners' language skills, the HMC process that occurs during their interaction with the technology, and the formulation of pedagogical principles underlying the VA's application are just a few issues that have been understudied and underanalysed in CALL. Equally importantly, since the implementation of Web 2.0 tools and examinations of computer-mediated communication have been high on the task-based language learning (TBLL) research agenda (Gonzales-Lloret, 2015), the integration of voice-assisted technologies, such as VAs, with tasks and investigation of HMC have remained *terra incognita*. Even though TBLL has been considered a perfect setting for exploiting the potential of technological innovations in language learning (Gonzales-Lloret, 2017; Ziegler, 2016), as demonstrated by the results of empirical studies indicating the greater effectiveness of CALL modes of TBLL over conventional TBLL in enhancing participants' language skills (e.g., Abdallah & Mansour, 2015; Fang et al., 2021; Mulyadi et al., 2021; Widiastuti et al., 2022) thanks to the personalisation (Tavakoli & Loth, 2021), attractiveness (Kruk, 2011), authenticity (Mulyadi et al., 2021), and multimodality (Shokrpour et al., 2019) underlying technology-supported language learning environments,

it is not clear what the effects of integrating tasks and VAs on the development of learners' language skills would be, how students would communicate with the technology during performance of the task, and what their perceptions of the technology's application in task-based learning would be.

Given the abovementioned research gaps, we carried out a small-scale pilot study in order to provide preliminary findings on the novel combination of dialogue-based CALL and TBLL. Specifically, we aimed to examine the phenomenon of speech-based HMC occurring during TBLL to mediate texts by (1) measuring the effects of using Alexa on Polish teenage learners' ability to mediate EFL texts, (2) examining the case of speech-based HMC, and (3) exploring the participants' views on using the technology when performing out-of-class mediation tasks. To the authors' best knowledge, the current study is the first attempt made in the field to report the research results of integrating voice-assisted technology (in this case, Amazon's Alexa) and task performance. This, despite the main limitation of the current research procedure (i.e., limited sample size), allows us to shed some light on the innovative application of VAs in TBLL.

Literature review

A large part of dialogue-based CALL research on VAs has focused on learners' opinions of voice assistance and has generally revealed very positive findings, with the participants strongly emphasising self-perceived linguistic and affective benefits accruing from the implementation of voice-controlled technology. In the study conducted by Moussalli and Cardoso (2016) in Canada, four ESL students considered Alexa to be user-friendly, enjoyable, and useful in the development of their target language. The findings of Dizon's (2017) study involving four Japanese EFL learners were similar, in that the students had a high opinion of Alexa because of the speaking and pronunciation practice the technology promotes. Despite HMC breakdowns, the use of Alexa, Siri, and Google Assistant also received warm approval from 11 child learners, who described their interaction with the voice recognition systems as very pleasurable, in a nine-month study conducted in a Spanish EFL context by Underwood (2017). Similarly, Japanese university students pointed out the twofold nature of HMC occurring during their interaction with Alexa, expressing satisfaction with the technology use and identifying the problem of comprehensibility issues related to accented speech, as indicated by Dizon and Tang (2019, 2020).

Research attempts have already been made with the intention of measuring the reliability of various VAs in comprehending the ESL/EFL speech of language learners, with the preliminary studies revealing inconsistent findings about the technology's capacity to correctly understand accented utterances. Daniels and Iwago's (2017) examination, in which Google Assistant and Siri were compared in the context of their ability to transcribe the language input of 41 Japanese ESL learners, suggested that the mean accuracy scores of the two were rather low, ranging from 82% to 66% respectively. Dizon (2017) revealed that Alexa reached an even lower accuracy score of 50% when its transcriptions were

compared with the spoken commands of four Japanese EFL learners. In their follow-up study, Moussalli and Cardoso (2019) reported that Alexa succeeded in understanding the 11 ESL learners 83% of the time, yet it fell behind the human raters, who were able to transcribe 95% of the learners' utterances.

The analysis of HMC occurring during learners' interaction with VAs with regard to the strategies implemented by the participants to handle communication breakdowns has also come into the focus of research. On the one hand, the findings have been consistent, indicating that repetition, reformulation, and abandonment were normally used (Chen at al., 2020; Dizon & Tang, 2020; Moussalli & Cardoso, 2019). On the other hand, depending on the study and the research procedure adopted therein (video-recording of classroom VA use versus using VA in autonomous ESL learning), discrepancies in the frequency of the selected strategies implementation arose, with repetition (43%)(Moussalli & Cardoso, 2019) and abandonment (63%)(Dizon & Tang, 2020) being used most often by the participants.

The effects of VA applications on the linguistic and affective aspects of language learning have already been submitted to scientific scrutiny in a handful of experimental research studies, whose findings revealed a positive impact of using the technology. Dizon (2020), who investigated the influence of HMC on 13 Japanese EFL students' listening and speaking development in a 10-week long treatment, demonstrated that the Alexa application led to significant improvements in the learners' speaking skills. In their 10-week quasi-experimental study, Tai and Chen (2022) revealed that Google Nest Hub enhanced the participants' listening comprehension development thanks to its multimodal responses. In their two-week study conducted with a group of 112 Taiwanese high school students, Tai and Chen (2020) evaluated the influence of HMC on participants' willingness to communicate. The experimental group of students, who performed activities with Google Assistant, reported lower levels of anxiety and higher levels of confidence. Similarly, the findings of the study undertaken by Chen et al. (2020) on 29 Taiwanese university students who communicated with Google Assistant indicated the students' increased motivation and reduced foreign language anxiety.

These findings, however, can be called into question because of the methodological limitations of the research procedures underlying the studies mentioned above. Their first weakness is the small sample size, with the selected studies involving only two (Dizon & Tang, 2019) and four (Dizon, 2017; Mousalli & Cardoso, 2016) participants. The second problem relates to their short duration, limited, for example, to a single session lasting from 20 to 60 minutes (Chen et al., 2020; Dizon, 2017; Moussalli & Cardoso, 2016). Other controversial methodology-related issues arise from the invasive data collection procedures (Moussalli & Cardoso, 2019), neglect of the triangulation principle (Tai & Chen, 2020, 2022), and nonrandom sampling (Dizon, 2017, 2020).

In view of the above, the current pilot study was conducted to overcome some of the limitations underlying the preliminary VA research (e.g., the lack of triangulation of sources and methods) and fill gaps in TBLL, as the process of HMC during task performance and its effects on language learning have

not been examined, by carrying out a mixed-method investigation of the HMC established between Polish EFL learners and Alexa in out-of-class task-based text mediation practice. Given its research and pragmatic value, this multi-methodological examination is necessary in view of the research gaps and isolated cases of out-of-class interaction of foreign language learners in the target language. For these reasons, the study sought answers to three research questions (RQs):

1. What is the effect of the participants' interaction with Alexa when performing the task on their ability to mediate texts?
2. How do the participants handle communication with Alexa when performing the task?
3. How do the participants assess the application of Alexa when performing the task?

Methodology

The aim of this small-scale pilot project was to address the CALL and TBLL research gaps by investigating Polish EFL learners' interaction with Alexa when performing a mediation task. Since no such an endeavour has been made in either of the fields so far, the current study, despite its serious limitation resulting from the small sample size, allows us to gain a broader perspective on the use of VA, namely its application in TBLL to mediate texts.

Research design

A mixed-methods approach to research design involving the quantitative and qualitative methods of data collection and analysis (Dörnyei, 2007) was followed in the pilot study to ensure its validity through methodological triangulation (Fraenkel et al., 2012), inform the process of instrument design, and widen the scope of the treatment evaluation through product and process assessment (Greene et al., 1989).

The quasi-experimental research was undertaken to answer RQ1 by examining the cause-and-effect relationships between the independent and dependent variables (Konarzewski, 2000), namely CALL-mediated TBLL and the participants' mediating skills, respectively. The factor rotation technique was employed, under the influence of which, the quasi-experiment was divided into two phases to account for the group differences inherent in intact grouping (Sirotova et al., 2021). In Phase 1, the impact of the experimental intervention was measured in Group A, which was the experimental group (ExG), whereas Group B served as the control group (CtrlG). In Phase 2, the groups switched.

The descriptive research was conducted in response to RQ2 and RQ3. The convergent parallel mixed-methods approach was adopted (Creswell, 2014), which means that both qualitative and quantitative data on HMC were collected and analysed irrespective of each other to verify the findings through triangulation (QUAL + QUAN)(Dörnyei, 2007). The explanatory sequential mixed-methods approach was taken (Creswell, 2014) in line with the QUAN + QUAL

technique (Dörnyei, 2007) to investigate the learners' views on using Alexa during text mediation practice. The quantitative results were used to select the participants who were interviewed during the qualitative phase, which helped the author transcend the limitations of quantitative research.

Participants

Two groups of Polish intermediate EFL learners (n = 16) attending a private secondary school in Lublin, Poland, participated in the pilot study. They were selected according to the convenience sampling strategy and then randomly assigned to Group A or Group B. The two groups were similar with regard to a number of significant aspects, as evident in their responses to the background questionnaire and the mean scores for their pre-tests (see Table 1).

Table 1. Participants in the pilot study.

	Group A	Group B
Number of participants	7	9
Number of EFL classes a week	5	5
Duration of EFL education	11.22	11.41
Their EFL proficiency level	B1	B1
Number of students with (very) positive attitudes to EFL learning	6	7
Number of students with (very) frequent out-of-class use of English	6	8
Number of students attending private EFL tuition	7	9
Mean score for mediation pre-tests	5.17	5.25

Oral consent from the school administration to carry out the pilot study was obtained before the treatment began. To reduce the Hawthorne effect, the students were not informed about their participation in the quasi-experiment until it was performed. Once the information about the study had been disclosed, they could give or refuse consent for their data to be subjected to analysis. In total, 14 students agreed to participate in the study by giving their oral consent to have their data examined in the current investigation.

Treatment

The action-oriented approach (Council of Europe, 2001) was combined with technology-facilitated oral homework (Shanks, 2021) to enhance Polish EFL learners' ability to mediate EFL texts by engaging them in action under the guise of performing real-life tasks (Piccardo et al., 2011) in low-stakes home surroundings. In keeping with the TBLL theory (Long, 2005; Gonzales-Lloret, 2014, 2015) and the backwards design principle (Howell, 2017), authentic tasks ensuring relevance and focus on meaning (Willis, 2004) were selected on the basis of the multimethodological analysis of the participants' needs, which, among others, involved data collection methods such as interviews, written retrospection, and questionnaires.

A set of tasks covering the topics of travelling and popular culture, to be performed by the students on a weekly basis, was developed for the ExG and CtrlG (consisting of two tasks each), which were based on the results of the needs questionnaire. The text mediation concept (i.e., relaying sets of instructions and information from announcements and texts) (Council of Europe, 2020) was operationalised in the treatment design. The participants were asked to take on the role of language mediators, which required them to present selected pieces of information from the source texts in the target texts (Coste & Cavalli, 2015) to help imaginary interlocutors (i.e., friends, relatives) overcome physical, technical, and linguistic difficulties in gaining access to the source texts (Stathopoulou, 2015). There were either six or seven gaps in each target text (for example, see Figure 1).

Step 1: Read the instructions carefully

- You're on holidays in Rome. Your English-speaking friend from Manchester has always dreamt of visiting Italy and now wants you to tell her more about this destination (Text A). Ask Alexa for help in order to answer her questions. Using the information presented by Alexa fill in the gaps in Text B.

Step 2: Do the mediation task by filling in the gaps in Text B

- **Text A**: Hello! I've heard from Joanne that you're spending your winter break in Italy. I'd like to be there too, but I now have to study to retake my exam. Tell me at least a few things about this beautiful country. How many people live there? What countries does Italy border? I'm pretty certain that one of the neighbouring countries must be Switzerland. I remember that from our geography classes. LOL! Are there many mountain ranges in Italy? I know that the Italian flag is green, white and red, but I've no idea why it is so. Do you know anything about it? In addition to the Colosseum, what else is Italy famous for? By the way, have you seen any famous landmarks? Don't forget to upload some photos. I'm going to plan my summer holidays soon – can you tell me what the best time of the year is to visit Rome? Hugs, Betty.

- **Text B:** Hi Betty! I feel wonderful here in Italy. I hope you're fine there in Manchester too. Good luck with your exam preparations! Let me tell you a few facts about Italy. The population of Italy is estimated to be (1)..... You're right that Italy borders Switzerland. It also borders (2)..... Almost eighty percent of the country is mountainous, with (3) .....being the highest peak. There are many theories about the origins of the flag. One says that the colours represent the (4) .... In addition to the Colosseum that you've mentioned, some other popular attractions in Rome include (5) ..... It's been warm in Italy recently, but from what I've heard the best time to visit Rome is (6) ..... This evening I'll upload some photos on WhatsApp! I must go now. XOXO

Step 3: Answer the questions

- Are you satisfied with Alexa's responses to your questions? If not, try to do the task again.

Figure 1. Task 1 designed for the ExG by the author

There was a fundamental difference between the procedure of performing the task applied in the ExG and CtrlG. In the ExG, the discourse environment was limited to Alexa. The ExG students were involved in language reception, interaction, and production, as they were expected to read the target text, identify

the kind of information missing (e.g., date, temperature, name, etc.), negotiate meanings with Alexa by asking the VA an appropriate question, and fill in the gaps in the target text with the missing information. The CtrlG students did not use the VA, as semi-authentic materials (i.e., YouTube videos, blog posts, travel websites) provided the basis for the source texts in the two tasks targeting listening and reading. As opposed to the ExG, the controls were only involved in language reception and production.

Research instruments

Three written tests with a similar level of difficulty (B1+/B2) were designed to gather data in response to RQ1. Each of the tests consisted of two text mediation activities. They involved the participants in presenting selected information from the spoken and written source texts in writing based on the Companion Volume's (Council of Europe, 2020) "Relaying specific information" descriptor, which is the same descriptor underlying the task design process to ensure content and construct validity (Brown, 2004). Group A and B took each version of the test only once, in accordance with the split-block testing scheme, which was followed to control the test practice effect (Kruk, 2015).

Data from the learners' logs and the Alexa app were collected to triangulate the findings related to RQ2. The logs consisted of three close- and two open-ended questions to elicit learners' responses to how they assessed HMC. In Question 1 (Q1) and Question 2 (Q2), the students evaluated their interaction with Alexa using a six-point Likert scale from negative (1) to positive (6) (i.e., from *very unsuccessful* to *very successful*) and justified their decision. Next, in Question 3 (Q3), they were asked to decide whether they had experienced HMC breakdowns by selecting one of the options provided (i.e., either *yes* or *no*). If their answer was affirmative, they needed to assess the frequency of HMC breakdowns using a six-point Likert scale (i.e., from *very rare* to *very frequent*) in Question (Q4). In the Question 5 (Q5), they listed their reactions to HMC breakdowns. To gain a proper perspective on the HMC process within the treatment, the author triangulated the data from the learner logs by asking the ExG to provide her with their login and password to the Alexa app, which contained the spoken commands that the participants had given the VA and the responses that the technology made.

To answer RQ3, two methods of data collection were used. First, the participants completed the questionnaire, which was developed to assess their perceptions of using Alexa for learning to mediate in EFL. The 12 items to the questionnaire were originally written in English by Dizon and Tang (2019, 2020) and then translated into Polish by the author to prevent comprehension problems. Students' acceptance of Alexa, standing here for the self-perceived usability (Statement 1–4) and effectiveness of Alexa in learning to mediate (Statements 5–8) as well as participants' satisfaction with the technology (Statement 9–12) was measured according to the technology acceptance model (TAM) of Davis (1989). A six-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (6) was used for that purpose. In addition, an interview guide prepared

by the author comprising five questions relating to the students' acceptance of Alexa based on the TAM (Davis, 1989) and students' self-perceived advantages and disadvantages of using the VA was used during the semi-structured interviews (SSIs).

Data collection and analysis

The one-month pilot study began with the administration of the background questionnaire and pre-tests (T_1 and T_2) (see Figure 2). In the following two weeks, Group A and Group B took on the status of the ExG and CtrlG, respectively. A week prior to the introduction of the treatment, the ExG received the guide on how to create an Amazon account and install the Alexa app. The document was written in Polish to avoid comprehension problems. The controls received only two text mediation tasks that they were assigned to complete on a weekly basis. In Week 2, the second round of text mediation tests was administered to Group A (T_3) and Group B (T_4). Once the tests had been conducted, the groups switched their roles, and the same procedure was reimplemented with Group A as the CtrlG and Group B as the ExG. In the last week, the final round of mediation tests took place in Group A (T_5) and Group B (T_6). The data collected in the text mediation tests were subjected to quantitative analysis. The mean, median, and standard deviation were calculated to select an appropriate method of statistical analysis and measure the impact of the treatment on the participants' mediating skill at the within- and between-group levels (RQ1). The 0.05 level of statistical significance was chosen.

To examine HMC (RQ2), the ExG students filled in learners' logs after completing two mediation tasks. Mean, median, mode, standard deviation, and minimum and maximum values were calculated for three close-ended questions (Q1, Q3, Q4) in addition to the percentage of scores indicating agreement. The students' answers to open-ended questions were analysed by the author in accordance with the principles of content analysis (Krippendorf, 2019), categorised into themes with Taguette, and reported in a narrative form. To analyse the learners' usage data from the app provided by four participants who agreed to share their login and password, the procedure described by Dizon and Tang (2020) was adopted. Once the commands had been counted, their recordings were compared with the transcriptions: a match between their audio and text versions was considered a HMC success. A situation in which the users did not receive a piece of information that they had asked for and the transcription deviated from the recording was regarded as a failure. Students' responses to HMC breakdowns were categorised into "repeat", "rephrase", and "abandon" according to Moussalli and Cardoso's (2019) classification.

Evaluation of the programme was conducted in line with Norris' (2009) TBLL framework, modified by the author. One class after the students had taken the post-tests, they filled out the questionnaire. Their responses were analysed quantitatively and used to select two proponents and two opponents of Alexa, who were requested to participate in the SSIs in line with the purposive maximum variation sampling to establish a more comprehensive picture

of voice assistance in TBLL (Etikan at al., 2015). Four online meetings were scheduled to take place via the Zoom Cloud Meetings program. They lasted for approximately 20 minutes and were recorded with Apowersoft with the oral consent of the interviewees. The language used during the SSIs was Polish. Qualitative content analysis, "a research method for subjective interpretation of the content of text data through the systematic classification process of coding and identifying themes or patterns" (Hsieh & Shannon, 2005, p. 1278), was used. An inductive approach to data analysis was adopted, which means that the themes were directly taken from the data "without preconceived notions and categories" (Kondracki et al., 2002, p. 225). The qualitative data were categorised with the assistance of Taguette.

Figure 2. Stages of the pilot project

Results

RQ1: What is the effect of the participants' interaction with Alexa when performing the task on their ability to mediate texts?

The mean scores and median values for the pre- (T_1, T_2, T_3, T_4) and post- (T_3, T_4, T_5, T_6) tests were calculated to assess whether the experimental intervention had a statistically significant effect on the participants' oral (Activity 1) and written (Activity 2) text mediation ability at the within- and between-group levels (see Table 2).

Table 2. Descriptive statistics for the pre- and post-tests in Groups A and B.

			Activity	Mean	Median	SD
Group A	*n* = 6	T_1	1	2.50	2	0.837
			2	2.67	3	0.816
		T_3	1	2.50	2	1.378
			2	2.67	3	0.516
		T_5	1	3.00	3	0.632
			2	2.33	2	0.816
Group B	*n* = 8	T_2	1	2.63	3	0.518
			2	2.63	2	1.061
		T_4	1	2.50	2	1.195
			2	2.00	2	0.756
		T_6	1	2.75	3	0.886
			2	2.13	2	0.641

To choose an appropriate method of statistical analysis for examining the effect of the treatment at the within-group level, the Shapiro–Wilk test was performed for each dataset, namely Group A (T_1, T_3, and T_5) and Group B (T_2, T_4, and T_6) variables, to check whether they followed a normal distribution. The results of the normality test (see Table 3) showed that the distribution of ExG (T_1, T_3, and T_5) and CtrlG (T_2 and T_6) variables departed significantly from normality ($p < 0.05$).

Table 3. Results of the Shapiro–Wilk test for the T1–T6 variables.

		Activity	Pre-test	W	*p*	Post-test	W	*p*
Group A (n = 6)	Phase 1	1	T_1	0.70	0.006*	T_3	0.86	0.177
		2		0.82	0.09		0.64	0.001*
	Phase 2	1	T_3	0.86	0.177	T_5	0.79	0.010*
		2		0.64	0.001*		0.50	0.00002*
Group B (n = 8)	Phase 1	1	T_2	0.64	0.005*	T_4	0.85	0.090
		2		0.68	0.001*		0.85	0.092
	Phase 2	1	T_4	0.85	0.090	T_6	0.82	0.049*
		2		0.85	0.092		0.82	0.049*

Wilcoxon's signed rank test was then run to assess whether the treatment had an effect on the learners' ability to mediate oral and written texts at the within-group level by comparing the sample medians of the pre- and post-test scores. The treatment introduced in Phase 1 for Group A ($Z = 0.00$; $p = 1.000$) and in Phase 2 for Group B ($Z = 0.67$; $p = 0.5002$) did not lead to statistically significant changes in the ExG students' ability to mediate oral texts (Activity 1). Moreover, it did not result in statistically significant changes in the ExG students' ability to mediate written texts (Activity 2) in Phase 1 for Group A ($Z = 0.00$; $p = 1.000$) and in Phase 2 for Group B ($Z = 0.45$; $p = 0.6547$). The ExG students did not perform better on the oral and written text mediation tasks having been exposed to the treatment.

To assess the impact of the experimental intervention at the between-group levels, the mean gain scores for Activity 1 and 2, which is the difference in the scores on the post-tests (T_3, T_4, T_5, and T_6) and pre-tests (T_1, T_2, T_3, and T_4), were calculated. To choose an appropriate method of statistical analysis, the Shapiro–Wilk test was run on each dataset (T_1–T_6) to detect any departures from normality. The normality hypothesis could not be challenged in case of the T_3–T_1 and T_4–T_2 variables ($p > 0.05$), as the results of the normality test indicated that these variables were normally distributed (see Table 4).

Table 4. Results of the Shapiro–Wilk test for the T_3–T_1, T_4–T_2, T_5–T_3, and T_6–T_4 variables.

		W	p
T_3–T_1	1	0.91	0.47
	2	0.85	0.17
T_4–T_2	1	0.90	0.269
	2	0.89	0.24
T_5–T_3	1	0.86	0.18
	2	0.82	0.09
T_6–T_4	1	0.74	0.007*
	2	0.68	0.001*

In addition to the normality of the data's distribution, the homogeneity of variance for a variable needed to be fulfilled for an appropriate parametric test to be run. Levene's F-test was run to assess the equality of variances for T_3–T_1 and T_4–T_2. The F-test was not statistically significant for the two variables in Activity 1 (F = 1.49, p = 0.6066) and Activity 2 (F = 1.76, p = 0.5510). In view of this, the parametric independent samples T-test was performed to find out whether there is statistical evidence in favour of the fact that the mean gain scores were significantly different in Phase 1. The results indicated that the students' answers to Activity 1 (t = 0.14; p = 0.894) and Activity 2 (t = 1.08; p = 0.302) did not differ between Group A and B ($p > 0.05$). The negative mean gain scores of −0.13 and –0.63 points indicate that Group B performed slightly better on the text mediation tests before the two traditional homework assignments had been implemented.

Table 5. Results of the independent samples T-test: Group A versus Group B in Phase 1.

			Mean T_3–T_1	Mean T_4–T_2	*t*	*p*
ExG (Group A) versus CtrlG (Group B) in Phase 1	T_3–T_1 versus T_4–T_2	1	0.0	−0.13	0.14	0.894
		2	0.0	−0.63	1.08	0.302

Since the normality hypothesis for the T_6–T_4 variable was refuted ($p < 0.05$; see Table 4), the nonparametric Mann–Whitney U-test was used to compare the median values for Activities 1 and 2 achieved by Group A and Group B in Phase 2. The difference in the CtrlG and ExG students' ability to mediate oral (Z = 0.63; p = 0.529) and written (Z = –1.01; p = 0.31) texts also lacked statistical significance ($p > 0.05$). The difference in the median values for T_5 and T_3 in Activity 2 was reported to be negative (–1.0), which indicates that Group A performed better before the traditional homework was assigned.

Table 6. Results of the Mann–Whitney U-Test: Group A versus Group B in Phase 2.

			Median T_5–T_3	Median T_6–T_4	*Z*	*p*
CtrlG (Group A) versus ExG (Group B) in Phase 2	T_5–T_3 versus T_6–T_4	1	1.0	1.0	0.63	0.529
		2	–1.0	0.0	–1.01	0.31

RQ2: How do the participants handle communication with Alexa when performing the task?

The learners (n = 14) positively assessed their HMC, as evident in the high mean score of 4.64 points and the modal value of 5 points (Q1). They regarded their interaction with the VA as (very) successful, explaining that they easily managed to receive the answers to their questions, as originally written "Alexa answered all my questions both in speech and writing. It also displayed photos. It was easy and fun to use" and "I had no difficulties using Alexa. I asked it a question and then it gave me an answer" (Q2). No HMC breakdowns were reported to have taken place during the performance of homework assignments according to 36% of the logs (Q3). In the remaining 64% of the cases when the participants worked on the mediation tasks, they experienced HMC breakdowns; however, the breakdowns were, in their opinion, (very) rare, as demonstrated by the low mean scores (2.44) and the modal value (1 point) (Q4).

Table 7. Descriptive statistics for the close-ended question in the learners' logs.

	n	Mean	Median	Mode	SD	Min	Max
Q1: Assessment of HMC	28	4.64	5.00	5.00	1.193	2.0	6.0
Q3: Occurrence of HMC breakdowns		1.32	1.0	1.00	0.475	1.0	2.0
Q4: Frequency of HMC breakdowns	18	2.44	2.5	1.00	1.464	1.0	5.0

SD, standard deviation.

All the participants unanimously reported having employed two strategies to solve HMC breakdowns (Q5). Rephrasing (82%) and repetition (43%) were used by the students, as illustrated by statements such as "I checked where I made a mistake and asked the question again", "I came up with a different question, which has a meaning similar to the previous one", and "I tried to ask the question again but in a more accurate way."

The analysis of the learners' usage data that came from four participants revealed that there were 102 cases of HMC in total, including 47 instances of HMC breakdowns caused exclusively by the learners (46%). The common causes included, in descending order of frequency, hesitations and excessively slow speech (40%; e.g., "How…uhhh… do many… persons… live in Italy?, Why is… ummm… the Palace of Culture and Science… ummm…controversial?"), mispronunciations resulting from segmental errors (23%; e.g., "What countries

does Italy [/bəʊldə]?, What were added to the top of building the [/pæls/] of Culture and Science?"), and grammatical incorrectness (21%; e.g., "What Italian places worth visisted?" and "What were added to the top of Palace of Science and Culture in twenty zero zero?").

The students' most common reactions to HMC breakdowns included rephrasing (84%), which corroborates the findings of the logs detailed above. The strategy of repetition was used 15% of the time. None of the students was observed to have used the strategy of abandonment.

RQ3: How do the participants assess the application of Alexa when performing the task?

The participants (n = 14) positively welcomed the idea of using Alexa in EFL learning (see Table 8). They had favourable opinions on the usability of the VA. Specifically, there was relatively high agreement with the statement "My interaction with Alexa was clear and understandable during my performance of mediation-based tasks", as demonstrated by the mean score of 4.63 points. This implies that the students generally did not have problems understanding Alexa's responses. Nonetheless, the level of agreement was slightly lower for "It was easy for me to remember how to give Alexa commands" (3.93), which suggests that some of the students probably found it difficult to learn to use the VA.

Generally, the learners (n = 14) considered the technology an effective tool, with the statement "Using Alexa helped me a lot in my EFL learning to mediate" receiving a high mean score (4.67). On the other hand, the learners expressed doubt over the role that the VA could play in increasing their autonomy as an EFL learner. "Using Alexa gave me greater control over my learning of the English language" statement received the lowest mean score (3.60) out of all the 12 survey items.

The students' (n = 14) level of satisfaction with the use of the VA was also high. The mean values were all over 4, with the statements "It was interesting to use Alexa for my EFL learning to mediate" (5.27) and "I was satisfied with the functions offered by Alexa" (5.13) receiving the highest mean scores out of all the survey statements. This probably suggests that the students enjoyed communicating with Alexa within their EFL homework assignments.

Table 8. Descriptive statistics for the participants' answers (n = 14) to the post-study questionnaire.

		n	Mean	Median	Min	Max	SD
1.	Learning to use Alexa to complete mediation-based homework assignments was easy for me.	14	4.33	4	2	6	1.047
2.	It was easy for me to give Alexa commands to complete mediation tasks.		3.93	4	2	6	1.100
3.	My interaction with Alexa was clear and understandable during my performance of mediation tasks.		4.60	4	3	6	0.986
4.	Overall, it was easy to use Alexa to complete mediation tasks.		4.27	4	3	6	1.163
5.	Using Alexa helped me a lot in learning to mediate.		4.67	5	2	6	1.113
6.	Using Alexa gave me greater control over my EFL learning.		3.60	4	1	6	1.549
7.	Using Alexa improved my EFL ability to mediate.		4.07	4	2	6	1.163
8.	Overall, Alexa was useful in learning to mediate.		4.53	5	1	6	1.246
9.	It was interesting to use Alexa for learning to mediate.		5.27	5	4	6	0.704
10.	I am willing to continue using Alexa for my learning.		4.27	4	1	6	1.486
11.	I was satisfied with the functions offered by Alexa.		5.13	5	4	6	0.834
12.	Overall, I was satisfied with Alexa for learning to mediate.		4.93	5	3	6	0.884

SD, standard deviation.

The reliability of the instrument was evaluated for the sample of 14 participants to ensure validity. A correlation matrix was analysed to check whether the questions should be reversed or not. The Cronbach's alpha coefficient was calculated to examine the internal consistency of the scale by describing the extent to which all the items measured the same latent variable (i.e., the acceptance of Alexa by the students). The analysis of the correlation matrix revealed no need to reverse the questions. The correlation coefficients between Q4 and Question 6 (Q6) as well as Question 9 and Q6 were negative; however, that did not affect the internal consistency of the questionnaire. The Cronbach's alpha coefficient value was reported to be 0.89, which shows that the survey had a high level of reliability.

Several themes were identified by two proponents of Alexa who responded to the interview questions. As far as the self-perceived usefulness of the VA is concerned, interaction, information, and skills were commented on by the learners. They were convinced that thanks to the technology, they had a chance not only to communicate in English and improve their pronunciation and vocabulary knowledge, but also to easily and quickly access the information

they needed to complete the mediation tasks. Interaction, information, and skills were three themes discussed in the context of the effectiveness of the VA. The students argued that Alexa increased their willingness to communicate in English and enabled them to develop their ability to speak. They also added that it was able to quickly find information online. The students were satisfied with using Alexa: what the students liked about the VA was that it gave them instantaneous access to the information they needed to perform the tasks. The interviewees were able to identify more benefits than challenges related to the use of the technology. They listed four advantages of interacting with Alexa, namely gaining unrestricted access to information online, improving their English vocabulary and pronunciation, developing their ability to communicate, and entertainment. They were also able to suggest some potential problems related to the implementation of the VA when performing the tasks, expressing doubts over such issues as reliability and understanding. All the sample quotations taken from the SSIs have been translated from Polish into English by the author and are presented in Table 9.

A different perspective on using the VA was provided by the opponents of Alexa (n = 2). Generally, they considered the technology to be useless, given the problems with accessing information and the disruptive influence of multimodal responses provided by the technology on learning. The interviewees were convinced that their interaction with Alexa made the task more challenging and time-consuming. They also considered the VA ineffective in the context of a language learning process in view of their over-reliance on technology in everyday life. Since the students spent a lot of time on their mobile devices, they were in favour of traditional ways of learning to reduce their screen time. The participants were dissatisfied with Alexa, stressing psychological and time management issues. They felt demotivated by the task involving HMC and feared that they would not be understood by the technology. When asked about the self-perceived benefits related to using VA during the treatment, only one of the interviewees was able to identify one, namely the opportunity to practice EFL reception. The findings revealed that the two interviewees had faced three challenges that were connected with understanding, downloading, and issues of artificiality. Examples of quotations are presented in Table 10.

Table 9. Evaluation of HMC by the proponents of the VA (n = 2) during the post-study SSIs.

	Theme	Description	*n*	Examples
Usefulness	Interaction	I could converse in English.	2	Thanks to Alexa, I could hold conversations in English whenever and wherever I wanted.
	Information	I could easily access information.	2	I believe that Alexa was very useful, as it quickly found the information I needed to do my homework.
	Skills	It improved my pronunciation and vocabulary.	1	Alexa helped me to learn how to correctly pronounce selected words.
			1	I learned some new words by talking to Alexa.
Effectiveness	Interaction	It promoted EFL communication.	2	What I love about Alexa is that it encouraged me to speak English.
	Information	It allowed me to gain immediate access to information.	2	Alexa searched for information online. It is faster and easier than looking for information on your own.
	Skills	It improved my English speaking.	2	I could finally improve my ability to speak English.
Students' satisfaction	Information	It allowed me to gain easy access to information.	2	I quickly managed to find [the] information that I needed to complete [my] homework
Benefits	Interaction	I practised communication in English.	2	I think that I am able to communicate in English better.
	Skills	I improved my pronunciation and vocabulary.	1	I found out how to pronounce some English words.
			1	I learned the meaning of new words in English.
	Information	It offered quick access to information.	2	When you find information quickly, you can learn faster.
	Entertainment	It was fun and amusing to talk to Alexa.	1	I enjoyed talking to Alexa. Whenever I felt bored, I could talk to her, in addition to doing the homework.
Challenges	Reliability	It was difficult to be understood by Alexa.	1	No matter how many times I repeated the question, Alexa could not answer me and talked off-topic.
	Understanding	It was difficult understanding Alexa.	1	Sometimes I had problems understanding Alexa. I could not understand some words that were displayed on the screen either.

Table 10. Evaluation of HMC by the opponents of the VA (n = 2) during the post-study SSIs.

	Theme	Description	n	Examples
Uselessness	Information	It limited access to information.	2	I found it difficult to complete the tasks. Alexa didn't always give me the information I needed. I would have done the tasks faster if I had searched for the information online on my own.
	Learning	It distracted attention.	1	I couldn't concentrate on the task because Alexa kept reading out its answers, which was irritating.
Ineffectiveness	Learning	It did not meet educational needs.	2	Because I overuse my smartphone, I would like to reduce the amount of time I spend on it. That is why I am in favour of traditional ways of studying.
Students' dissatisfaction	Motivation	It caused frustration.	1	I felt depressed by the prospect of not being understood by Alexa and the necessity to repeat my questions.
	Time	It increased the time burden.	2	Completing homework with the help of Alexa was time-consuming. Instead of asking six or seven questions, I usually had to come up with many more because it couldn't answer me properly.
Benefits	None	–	2	I prefer learning English in a traditional way by doing vocabulary and grammar exercises.
	Language skill	It created opportunities for EFL reception.	1	Thanks to Alexa, I could practice both listening and reading, as the written version of the answer was displayed on the screen and the text was read out loud by the technology.
Challenges	Understanding	It was difficult to understand Alexa's responses.	1	Sometimes, I had problems understanding Alexa's oral responses. Luckily, their written version helped me to overcome such problems.
	Downloading	It was bothersome to install the app.	1	I found it a waste of time to download the Alexa app, as I own a smartphone with iOS software, which included the voice assistant called Siri.
	Artificiality	It was unnatural to talk to a device.	1	I found it strange to have to talk to my phone.

Discussion

The statistical analysis of the data gathered in response to RQ1 revealed that the treatment where the participants interacted with Alexa during the mediation task had no statistically significant effect on their ability to mediate EFL texts. Understanding Alexa, whose target users comprise native speakers, might have been a challenge for some participants who were not able to successfully deal with all the cases of HMC underlying the task because of the higher level of vocabulary used by the technology and its faster speech rate (Dizon, 2020). The current findings conflict with previous research which demonstrated the

positive impact of multimodal input (e.g., Chang, 2009; Guichon & McLornan, 2008) and student-generated questions (e.g., Khaki, 2014; Urlaub, 2012) on the participants' comprehension and indicated a beneficial influence of using the spoken dialogue systems on the students' language reception (Kim, 2018; Tai & Chen, 2022) and production (Dizon, 2020). This statistically nonsignificant difference could have been caused not only by the methodological limitations, including the small sample size, the short duration, and the low number of mediation tasks, but also by the treatment design. The students complained that the Alexa user guide had been handed out too late and that the task instructions were excessively long. Adding preparation activities to the treatment, abbreviating the instructions, and giving learners more time (e.g., two weeks) to familiarise themselves with the basics of HMC could all have had a positive impact on the study results. Even though scant evidence supporting the alternative hypotheses that the treatment improved the participants' mediation skills was provided, some small gains in the mean test scores of the ExG students in Phase 2 could be observed. It is thus possible that changes in the treatment design, combined with a longer period of implementation and a larger sample size, could lead to statistically significant differences in the participants' text mediation ability either at the within- or between- group level (or both).

The investigation of communication between the Polish students and Alexa during the task in the logs and the app carried out in response to RQ2 showed that breakdowns were intrinsically linked to the HMC process occurring in the study. Since a correlation between the learners' proficiency level and the VA's comprehensibility score was demonstrated by Chen et al. (2020), with the advanced Taiwanese EFL participants having been understood by the technology more frequently than intermediate and beginner learners, the relatively low level of language command (approximating B1) of the Polish learners participating in the current study might have been the cause of frequent HMC breakdowns resulting from their inaccuracy, mispronunciations, and hesitations. Because only 4 out of 14 participants agreed to have their history page accessible through the Alexa app analysed, not all the cases of HMC could be examined and therefore, it is not clear whether there were other causes of breakdowns, including learner-related (e.g., fast speech rate, atypical demands) (Moussalli & Cardoso, 2017) or technology-related ones (e.g., reliability) (Daniels & Iwago, 2017; Dizon, 2017; Moussalli & Cardoso, 2019), occurring during the task. Despite this, the students had favourable impressions of their interaction with the VA, which corroborates the results of previous CALL research on VAs (e.g., Dizon & Tang, 2019, 2020; Underwood, 2017). The analysis also revealed the technology's potential to engage EFL learners in negotiating meaning, as demonstrated by their frequent application of the paraphrasing and repetition strategies recorded in the logs and app. These findings, however, contradict the results of previous research, according to which, the out-of-class use of the VA application translated into students' abandonment of HMC (Dizon & Tang, 2020). In the current examination, the students' interaction with the VA was embedded in TBLL, which means that they were expected to ask Alexa questions to complete tasks, and therefore only rephrasing and repetition strategies

(not abandonment) were used, even though HMC was established outside of the classroom.

Having been exposed to the treatment, the 14 Polish EFL teenage learners positively assessed the idea of using Alexa to perform mediation tasks within their homework assignments, as demonstrated by the high mean scores for usability, effectiveness, and satisfaction in the post-study questionnaire conducted to answer RQ3. These findings confirm previous CALL research indicating students' positive attitudes towards the technology (Almekhlafi, 2006; Chen, 2013; Izadpanah & Alavi, 2016). The data collected in the post-study SSIs formed a comprehensive picture of the implementation of Alexa in TBLL, with the technology being simultaneously considered a source of linguistic, emotional, and practical support for the learners during the task and a time-wasting and dispensable piece of technological equipment. Similarly, the use of a VA in CALL has generally been a double-edged sword, with the study participants enjoying their interaction with VAs on the one hand, yet, on the other hand, experiencing many communication breakdowns, which devalued their overall experience of HMC (e.g., Moussali & Cardoso, 2016, Underwood, 2017; Chen et al., 2020). The atypical set of homework assignments used in the treatment, referred to as the "wow" factor (Murray & Barnes, 1998), caused mixed reactions among the study participants. It could be observed, however, that their highly favourable opinions significantly outnumbered the rare derogatory comments, which is in line with the process of technology adoption defined by Bax (2003) as a gradual transition from early adoption, ignorance, and scepticism, through to its implementation with fear and awe, to normalisation.

Implications and conclusions

Even though the results of the small-scale pilot project did not reveal any statistically significant improvements in the students' ability to mediate EFL texts (RQ1), the two-week treatment did involve the ExG participants in multimodal communication with Alexa. They needed to interact with the VA to complete two mediation-based tasks, which required the accurate production of oral language output and the reception of both written and oral (and often visual) language input by the learners. Despite the issues of understanding and reliability highlighted by the opponents of the technology, Alexa has potential to become an EFL language learning tool, as shown by the need to negotiate meanings with the VA by the participants (RQ2) and their very positive opinions about using the technology in their TBLL (for mediation) (RQ3).

On the basis of the interim results of our pilot study and past CALL research (e.g., Chen et al., 2020; Moussalli & Cardoso, 2016; Tai & Chen, 2022; Underwood, 2017), it can be concluded that the use of VAs in foreign language learning opens up welcome opportunities for language teachers and educators to innovate teaching by effecting the transition from teacher- to learner-centred learning. The technology use facilitates anytime-and-anywhere learning that is tailored to learners' personal needs (Yang, 2013) and maximises the chances for sustained language practice outside foreign language instruction settings

(Kukulska-Hulme, 2018). The application of VAs, which are accessible through mobile devices, is in line with the current theories of learning, such as constructivism and constructionism. The implementation of CALL promotes the development of learners' autonomy (Kukulska-Hulme et al., 2015) and facilitates experiential learning (Kukulska-Hulme & Viberg, 2018) through the involvement of students in authentic interactions with the voice-controlled technology anytime and anywhere.

The pilot study, in which dialogue-based CALL and TBLL were combined within the treatment, had several limitations, the most serious of which included a lack of randomisation, the small sample, and the short duration. Nonetheless, thanks to its mixed-methods research design, it suggested preliminary findings about the novel voice assistance phenomenon in TBLL and set new directions for further CALL research on VAs which, as already demonstrated, is still being neglected in this regard. Technology-mediated TBLL has created not only learning-by-doing environments, but also real-life contexts for interactive language use, simultaneously offering students individualised feedback on their utterances. More empirical research though is required to verify the current findings and fully explore the application of VAs in out-of-class TBLL, which is worth further investigation as an alternative to traditional homework assignments.

References

Abdallah, M., & Mansour, M. M. (2015). Virtual task-based situated language-learning with Second Life: Developing EFL pragmatic writing and technological self-efficacy. *Arab World English Journal, Special Issue on CALL*, 2, 150–182. https://files.eric.ed.gov/fulltext/ED558023.pdf

Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology and applications. *Machine Learning with Applications, 2*, 1–18. https://doi.org/10.1016/j.mlwa.2020.100006

Adobe Inc. (2019). *State of voice: Adobe digital insights 2019.* Digital Marketing Community. https://www.digitalmarketingcommunity.com/researches/state-of-voice-adobe-digital-insights-2019/

Almekhlafi, A. (2006). The effect of computer assisted language learning (CALL) on United Arab Emirates English as a foreign language (EFL) school students' achievement and attitude. *Journal of Interactive Learning Research, 17*(2), 121–142.

Bax, S. (2003). CALL: Past, present and future. *System, 31*(1), 13–28. https://eclass.uoa.gr/modules/document/file.php/ENL110/callpresentpastandfuture.pdf

Brown, H. (2004). *Language assessment. Principles and classroom practices.* White Plains: Pearson Education.

Chang, A. C.-S. (2009). Gains to L2 listeners from reading while listening vs. listening only in comprehending short stories. *System, 37*(4), 652–663. https://doi.org/10.1016/j.system.2009.09.009

Chen, X. (2013). Tablets for informal language learning: Student usage and attitudes. *Language Learning & Technology, 17*(1), 20–36. http://llt.msu.edu/issues/february2013/chenxb.pdf

Chen, H., Yang, C., & Kuo-Wei Lai, K. (2020). Investigating college EFL learners' perceptions toward the use of Google Assistant for foreign language learning. *Interactive Learning Environments, 31*(3), 1–16. https://doi.org/10.1080/10494820.2020.1833043

Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching and assessment.* Cambridge University Press.

Council of Europe (2020). *Common European framework of reference for languages: Learning, teaching, assessment. Companion volume.* Council of Europe Publishing.

Coste D., & Cavalli M. (2015). *Education, mobility, otherness. The mediation functions of schools.* Council of Europe. https://rm.coe.int/education-mobility-otherness-the-mediation-functions-of-schools/16807367ee

Creswell, J. (2014). *Research design: Qualitative, quantitative and mixed methods approaches.* Sage.

Daniels, P., & Iwago, K. (2017) The suitability of cloud-based speech recognition engines for language learning. *The JALT CALL Journal, 13*(3), 211–221. https://files.eric.ed.gov/fulltext/EJ1165166.pdf

Davis, F. (1989). Perceived usefulness, perceived ease of use and user acceptance of information technology. *MIS Quarterly, 13*, 318–341. https://doi.org/10.2307/249008

Dizon, G. (2017). Using intelligent personal assistants for L2 learning: A case study. *TESOL Journal, 8*(4), 29–42. https://doi.org/10.1002/tesj.353

Dizon, G. (2020). Evaluating intelligent personal assistants for L2 listening and speaking development. *Language Learning & Technology, 24*(1), 16–26. https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/49bc9b68-5ab6-4379-8eee-87018fc92d9b/content

Dizon, G. (2021). Affordances and constraints of intelligent personal assistants for second-language learning. *RELC Journal, 54*(3), 1–8. https://doi.org/10.1177/00336882211020548

Dizon, G., & Tang, D. (2019). A pilot study of Alexa for autonomous second language learning. In F. Meunier, J. Van de Vyver, L. Bradley & S. Thouesny (Eds.), *CALL and Complexity. Short Papers from EUROCALL 2019* (pp. 107–112). Research-publishing.net. https://doi.org/10.14705/rpnet.2019.38.994

Dizon G., & Tang, D. (2020). Intelligent personal assistants for autonomous second language learning. *The JALT CALL Journal, 16*(2), 117–130. https://doi.org/10.29140/jaltcall.v16n2.273

Dörnyei, Z. (2007). *Research methods in applied linguistics.* Oxford: Oxford University Press.

Fang, W., Yeh, H., Luo, B., & Chen, N. (2021). Effects of mobile-supported task-based language teaching on EFL students' linguistic achievement and conversational interaction. *ReCALL, 33*(1), 71–87. https://doi.org/10.1017/S0958344020000208

Fraenkel, J., Wallen, N., & Hyun, H. (2012). *How to design and evaluate research in education.* The McGraw-Hill Companies, Inc.

Gillett, F. (2020). Getting consumers beyond simple tasks on smart speakers is challenging. *Forrester.* https://www.forrester.com/blogs/getting-consumers-beyond-simple-tasks-on-smart-speakers-is-challenging/

Gonzalez-Lloret, M. (2014). The need for needs analysis in technology-mediated TBLT. In M. Gonzalez-Lloret & L. Ortega (Eds.), *Technology and tasks: Exploring technology-mediated TBLT* (pp. 23–50). John Benjamins Publishing Company.

Gonzalez-Lloret, M. (2015). *A practical guide to integrating technology into task-based language teaching.* Georgetown University Press.

Gonzalez-Lloret, M. (2017). Technology for task-based language teaching. In C. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 234–247). Wiley-Blackwell.

Greene J., Caracelli V., & Graham W. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 11*(3), 255–274. https://doi.org/10.3102/01623737011003255

Guichon, N., & McLornan, S. (2008). The effects of multimodality on L2 learners: Implications for CALL resource design. *System, 36*(1), 85–93. https://doi.org/10.1016/j.system.2007.11.005

Guzman, A.L. (2018). What is human–machine communication, anyway? In A. L. Guzman (Ed.)., *Human–machine communication: Rethinking communication, technology, and ourselves* (pp. 1–28). Peter Lang.

Guzman, A., & Lewis, S. (2020). Artificial intelligence and communication: A human–machine communication research agenda. *New Media & Society, 22*(1), 70–86. https://doi.org/10.1177%2F1461444819858691

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Howell, P. (2017). Cross-language mediation in foreign language teaching. *Hiroshima Studies in Language and Language Education, 20,* 147–155. https://doi.org/10.15027/42619

Izadpanah, S., & Alavi, M. (2016). The perception of EFL high school students in using of computer technology in the process of learning: Merits and demerits. *Advances in Language and Literary Studies, 7*(3), 146–156. http://dx.doi.org/10.7575/aiac.alls.v.7n.3p.146

Khaki, N. (2014). Improving reading comprehension in a foreign language: Strategic reader. *The Reading Matrix, 14*(2), 186–200. https://www.readingmatrix.com/files/11-m9371u67.pdf

Kim, N. (2018). A study on chatbots for developing Korean college students' English listening and reading skills. *Journal of Digital Convergence, 16*(8), 19–26. https://doi.org/10.14400/JDC.2018.16.8.019

Konarzewski K. (2000). *Jak uprawiać badania oświatowe: Metodologia praktyczna.* Wydawnictwa Szkolne i Pedagogiczne.

Kondracki, N., Wellman, N., & Amundson, D. (2002) Content analysis: Review of methods and their applications in nutrition education. *Journal of Nutrition Education and Behaviour, 34*, 224–230. https://doi.org/10.1016/s1499-4046(06)60097-3

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology.* Sage Publications.

Kruk, M. (2011). The use of the internet and instant messengers in assisting the acquisition of speaking skills in English lessons. In M. Pawlak, E. Waniek-Klimczak & J. Majer (Eds.), *Speaking and instructed foreign language acquisition* (pp. 258–267). Bristol: Multilingual Matters.

Kruk, M. (2015). CALL vs. traditional grammar instruction. The case of the past simple tense. In A. Turula & M. Chojnacka (Eds.), *CALL for bridges between school and academia* (pp. 129–142). Frankfurt am Main: Peter Lang.

Kukulska-Hulme, A. (2018). Mobile assistance for personal learning on a massive scale. In M. Orsini-Jones & S. Smith (Eds), *Flipping the blend through MOOCs, MALL and OIL: New directions in CALL* (pp. 1–7). Research-publishing.net. https://files.eric.ed.gov/fulltext/ED585289.pdf

Kukulska-Hulme, A. (2019). Intelligent assistants in language learning: Friends or foes? In C. Glahn, R. Power & E. Tan (Eds.), *Proceedings of World Conference on Mobile and Contextual Learning* (pp. 127–131). LearnTechLib. https://oro.open.ac.uk/75261/

Kukulska-Hulme, A., Norris, L., & Donohue, J. (2015). *Mobile pedagogy for English language teaching: A guide for teachers.* The British Council.

Kukulska-Hulme, A., & Viberg, O. (2018). Mobile collaborative language learning: State of the art. *British Journal of Educational Technology, 49*(2), 207–218. https://doi.org/10.1111/bjet.12580

Kukulska-Hulme, A., & Lee, H. (2020). Intelligent assistants in language learning: An analysis of features and limitations. In K. Frederiksen, S. Larsen, L. Bradley & S. Thouesny (Eds.), *CALL for widening participation: Short papers from EUROCALL 2020* (pp. 172–176). Research-publishing.net. https://doi.org/10.14705/rpnet.2020.48.1184

Long, M. (2005). Second language needs analysis. Cambridge University Press.

Moussalli, S., & Cardoso, W. (2016). Are commercial "personal robots" ready for language learning? Focus on second language speech. In S. Papadima-Sophocleous, L. Bradley & S. Thouesny (Eds), *CALL communities and culture: Short Papers from EUROCALL 2016* (pp. 325–329). Research-publishing.net. https://doi.org/10.14705/rpnet.2016.eurocall2016.583

Moussalli, S., & Cardoso, W. (2017). Can you understand me? Speaking robots and accented speech. In K. Borthwick, L. Bradley & S. Thouesny (Eds.), *CALL in a climate of change: Adapting to turbulent global conditions. Short papers from EUROCALL 2017* (pp. 217–221). Research-publishing.net. https://doi.org/10.14705/rpnet.2017.eurocall2017.716

Moussalli, S., & Cardoso, W. (2019). Intelligent personal assistants: Can they understand and be understood by accented L2 learners? *Computer Assisted Language Learning, 33*(1), 1–26. https://doi.org/10.1080/09588221.2019.1595664

Mulyadi, D., Wijayatiningsih, T. D., Singh, C. K. S., & Prastikawati, E. F. (2021). Effects of technology enhanced task-based language teaching on learners' listening comprehension and speaking performance. *International Journal of Instruction, 14*(3), 717–736. https://doi.org/10.29333/iji.2021.14342a

Murray, L., & Barnes, A. (1998). Beyond the "wow" factor: Evaluating multimedia language learning software from a pedagogical viewpoint. *System, 26*(2), 249–259. https://doi.org/10.1016/S0346-251X(98)00008-6

Norris, J. (2009). Task-based teaching and testing. In M. Long & C. Doughty (Eds.), *The handbook of language teaching* (pp. 578–594). John Wiley & Sons.

Piccardo, E., Berchoud, M., Cignatta, T., Mentz, O., & Pamuła, M. (2011). *ECEP project guide. Pathways through assessing, learning and teaching in the CEFR.* Council of Europe. https://www.ecml.at/Portals/1/documents/ECML-resources/2011_08_29_ECEP_EN_web.pdf

Shanks, D. (2021). Technology-facilitated oral homework: Leveraging technology to get students speaking outside the classroom. In T. Beaven & F. Rosell-Aguilar (Eds), *Innovative language pedagogy report* (pp. 69–75). Research-publishing.net. https://doi.org/10.14705/rpnet.2021.50.1238

Shokrpour, N., Mirshekari, Z., & Moslehi, S. (2019). Learning vocabulary electronically: Does computer assisted language learning (CALL) instruction have any impacts on Iranian EFL learners? *Cogent Education, 6*(1), 1–20. https://doi.org/10.1080/2331186X.2019.1702827

Sirotova, M., Michvocikowa, V., & Rubacha, K. (2021). Quasi-experiment in the educational reality. *Journal of Education, Culture and Society, 1*, 189–201. https://doi.org/10.15503/jecs2021.1.189.201

Stathopoulou, M. (2015). *Cross-language mediation in foreign language teaching and testing.* Multilingual Matters.

Social Lens Research (2018). *Voice commands: Current state.* https://www.sociallensresearch.com/wp-content/uploads/2018/10/FINAL-SLR-Voice-Command-Study-10.18.pdf

Tai, T., & Chen, H. (2020). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments, 31*(3), 1–18. https://doi.org/10.1080/10494820.2020.1841801

Tai, T., & Chen, H. (2022). The impact of intelligent personal assistants on adolescent EFL learners' listening comprehension. *Computer Assisted Language Learning, 37*(3), 1–28. https://doi.org/10.1080/09588221.2022.2040536

Tavakoli, H., & Loth, A. (2021). The impact of CALL-mediated TBLT on L2 reading comprehension. *Journal of Educators Online, 18*(2), 142–157.

Terzopoulos, G., & Satratzemi, M. (2019, September 26–28). *Voice assistants and artificial intelligence in education* [Paper presentation]. Ninth Balkan Conference on Informatics, Sofia, Bulgaria. https://doi.org/10.1145/3351556.3351588

Underwood, J. (2017). Exploring AI language assistants with primary EFL students. In K. Borthwick, L. Bradley & S. Thouesny (Eds.), *CALL in a climate of change: Adapting to turbulent global conditions. Short Papers from EUROCALL 2017* (pp. 317–321). Research-publishing.net. https://doi.org/10.14705/rpnet.2017.eurocall2017.733

Underwood, J. (2021). Speaking to machines: Motivating speaking through oral interaction with intelligent assistants. In T. Beaven & F. Rosell-Aguilar (Eds.), *Innovative language pedagogy report* (pp. 127–132). Research-publishing.net. https://doi.org/10.14705/rpnet.2021.50.1247

Urlaub, P. (2012). Reading strategies and literature instruction: Teaching learners to generate questions to foster literary reading in the second language. *System, 40*(2), 296–304. https://doi.org/10.1016/j.system.2012.05.002

Widiastuti, O., Ivone, F. M., Sulistyo, T., Hartono, D., Sudarwati, E., & Prastiyowati, S. (2022). CALL-mediated task-based language teaching: A speaking project with online audiences in Indonesia. *Indonesian Journal of Applied Linguistics, 12*(1), 224–234. https://doi.org/10.17509/ijal.v12i1.46427

Willis, J. (2004). Perspectives on task-based instruction: Understanding our practices, acknowledging different practitioners. In B. Leaver & J. Willis (Eds.), *Task-based instruction in foreign language education: Practices and programs* (pp. 3–44). Georgetown University Press.

Yang, J. (2013). Mobile assisted language learning: Review of the recent applications of emerging mobile technologies. *English Language Teaching, 6*(7), 19–25. https://doi.org/10.5539/elt.v6n7p19

Ziegler, N. (2016). Taking technology to task: Technology-mediated TBLT, performance, and production. *Annual Review of Applied Linguistics, 36,* 136–163. https://doi.org/10.1017/s0267190516000039