Arigatō: Effects of Adaptive Guidance on Engagement and Performance in Augmented Reality Learning Environments

Maheshya Weerasinghe, *Student Member, IEEE*, Aaron Quigley, Klen Čopič Pucihar,
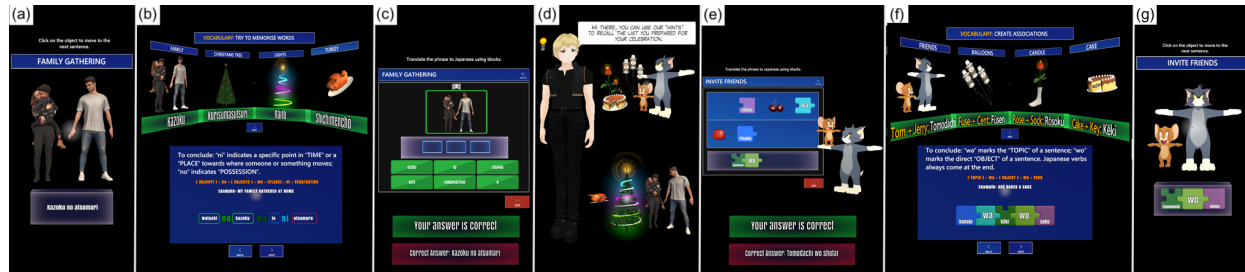Alice Toniolo, Angela Miguel and Matjaž Kljun

Fig. 1. Arigatō prototype and its experiential learning stages for two different topics: Christmas and birthday celebration. (a) Reflective observation (RO), (b) Abstract conceptualisation (AC), and (c) Active experimentation (AE), stages for fixed-associations condition on the Christmas celebration topic. (d) Concrete experience (CE) for both topics. (e) Active experimentation (AE), (f) Abstract conceptualisation (AC), and (g) Reflective observation (RO) stages for adaptive-associations condition on a birthday celebration topic. To avoid excessive clutter of the AR scene the black background was selected.

Abstract— Experiential learning (ExL) is the process of learning through experience or more specifically "learning through reflection on doing". In this paper, we propose a simulation of these experiences, in Augmented Reality (AR), addressing the problem of language learning. Such systems provide an excellent setting to support "adaptive guidance", in a digital form, within a real environment. Adaptive guidance allows the instructions and learning content to be customised for the individual learner, thus creating a unique learning experience. We developed an adaptive guidance AR system for language learning, we call Arigatō (Augmented Reality Instructional Guidance & Tailored Omniverse), which offers immediate assistance, resources specific to the learner's needs, manipulation of these resources, and relevant feedback. Considering guidance, we employ this prototype to investigate the effect of the amount of guidance (fixed vs. adaptive-amount) and the type of guidance (fixed vs. adaptive-associations) on the engagement and consequently the learning outcomes of language learning in an AR environment. The results for the amount of guidance show that compared to the adaptive-amount, the fixed-amount of guidance group scored better in the immediate and delayed (after 7 days) recall tests. However, this group also invested a significantly higher mental effort to complete the task. The results for the type of guidance show that the adaptive-associations group outperforms the fixed-associations group in the immediate, delayed (after 7 days) recall tests, and learning efficiency. The adaptive-associations group also showed significantly lower mental effort and spent less time to complete the task.

Index Terms—Experiential learning, instructional guidance, adaptive learning systems, augmented reality, engagement, language Learning

◆

1 INTRODUCTION

Experiential learning (ExL) is a well-known learning approach used in education, training, facilitation, coaching and organisational development [28, 52, 57]. ExL refers to the process of learning through experience or "learning by doing". One of the most influential models in ExL [57] is the Kolb's learning model [50]. This model defines ExL as "the process whereby knowledge is created through the transformation of experience. Knowledge results from the combination of grasping and transforming experience" [50]. The model is repre-

sented by four cyclical learning stages as seen in Fig. 2: (1) concrete experience (feeling), (2) reflective observation (watching), (3) abstract conceptualisation (thinking), and (4) active experimentation (doing). One can enter the learning cycle at any stage, but all stages in the cycle must be addressed for meaningful learning to occur.

In learning, including experiential learning, guidance is beneficial. The term guidance has a broad meaning, but we adopt the generally understood definition of, guidance as "any form of assistance offered to users so they can achieve a learning goal". In more precise terms, the role of guidance is "to simplify, provide a view on, elicit, supplant, or prescribe the scientific reasoning skills" [55] involved in pursuing a goal. It can be explicit, as when a teacher provides instructions on how to solve a mathematical problem with all the in-depth explanations of concepts and skills students need to learn, or provides feedback on the current performance. Or it can be implicit, as it is provided within the environment for users to find, interpret and use in order to progress in the task at hand [7, 16]. In either case the amount and the type of guidance play an essential role in engagement and consequently, in the learning outcomes [17, 44, 96].

While several educational definitions of guidance exist [5, 19, 55], in the context of this study we focus on instructional "guidance", which is defined as "providing users with accurate and complete procedural information (and related declarative knowledge) that they have not

- *Maheshya Weerasinghe is with the University of Primorska, Slovenia and the University of St Andrews, United Kingdom. E-mail: amw31@st-andrews.ac.uk .*
- *Aaron Quigley is with the University of New South Wales, Australia. E-mail: a.quigley@unsw.edu.au .*
- *Alice Toniolo and Angela Miguel are with the University of St Andrews, United Kingdom. E-mail: {a.toniolo | arm14}@st-andrews.ac.uk .*
- *Klen Čopič Pucihar and Matjaž Kljun are with the University of Primorska, Slovenia. E-mail: {klen.copic | matjaz.kljun}@famnit.upr.si .*

yet learned in a demonstration about how to perform the necessary sequence of actions and make the necessary decisions to accomplish a learning task" [19]. Guidance also has different dimensions such as "how much" (amount) instructional support it provides [64, 95, 105], and "what kind" (type) of support it provides [5]. Digital technologies such as augmented reality (AR) appear ideal to support adaptation of guidance. AR can provide an environment for simulating various experiences by combining the physical environments with in-context adaptive digital elements that can be manipulated by the user. AR can thus replicate the real-world ExL scenario (e.g. learning by decorating a room for Christmas celebrations with AR objects as can be done with real objects) and show information in a coherent and meaningful way within the real world context.

Researchers have previously explored AR systems as tools to guide or teach different skills and activities that can be learned best through experience [37, 67]. Most of these projects have primarily focused on creating simulated environments to support different stages of the Kolb's experiential learning cycle [12, 37, 67, 68, 97]. However, these projects explored generic guidance only, in which all the users receive the same set of general instructional cues, with the same amount, in the same way [12, 41, 60, 63, 101]. Therefore, there are several questions that remain unanswered, such as: How to design and implement guidance in learning environments where the learning system needs to take over some or all of the tasks commonly conducted by educators? How the adaptation of guidance (the amount and the type) would affect the engagement and consequently the learning outcomes?

To this end, we built the Arigatō – an AR prototype, which aims to effectively support learners to proceed through the ExL cycle facilitated by adaptive guidance. We deliberately selected AR as the most probable future technology that will be used in the classroom since it better supports real-world in-person communication and group collaboration compared to a desktop, tablet, and VR [69, 80]. And rather than comparing the AR to other technologies our goal was to explore the design space of the AR.

With the Arigatō prototype that provides options for a fixed and adaptive amount and type of guidance, we aimed to answer the following research questions:

RQ1 How do the dynamic adaptations of AR guidance influence learners' (RQ1a) recall of previously learned information, (RQ1b) mental effort, (RQ1c) task completion time, and (RQ1d) instructional efficiency?

RQ2 How does learners' engagement with AR content in terms of (i.e., task completion time, mental effort and motivation) affect their performance (i.e., recall and efficiency)?

The focus on AR in language learning covers the exploration of XR applications in the education and training domains making this study strongly relevant to the XR research community.

2 RESEARCH BACKGROUND

We first identify and analyse the relevant literature on experiential learning, instructional guidance and in particular the amount and type with a focus on language learning, user engagement, and language learning in AR.

2.1 Experiential Learning

Experiential learning theory is based on the idea that learning is a process of acquiring knowledge through experience or "learning through reflection on doing" [50, 57]. Theoretical work on experiential learning continues to evolve and several models as "theories-in-use" have been proposed such as the Kolb's model [50], Boud and Walker's model [13], Joplin's five stage model [42], etc. Among these, the Kolb's experiential learning model is arguably the most influential model on educational scholarship [57]. As shown in Fig. 2, the model describes experiential learning as a four-part process, where the learner is asked to engage in a new experience, actively reflect on that experience, conceptualise

it, and integrate it with prior experiences and knowledge. After completing this cycle, future decisions can be based on the newly acquired concepts.
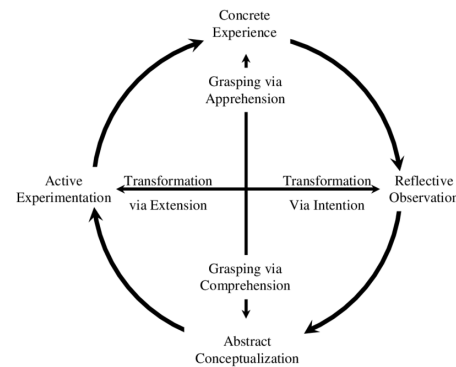
Fig. 2. Experiential learning cycle by Kolb [50]

A range of prior work has explored the use of simulated environments and AR technology to support experiential learning [3, 12, 37, 40, 67, 68, 70, 93, 97, 102]. Most of these approaches have primarily focused only on the learning stages of the Kolb's experiential learning cycle, allowing learners to perform an action and providing feedback, which learners could reflect on [12, 32, 37, 67, 68, 97]. However, according to Kolb, the learner must continue cycling through all the four stages, thus creating a "learning spiral of ever-increasing complexity" in order to gain a better learning outcome [51]. Currently, we are unaware of any work that focuses on providing guidance to support learners move from one stage to another in the experiential learning cycle and progress on the learning spiral.

2.2 Instructional Guidance

In the past, the instructional support provided during learning has been categorised and explained using different concepts and terms, such as learning paradigms (e.g. behaviourism, cognitivism and constructivism, etc.) [65, 95], instructional strategies (e.g. direct instructions, indirect instructions, interactive instructions, etc.) [66, 103] and instructional methods (e.g. lectures, case studies, peer feedback, quizzes, etc.) [22, 94]. Such paradigms, strategies and methods often emerge from different learning theories, and sometimes various belief systems and philosophies [22, 35, 64, 65, 105]. Therefore, when trying to employ a construct, such as "instructional guidance", we depend on the fact that the use of different conceptual frameworks and theories tend to define and operationalise instructional support in different ways. In the following sections, we will focus on the amount and type of guidance as described in the literature.

2.2.1 Amount of Guidance

Debates about the impact of instructional guidance and "how much" instructional support needs to be provided in a learning environment have been ongoing for at least the past century [47, 54, 64, 95]. For example, is it better to instruct learners on what they need to know by presenting them the essential information, or is it better to allow learners to discover or construct essential knowledge for themselves? Koedinger and Aleven called this issue the "assistance dilemma" [48], i.e., deciding whether to provide or withhold assistance. The contrast between the two practices can be better understood as a continuum. On one side of this continuum is the hypothesis that people learn best in an unguided or minimally guided environment [15, 90]. On the other side is the hypothesis that learners should be given instructional guidance on the concepts and procedures required for a particular task [22, 47, 64].

In a minimally guided learning environment, students are seen as active learners and are given opportunities to digest content for themselves rather than as passive learners who merely follow instructions. While these approaches have been adopted by some teachers and educators [104], there have also been decades of efforts to discourage

educators from using minimally guided learning approaches. The past half-century of empirical research on the topic has provided overwhelming and unambiguous evidence that minimal guidance during learning is significantly less effective and less efficient compared to guided learning [47]. The superiority of the latter is explained in the context of human cognitive architecture and expert-novice differences [47]. Wise and O'Neill present a review of evidence to support the view that the optimal amount of guidance is often somewhere in the middle of the aforementioned continuum and that the *granularity* of the advice provided in a task (i.e. the level of details) is equally important [105].

2.2.2 Type of Instruction

Research on memory and learning has shown that comprehension and recall depend on different types of instructional methods and techniques that can be used to process and store information [25]. The mnemonic techniques have proven to be extremely effective in improving memory and recall, especially in the area of foreign language learning [6, 20, 72, 76]. Mnemonic is an instructional strategy designed for enhancing both memory and recall [62, 74, 75, 106]. This technique connects new learning to prior knowledge through the use of visual and/or acoustic cues. The basic types of mnemonic techniques rely on the use of key words, rhyming words, or acronyms [62, 75].

In the field of language learning, mnemonics have mostly been used for vocabulary learning. Despite the great variety of techniques for presenting mnemonic, the "keyword method" [11] for creating mental associations to a known language has proven to be effective in the memorisation of vocabulary. In the keyword method, learners associate the sound of a word they want to learn to one they already know in either their first language or the target language. They then mentally create an image of the known word to memorise the association [74]. This association based technique provides a powerful tool for words that have a high degree of "imagenability" [79], or for word pairs between which the learner can form some kind of semantic link [26]. The important aspect is that the keyword should clearly relate to the thing being remembered. This method also motivates learners to be more creative and use their minds more productively.

2.3 User Engagement

Engagement is a complex and multi-dimensional process [81, 83, 96] intertwined with a learners' internal indicators such as motivation, feelings, etc. (affective dimension) [30], mental effort, perceptions, etc. (cognitive dimension) [30], and observable actions such as performing various activities, interacting with a system, etc. (behavioural dimension) [8, 30]. In addition, Reeve and Tseng suggested incorporating agentic engagement as a fourth dimension of engagement [77]. Agentic engagement refers to the proactive and intentional activity of the learner to personalise the conditions of learning and to enrich external learning goals.

Previous studies have shown that behavioural, cognitive and affective dimensions predict learner' performance both separately and in unison [4, 96]. This means that the dimensions of user engagement are interrelated and simultaneously affect human behaviour [30].

Kearsley and Shneiderman have stressed that engagement can be stimulated without technology, but digital technology opens up novel possibilities that are hard to achieve in a physical form [46]. Accordingly, research in this area has explored engagement with images [29], video clips [2, 89], music [39, 49], and real-life scenarios [45, 99]. However, while there are a limited numbers of studies such as [33, 96], there is a lack of research on learning and engagement in AR environments.

2.4 Language Learning in AR

A considerable body of literature focuses on guidance in AR environments to support learning in general [12, 60, 63, 93, 101] and language learning in particular [9, 23, 24, 38, 87, 107]. Prior work on language Learning in AR is summed in Table 1. We categorised it based on the following dimensions: (i) *Hardware used:* mobile AR, HMD, sensors; *Learning focus:* vocabulary and/or grammar; *Guidance method:* generic or adaptive; and *Learning method:* experiential, contextual,

game-based learning, etc. [100]). Importantly, most of the existing approaches deliver generic guidance, in which all users receive the same set of general instructional cues, and in the same way [12, 60, 63, 101].

Table 1. Selected prior work related to language learning in AR environments and how our work differs alongside different dimensions.

Study	Hardware used	Learning Focus	Guidance Method	Learning Method
Draxler et al. (2020) [24]	Mobile	Grammar	Generic	Context-based learning
Arvanitis et al. (2020) [9]	Mobile	Vocabulary	Generic	Self-directed learning
Yang & Mei (2018) [107]	Mobile	Vocabulary	Generic	Game-based learning
Ibrahim et al. (2018) [38]	HMD	Vocabulary	Generic	Context-based learning
Vazquez et al. (2017) [98]	HMD	Vocabulary	Generic	Context-based learning
Dita (2016) [23]	Mobile	Vocabulary	Generic	Game-based learning
Seedhouse et al. (2014) [87]	Sensors	Vocabulary	Generic	Experiential learning
Liu & Tsai (2013) [59]	Mobile	Vocabulary	Generic	Context-based learning
Arigatō (2022)	HMD	Vocabulary Grammar	Adaptive	Experiential learning

However, as pointed out in Sect. 2.2, the amount and the type of guidance play a vital role when providing instructional guidance. In addition, the adaptation of guidance to user needs is also important as noted in Sect. 2.2.1. Together they affect the engagement and consequently the learning outcomes of each individual learner. While there are studies such as [36, 93] that provide adaptive guidance for performing a physical task, generally there is a lack of research on providing adaptive guidance based on the amount or type of instructions. Moreover, we are unaware of any work that focuses on the "keyword method" as the type of guidance used to support language learning in AR environments.

In order to address these gaps, we developed an adaptive guidance AR system for language learning called Arigatō. We employ this prototype to investigate the effect of the amount (fixed vs. adaptive) and the type of associations (fixed vs. adaptive) on the engagement and consequently the learning outcomes of language learning in an AR environment. The research method to investigate the aforementioned effects is presented in the next section.

3 RESEARCH METHOD

The Arigatō prototype for language learning was developed to answer our research questions. In this section, we present the prototype and we describe the study conditions, study design, study procedure, participants' sampling, data collection, and analysis.

3.1 Study Conditions

The language selected for this study was Japanese, since it is unrelated to the Indo-European family of languages and we expected people would not be familiar with its grammar and vocabulary. We designed four different study conditions based on two different aspects of instructional guidance that can be adapted: the AMOUNT of guidance (FIXED-AMOUNT, which is the same for all, and ADAPTIVE-AMOUNT, which decreases based on the learners' performance), and the TYPE of instructions, (FIXED-ASSOCIATIONS using predefined 3D AR models of the vocabulary being learnt and ADAPTIVE-ASSOCIATIONS using self selected 3D AR models to create associations ("keyword method") to the vocabulary being learnt).
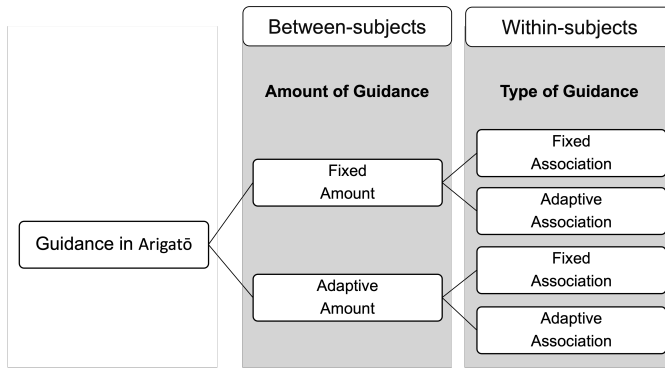
Fig. 3. Study design and conditions.

The structure of the design of the study with all four (4) study conditions is illustrated in Fig. 3. The study was planned as a 2 x 2 mixed design study (including both within and between-subjects) envisaged to take approximately 60 to 75 minutes. A common within-subjects design would make this cognitively demanding learning study even longer (approx two to three hours), which might hinder participants' performance and negatively affect the results. Other options such as splitting the study into several sessions would also introduce other biases (e.g. users might study between sessions, day to day performance might vary) and practical issues (e.g. getting all users back for the following session).

The AMOUNT of guidance was thus evaluated as a within-subjects variable while the TYPE of instructions as a between-subjects variable. This means that each participant either received the FIXED-AMOUNT or the ADAPTIVE-AMOUNT of guidance, but all participants experienced both FIXED-ASSOCIATIONS and ADAPTIVE-ASSOCIATIONS.

In each condition, participants had to learn Japanese vocabulary and grammar around a particular topic (Christmas or birthday celebration) by learning and understanding four (4) distinct phrases and their structure, and recall them successfully. The phrases related to Christmas were: family gathering, Christmas tree preparation, turkey dinner and lights decorations. The phrases related to birthday celebration were: invite friends, blow balloons, bake a cake and light up candles.

In the FIXED-AMOUNT condition, at each stage of the learning cycle participants received all the instruction for all the content needed to be learnt for all four (4) phrases repeatedly through consequent cycles until all phrases were correctly recalled. In the ADAPTIVE-AMOUNT condition, at each stage of the learning cycle participants only received instructions for the phrases that were not recalled correctly in previous cycles. Thus, once a phrase was recalled correctly in the CE stage, the guidance for that phrase was not shown in the next cycles. In the FIXED-ASSOCIATIONS condition, participants were presented with predefined 3D AR models of objects corresponding to vocabulary of the phrases being learned. In the ADAPTIVE-ASSOCIATIONS condition, the participants could self-select 3D AR models to create associations for the corresponding vocabulary of the phrases being learned.

To avoid the "order effects" (the influence of the order in which the conditions are presented on participants' performances [86]), the order of TYPE of instructions (FIXED and ADAPTIVE) as well as the order of the topic being learnt (Christmas and birthday celebration) was balanced among the participants.

3.2 The Arigatō Prototype

The prototype was developed for the Magic Leap one AR head mounted display (HMD) with controller inputs [1], using the Unity3D game development environment [2]. The Lumin SDK 0.26 and Magic leap tools were used for setting up the development environment with Lumin OS 0.98.3. The MRTK Mixed Reality tool kit [3] was used for integrating

the inputs and object manipulation techniques. Speech recognition was implemented with the IBM Watson speech to text API [4].

The prototype replicates all four stages of Kolbs' experiential cycle. Before entering the cycle, the learner first sees an introduction window including all four (4) phrases related to the topic being learnt in English in order to familiarise themselves with the intent and content of the study. Besides text instructions and explanations, audio instructions and explanations are also provided throughout the cycle.

Fig. 4. The study setup where the participant tries to finish the learning tasks visible in AR HMD and researcher follows the process on the laptop screen.

The system starts the learning cycle in the abstract conceptualisation (AC) stage. The AC stage provides the description of the vocabulary and the basic grammar rules related to phrases being learnt. The vocabulary is presented by either (i) FIXED-ASSOCIATIONS (for example, for the word lights – "raito" in Japanese – a 3D virtual object of a real world light set is shown by the word as visible in Fig. 1b), or (ii) ADAPTIVE-ASSOCIATIONS (for example, for the word candle – "rosoku" in Japanese – a merged 3D virtual object/a combined metaphor of a rose inside a sock is shown by the word as visible in Fig. 1f). After familiarising themselves with the content in the AC stage, the learner proceeds to the active experimentation (AE) stage.

In the AE stage, the system shows each English phrase with a puzzle task to generate the corresponding Japanese phrase made of three parts from six (6) possibilities offered. The puzzle is also presented in two ways based on the TYPE of instructions: (i) a word puzzle with text components only (Fig. 1c) or (ii) a word puzzle together with 3D virtual objects for the learner to create associations (Fig. 1e) from a given set of objects. If the learner does not know the solution to the puzzle they can skip that phrase. Feedback for the correct and incorrect phrases is given as text and highlights as shown in Fig. 1c and Fig. 1e. There, if the answer is incorrect, it highlights the given answer in red and shows the correct answer separately.

After completing the puzzles related to all four phrases, the system nudges the learner to proceed to the concrete experience stage (CE). A 3D avatar is shown asking the learner to recall each phrase in Japanese (Fig. 1d) by speaking it aloud. To evaluate the spoken answers, the IBM Watson speech recognition was not accurate enough, so a Wizard of Oz technique was used for the experiment. If the phrase is correctly voiced, the corresponding 3D AR model is displayed. The learner can manipulate the object and place it anywhere in the room as they would be able to do it in a real-world experiential learning scenario with physical objects.

If all the phrases are not recalled correctly in the CE stage, the system automatically moves to the reflective observation (RO) stage. There, the learner can first go through all the phrases that include corresponding visual models and audio pronunciation, and then recall

---

[1] https://www.magicleap.com/

[2] https://unity.com/

[3] https://docs.microsoft.com/en-us/windows/mixed-reality/mrtk-unity

[4] https://cloud.ibm.com/apidocs/speech-to-text

them by selecting the correct Japanese phrase (out of four phrases) corresponding to the given English phrase (Fig. 1a and g). The system then returns back to the AC stage.

In this way, the system nudges the learner to proceed further in the experiential learning cycle and consequently on the learning spiral until all four (4) phrases are recalled correctly in the CE stage. In the FIXED-AMOUNT condition, all learners have to repeat all phrases in all stages in each cycle in the same way. In the ADAPTIVE-AMOUNT condition, the learner has to go only through and repeat previously incorrectly recalled phrases (in the CE stage) as the content related to previously correctly recalled phrases is no longer shown.

One of the main features of the Arigatō prototype is the possibility to move digital objects around the physical space as well as move various elements of the interface (e.g. puzzle) in order to complete the tasks given. It is this manipulation of the digital elements that can be brought in the visual field and removed that supports experiential learning or learning by doing.

3.3 Participants

In total, 28 participants were recruited for the study by invitation in a mailing list and web page as well as by snowball sampling. All were students and staff members from our university whose first language is one of the Indo-European languages. None of the participants had any prior knowledge in the Japanese language (identified via a short competency test questionnaire). The between-subject experimental sample comprised 14 participants in the FIXED-AMOUNT condition with 6 (46%) females and 14 participants in the ADAPTIVE-AMOUNT condition with 8 (54%) females. All participants were between 18 to 31 years old ($\bar{x} = 25$) and randomly assigned to one of the two conditions. The study was approved by the local Research Ethics Board.

3.4 Procedure

Participants were first given a consent form to sign, together with the Participant Information Sheet (PIS) outlining the entire research process, and were given an opportunity to ask any question related to the study. They were also instructed that they could abandon the study at any stage. Next, they were requested to fill out the Questionnaire on Current Motivation (QCM).

Before starting the actual task, they completed a five minute training session on a demo application to understand the interface and the interaction with the system. Participants were then instructed to complete a language learning task on one topic (Christmas or birthday celebration) in one condition (either FIXED-ASSOCIATIONS or ADAPTIVE-ASSOCIATIONS) and the other topic in the other condition (see within-subject part of the Fig. 3). After finishing each topic, participants filled out a mental effort questionnaire and a recall questionnaire (to assess their immediate recall) asking participants to remember phrases they just tried learn. They were given a 5 minutes break in between the topics.

In addition, at the end of the study, participants answered two standard questionnaires: a system usability (SUS) [56] and a user experience questionnaire (UEQ) [85]. They also filled out a short post-questionnaire with demographic questions, questions about previous experience with AR technology, and questions about their vision. The whole experiment lasted from 60 to 75 minutes.

One week after the study, participants were again requested to answer the same recall questionnaire (as just after finishing the study) to assess their delayed recall.

3.5 Data Collection

In all conditions, the performance progression data and the time stamp data were logged by the system as a part of behaviour engagement. To measure the motivation, the short form of the Questionnaire on Current Motivation (QCM) with 12 items/questions [31, 78] was used. QCM measures anxiety, challenge, interest, and probability of success on a five-point Likert scale ranging from 1 ("strongly disagree") to 5 ("strongly agree"). Instead of focusing in individuals sub-dimensions (i.e., anxiety, challenge, interest, and probability of success), we used the mean score of the 12 items as an indicator of the overall motivation.

For measuring mental effort we used a standard 8-questions questionnaire from [43] that focuses on mental load for learning contexts. To ensure comparability with other studies while maintaining coherence with our tasks, we just changed the wording to reflect the task at hand and better fit our learning scenario. In addition, we added 5 questions relevant to our study (How much effort did you invest in (i) memorising the words, (ii) understanding particles, (iii) recalling the memorised words, (iv) building the sentences with puzzle blocks, and in (v) speaking the sentences out loud?). All 13 questions were answered on a nine-point scale. The mean scores of the 8 (MEQ8) and the 5 (MEQ5) answers as well as the overall mean score (MEQ13) were used as indicators for the mental effort.

The recall questionnaire (immediate and delayed) included Japanese phrases the participants practised during the learning tasks (e.g., "How do you say 'invite friends' in Japanese?", "How do you say 'family gathering' in Japanese?" with the result marked as either correct or incorrect).

To measure the usability of the system, we used the System Usability Scale (SUS) questionnaire [14] (10 questions answered on a five-point Likert scale). The SUS scores were calculated with the standard SUS analysis. For measuring the user experience, we used the short version of the User Experience Questionnaire (UEQ-S) [84, 85] with eight items/questions. The first four represent pragmatic qualities (Perspicuity, Efficiency and Dependability) and the last four hedonic qualities (Stimulation and Novelty) [84]. The results are converted to a range between -3 to 3.

To assess the reliability of motivation and mental effort questionnaires, we performed Cronbach's alpha test. Estimated reliability for each questionnaire (motivation Cronbach's $\alpha = 0.71$ and mental effort $\alpha = 0.89$) was acceptable for the research purposes [10]. For the recall questionnaire, the reliability was measured using the Kuder-Richardson 20 test [53] because of the binary nature of the results (correct/incorrect). The $KR = 0.76 > 0.5$ value indicates that the reliability of the recall questionnaire was also acceptable.

3.6 Data Analysis

Each data set collected in the study was first checked for normality using the Shapiro–Wilk normality test [88].

The analysis was done in R studio, using "WRS2" R package. For immediate recall, delayed recall, mental effort, task completion time and learning efficiency, the statistical significance was examined using a mixed between-within subjects ANOVA on the 20% trimmed means–"bwtrim" [61]. The main between-subjects effect (group comparisons), the main within-subjects effect (e.g., due to repeated measurements), and the interaction effect, were computed using the "sppba", "sppbb", and "sppbi" functions respectively [61]. Statistical significance for motivation was examined using the Mann-Whitney U test [91].

The resulting $p < 0.05$ are reported as statistically significant. All boxplots use a 1.5xIQR (interquartile range) rule and Tukey's fences [92] for whiskers and identified outliers. Asterisk notation is used in figures to visualise statistical significance (ns: $p > 0.05$, *: $p < 0.05$, **: $p < 0.01$ and ***: $p < 0.001$).

We also conducted a power analysis to check and validate the results and findings of the study. We calculated the effect size (Cohen's d) for each data set collected [21], selected the minimum effect size (Cohen's $d = 1.251$) and estimated the statistical power ($1 - \beta = 0.8$) of data to check whether the type II error probability (β) is within an acceptable range for a given sample size ($n = 14$ per group) and a significance level ($\alpha = 0.05$). The estimated power value 0.8 shows that with the given sample size, we can have a 80% chance that we correctly reject the null hypothesis with a significance level 0.05.

Finally, we used a Pearson's multiple correlation test [73] to find out whether there were any correlations between a learner's mental effort, task completion time, motivation and their performances (i.e., recall and efficiency).

4 RESULTS

The results of the analysis based on the input from the 28 participants who completed the study is presented here. Before dwelling further into

the effects of guidance on various measures we have checked whether the motivation has had an effect on the study. The results show that the motivation had no statistically significant effect on the rest of the results presented in this section (Mann-Witney U test, $U(N_{\text{FIXED-AMOUNT}} = 14, N_{\text{ADAPTIVE-AMOUNT}} = 14) = 90.20$, $p > 0.05$). The following sections present the effect of adaptive guidance (amount and type) on learning performance (immediate and delayed recall), mental effort, task completion time, and learning efficiency (immediate and delayed). The last section presents correlations between these measures.

4.1 The Effect of Guidance on Learning Performance

Learning performance was measured with the recall questionnaires right after the study (immediate) and after 7 days (delayed). The results are presented in the following two sections.

4.1.1 Immediate recall

The effect of the adaptive guidance – i.e., the AMOUNT (FIXED and ADAPTIVE) and TYPE (FIXED and ADAPTIVE) – on immediate recall is shown in Fig. 5 (top left). The data summarised in Fig. 5 (bottom left) is analysed using a mixed between-within subjects ANOVA on the 20% trimmed means [61].

The results in Fig. 5 (bottom left) show that the effects of the AMOUNT and TYPE of guidance on immediate recall are statistically significant. The effect of the AMOUNT of guidance alone ($df = 14.19$, $p < 0.001$) is significant. Fig. 5 (top left) indicates that the immediate recall in the FIXED-AMOUNT condition ($\bar{x} = 83.04\%$, $SD = 16.44$) is significantly better compared to the ADAPTIVE-AMOUNT condition ($\bar{x} = 66.7\%$, $SD = 19.15$).

Similarly, the effect of the TYPE on immediate recall ($df = 17.88$, $p < 0.001$) is also significant. Furthermore, Fig. 5 (top left) indicates that the immediate recall in the ADAPTIVE-ASSOCIATIONS condition ($\bar{x} = 86.61\%$, $SD = 11.57$) is significantly better compared to the FIXED-ASSOCIATIONS condition ($\bar{x} = 62.5\%$, $SD = 14.42$). The interaction effect of the AMOUNT and TYPE on the immediate recall ($df = 17.88$, $p > 0.05$) is not statistically significant.

4.1.2 Delayed recall

The effect of the adaptive guidance (AMOUNT and TYPE) on delayed recall is shown in Fig. 5 (top centre). The data summarised in Fig. 5 (bottom centre) is analysed using a mixed between-within subjects ANOVA on the 20% trimmed means [61].

The results show that the effects of the AMOUNT and TYPE of guidance on delayed recall are statistically significant. The effect of the AMOUNT of guidance alone ($df = 15.13$, $p < 0.01$) is also significant. Fig. 5 (top centre) indicates that the delayed recall in the FIXED-AMOUNT condition ($\bar{x} = 70.54\%$, $SD = 23.20$) is significantly better compared to ADAPTIVE-AMOUNT condition ($\bar{x} = 50.00\%$, $SD = 23.15$).

The effect of the TYPE on the delayed recall ($df = 17.68$, $p < 0.001$) is also significant. Fig. 5 (top centre) indicates that the delayed recall in the ADAPTIVE-ASSOCIATION condition ($\bar{x} = 73.21\%$, $SD = 18.12$) is significantly better compared to FIXED-ASSOCIATION condition ($\bar{x} = 47.32\%$, $SD = 20.24$). The interaction effect of the AMOUNT and TYPE on the delayed recall ($df = 17.68$, $p > 0.05$) is not significant.

4.2 The Effect of Guidance on Mental Effort

The effect of the adaptive guidance (AMOUNT and TYPE) on mental effort is shown in Fig. 5 (top right). The data summarised in Fig. 5 (bottom right) is analysed using a mixed between-within subjects ANOVA on the 20% trimmed means [61].

The results show that the effects of the AMOUNT and TYPE of guidance on mental effort are statistically significant. The effect of the AMOUNT of guidance alone (MEQ13 $df = 17.47$, $p < 0.01$; MEQ08 $df = 16.09$, $p < 0.01$; MEQ05 $df = 14.73$, $p < 0.01$) is significant. Fig. 5 (top right) indicates that the overall mental effort in the ADAPTIVE-AMOUNT condition ($\bar{x} = 3.66\%$, $SD = 1.06$) is significantly lower compared to the FIXED-AMOUNT condition ($\bar{x} = 4.61$, $SD = 0.98$).

The effect of the TYPE alone is also significant (MEQ13 $df = 17.78$, $p < 0.001$; MEQ08 $df = 12.58$, $p < 0.01$; MEQ05 $df = 17.73$, $p < 0.01$). Fig. 5 (top right) indicates that the overall mental effort in the ADAPTIVE-ASSOCIATIONS condition ($\bar{x} = 3.39$, $SD = 0.63$) is significantly lower compared to the FIXED-ASSOCIATIONS condition ($\bar{x} = 4.88$, $SD = 0.73$). The interaction effect of the AMOUNT and TYPE on mental effort (MEQ13 $df = 17.78$, $p > 0.05$; MEQ08 $df = 12.58$, $p > 0.05$; MEQ05 $df = 17.73$, $p > 0.05$) is not statistically significant.

4.3 The Effect of Guidance on Task Completion Time

The results for the task completion time presented in Fig. 6 (bottom left) are analysed using a mixed between-within subjects ANOVA on the 20% trimmed means. The results indicate that the AMOUNT of guidance has no statistically significant effect on the task completion time ($df = 15.29$, $p > 0.05$). It can thus not be concluded that participants in the ADAPTIVE-AMOUNT condition completed the task faster compared to participants in the FIXED-AMOUNT condition.

In contrast, the effect of the TYPE on the task completion time is significant ($df = 17.88$, $p < 0.001$) as can be also observed in Fig. 6 (top left). This also shows that the task completion time in the ADAPTIVE-ASSOCIATIONS condition ($\bar{x} = 15.78\ min$, $SD = 3.17$) is significantly lower compared to the FIXED-ASSOCIATION condition ($\bar{x} = 27.13\ min$, $SD = 6.69$). The interaction effect of the AMOUNT and TYPE on the task completion time is not statistically significant ($df = 17.88$, $p > 0.05$).

4.4 The Effect of Guidance on Learning Efficiency

The learning efficiency for each study condition was determined using Equation 1 [18, 34, 71].

$$E = \frac{z_P - z_M}{\sqrt{2}} \qquad (1)$$

E = Learning efficiency
z_P = Average performance in Z-scores
z_M = Average task difficulty in Z-scores

We measured the performance of each study condition based on the recall scores participants obtained after completing the task for that study condition. We estimated the difficulty of the task based on the mental effort questionnaire for that study condition. We then calculated the learning efficiency for each of the four conditions: (a) FIXED-AMOUNT of guidance with FIXED-ASSOCIATIONS, (b) FIXED-AMOUNT of guidance with ADAPTIVE-ASSOCIATIONS, (c) ADAPTIVE-AMOUNT of guidance with FIXED-ASSOCIATIONS, and (d) ADAPTIVE-AMOUNT of guidance with ADAPTIVE-ASSOCIATIONS using Formula 1. This was done for both immediate recall (immediately after participants had completed the task) and delayed recall (a week after participants had completed the task).

4.4.1 Immediate Efficiency

The immediate recall learning efficiency results across the study conditions are depicted in Fig. 7 (left). It shows that, for both FIXED-AMOUNT and ADAPTIVE-AMOUNT, the ADAPTIVE-ASSOCIATIONS condition is more efficient (i.e., has a higher efficiency value) compared to the FIXED-ASSOCIATIONS condition.

The effects of the AMOUNT and TYPE of guidance, on immediate recall learning efficiency are shown in Fig. 6 (top centre). The data is analysed using a mixed between-within subjects ANOVA on the 20% trimmed means (Fig. 6 (bottom centre)). There is no significant effect of the AMOUNT of guidance conditions on participants' immediate recall learning efficiency ($df = 16.80$, $p = 0.460$).

By contrast, the effect of the TYPE of instructions on immediate recall learning efficiency is statistically significant ($df = 16.80$, $p < 0.001$). Fig. 6 (top centre) indicates that the immediate recall learning efficiency in the ADAPTIVE-ASSOCIATIONS condition is significantly higher compared to the FIXED-ASSOCIATIONS condition. The interaction effect of the AMOUNT and TYPE on the immediate recall learning efficiency is not statistically significant ($df = 16.80$, $p > 0.05$).
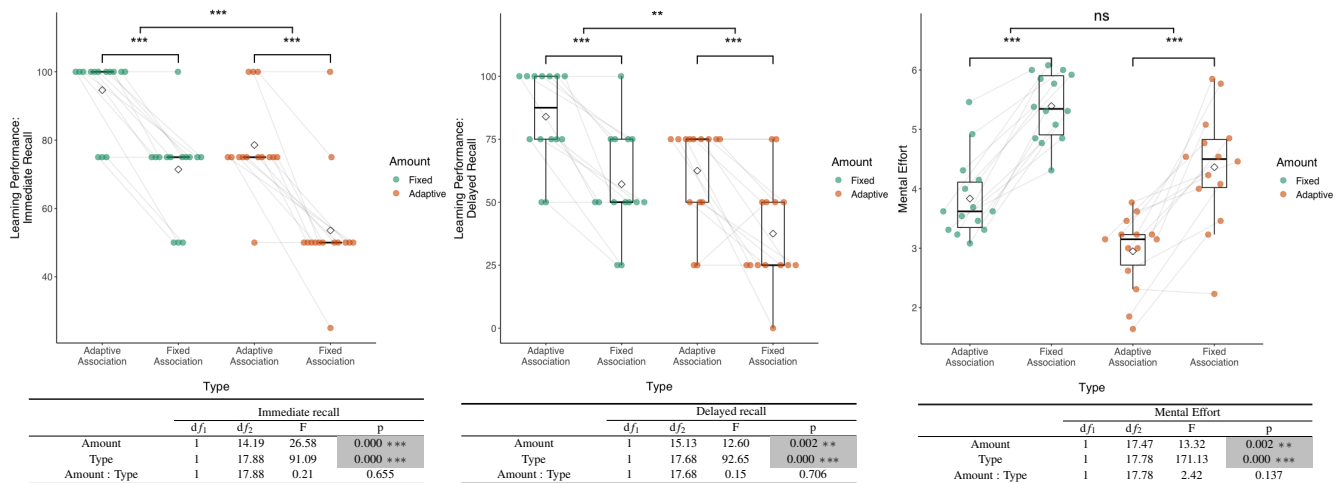
Fig. 5. Left: Immediate recall in percentage of correctly remembered phrases. Centre: Delayed recall in percentage of correctly remembered phrases. Right: Overall mental effort (MEQ13) invested during the task. Below are results form mixed between-within subjects ANOVA on the 20% trimmed means examining the between-subjects (AMOUNT), the within-subjects (TYPE) and, the interaction (AMOUNT:TYPE) effects, over dependent variables.
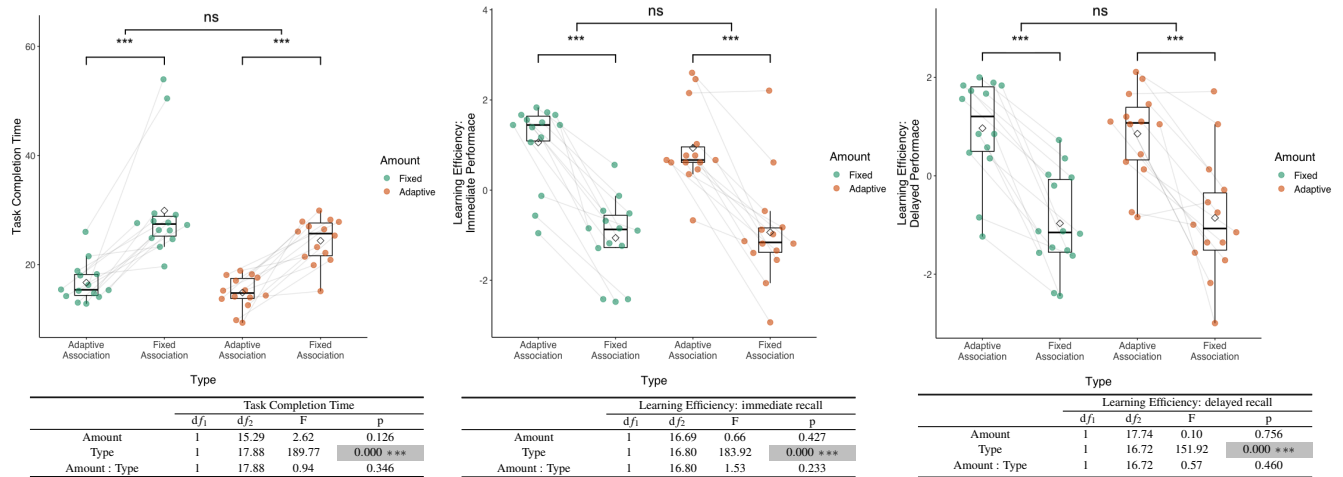

Fig. 6. Left: Task completion time in minutes. Centre: Average learning efficiency for immediate performance. Right: Average learning efficiency for delayed performance. Below are results form mixed between-within subjects ANOVA on the 20% trimmed means examining the between-subjects (AMOUNT), the within-subjects (TYPE) and, the interaction (AMOUNT:TYPE) effects, over dependent variables.
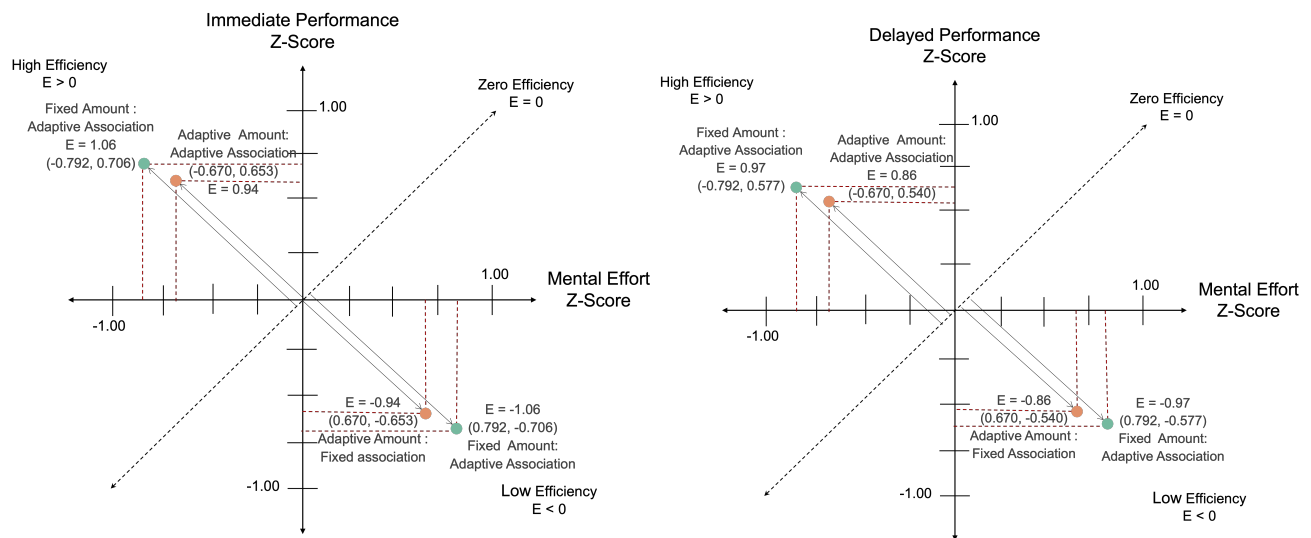

Fig. 7. Left: Immediate recall learning efficiency. Right: Delayed recall learning efficiency.

4.4.2 Delayed Efficiency

The delayed recall learning efficiency results across the study conditions are shown in Fig. 7 (right). The figure indicates that, for both FIXED-AMOUNT and ADAPTIVE-AMOUNT, the ADAPTIVE-ASSOCIATION is more efficient (i.e., has a higher efficiency value) compared to the FIXED-ASSOCIATIONS condition.

The effects of the AMOUNT and TYPE of guidance on delayed recall learning efficiency are illustrated in Fig. 6 (top right).

There is no significant effect of the AMOUNT of guidance condition on the delayed recall learning efficiency ($df = 16.72$, $p = 0.460$).

By contrast, the effect of the TYPE of instructions on the delayed recall learning efficiency is statistically significant ($df = 16.72$, $p < 0.001$). The results also show that the delayed recall learning efficiency is significantly higher for ADAPTIVE-ASSOCIATIONS compared to FIXED-ASSOCIATIONS condition. No statistically significant interaction effect was found between conditions ($df = 16.72$, $p > 0.05$).

4.5 System Usability and User Experience

The average SUS score shows that the system usability is in an acceptable range (FIXED-AMOUNT = 82.3, ADAPTIVE-AMOUNT = 85.8; $SUS > 70$). The results for UEQ show that pragmatic (FIXED-AMOUNT = 1.54, ADAPTIVE-AMOUNT = 1.75) and hedonic qualities (FIXED-AMOUNT = 2.33, ADAPTIVE-AMOUNT = 2.52), as well as overall user experience (FIXED-AMOUNT = 1.93, ADAPTIVE-AMOUNT = 2.14) are perceived as strongly positive in both conditions. In all cases we see that the adaptive-amount has higher values.

4.6 Correlations between Measures

A Pearson multiple correlation test was applied to find out whether or not there were any correlations between the measures (recall, efficiency, mental effort, task completion time). The results are presented in Table 2.

The results for the FIXED-AMOUNT condition indicate that in both FIXED-ASSOCIATIONS and ADAPTIVE-ASSOCIATIONS conditions, the mental effort is negatively related to both immediate and delayed recall as well as to both immediate and delayed efficiency. The task completion time measure does not show any relationship to other variables. The correlations between immediate/delayed and recall/efficiency variables was expected since these measures are all related.

Similar results can be observed for ADAPTIVE-AMOUNT condition. In both FIXED-ASSOCIATIONS and ADAPTIVE-ASSOCIATIONS conditions, the mental effort is negatively related to both immediate and delayed recall as well as to both immediate and delayed efficiency. The task completion time does also not show any relationship to other variables. While, as before, immediate/delayed and recall/efficiency correlations are not surprising.

5 DISCUSSION

Arigatō was developed to compare how learners perform in learning a new language, when presented with different adaptation factors of AR based instructional guidance in an experiential learning environment. The AR system was deliberately selected as the most probable future technology that will be used in the classroom. An AR classroom has several advantages over other types of digitised classrooms such as taking the real world into consideration and consequently embedding information directly into the user's field of view. These advantages better support real-world in-person communication and group collaboration compared to other technologies AR is often contrasted to, such as desktop, tablet computers, and VR [69, 80].

By taking the real world into consideration, our prototype did not just show content on the HMD, but allowed users to move and place AR objects in their physical surroundings. With this functionality, AR was used to replicate the real-world experiential learning scenario (e.g., learning by decorating a room for Christmas celebrations with AR objects as can be done with real objects) and show information in a coherent and meaningful way within the real world context.

Arigatō allowed us to vary (i) the AMOUNT of guidance, which is either FIXED or an ADAPTIVE-AMOUNT (the amount changes based

on participants' performance), and (ii) the TYPE of association, which is either FIXED or with ADAPTIVE-ASSOCIATIONS (participants can select associations by themselves from a predefined set). We used a 2×2 mixed design to evaluate the study conditions, as explained in Sect. 3.1. The AMOUNT of guidance was evaluated as a within-subjects variable while the TYPE of instructions as a between-subjects variable. In this paper, the impact of the guidance method was measured through performance (i.e., recall and efficiency) and engagement (i.e., task completion time, mental effort, and motivation).

We measured behavioural engagement for FIXED guidance and ADAPTIVE guidance by observing how participants interacted with Arigatō and how much time they spent in each condition. We could observe participants being more interested in the ADAPTIVE-ASSOCIATIONS than the FIXED-ASSOCIATIONS condition, even though they completed the task in less time. This was expected, since keyword methods have been suggested as motivating learners to be more creative and to enjoy using their minds more productively [79].

We measured cognitive engagement for all the study conditions by analysing how much mental effort participants invested in completing the task. We found that the mental effort with the ADAPTIVE-AMOUNT of guidance was significantly lower compared to the FIXED-AMOUNT of guidance. This is in line with the literature, as prior work has shown that repeating the same information could increase the amount of cognitive work and consequently the mental effort [82]. We further identified that the mental effort in the ADAPTIVE-ASSOCIATIONS condition was lower than in FIXED-ASSOCIATIONS condition. Again, this was somewhat expected, as the association method connects new learning to prior knowledge through the use of visual and/or acoustic cues [62, 75], which could lower the mental effort. However, we have shown that in the context of language learning the AR environment can support the creation of adaptive constructs not only in the form of the keywords (the "keyword method" [11]) and mental associations, but also in combination with 3D digital objects that users can manipulate with. It is this combination that supports experiential learning through manipulation of digital objects and that outperformed the traditional "keyword method" in our study.

We measured learning performance through analysis of recall (immediate recall after the study and delayed recall test 7 days after the study) and learning efficiency. Interestingly, we found that compared to the ADAPTIVE-AMOUNT of guidance, the FIXED-AMOUNT of guidance group scored better in the immediate and delayed recall tests, although they have invested a significantly higher mental effort during the task. In addition, we found that the ADAPTIVE-ASSOCIATIONS outperformed the FIXED-ASSOCIATIONS group, in the immediate, delayed recall tests, and learning efficiency, while investing a significantly lower mental effort and spending a less amount of time during the task. These findings suggest avenues for future work in archetypal user models, which might allow users to slowly adapt pre-developed personal models from a standard set of ideal users. What we would refer to as the "follow the leader" learning approach.

We have also investigated how different variables are correlated and how they could be combined in order to more accurately present engagement and learning performances of the participants' for each study condition. This contrasts to the existing literature in learning that often looks at these aspects individually. Our results show correlations between *mental effort and learning performances (both recall and efficiency)* in all conditions, which was to be expected since the two are considered related. Interestingly, there were no correlations found between *task completion time and learning performance* in any condition. Again, these findings suggest potential future work in "transfer motivation" between individual learners or between instructors and learners.

6 LIMITATIONS AND FUTURE WORK

The average age of participants in our study was 25 years, which presents a possible age bias. Despite this, this age group is worth studying as it is highly mobile, spending an extended period of time in a foreign speaking country (for example, the EU Erasmus+ programme alone funds more than half a million exchanges yearly [27]). As such,

Table 2. Correlations between different measures for the FIXED-AMOUNT of guidance condition (left) and the ADAPTIVE-AMOUNT of guidance condition (right). ME: mental effort, CT: task completion time, IR: immediate recall, DR: delayed recall, IE: immediate efficiency and DE: delayed efficiency.

FIXED-AMOUNT of guidance					
FIXED-ASSOCIATIONS					
ME	CT	IR	DR	IE	DE
ME 1					
CT −0.233 ns	1				
IR −0.627 *	−0.014 ns	1			
DR −0.819 ***	0.046 ns	0.797 ***	1		
IE −0.863 ***	0.097 ns	0.935 ***	0.891 ***	1	
DE −0.931 ***	0.125 ns	0.764 **	0.972 ***	0.920 ***	1
ADAPTIVE-ASSOCIATIONS					
ME	CT	IR	DR	IE	DE
ME 1					
CT 0.154 ns	1				
IR −0.846 ***	−0.075 ns	1			
DR −0.903 ***	−0.338 ns	0.745 **	1		
IE −0.963 ***	−0.121 ns	0.958 ***	0.860 ***	1	
DE −0.972 ***	−0.260 ns	0.812 ***	0.979 ***	0.931 ***	1

*** Correlation is significant at 0.001 level, ($p < 0.001$)
** Correlation is significant at 0.01 level, ($p < 0.01$)
* Correlation is significant at 0.05 level, ($p < 0.05$)
ns Correlation is not significant, ($p > 0.05$)

ADAPTIVE-AMOUNT of guidance					
FIXED-ASSOCIATIONS					
ME	CT	IR	DR	IE	DE
ME 1					
CT −0.546 ns	1				
IR −0.820 ***	0.278 ns	1			
DR −0.813 ***	0.393 ns	0.679 **	1		
IE −0.956 ***	0.436 ns	0.951 ***	0.784 ***	1	
DE −0.952 ***	0.493 ns	0.786 ***	0.953 ***	0.913 ***	1
ADAPTIVE-ASSOCIATIONS					
ME	CT	IR	DR	IE	DE
ME 1					
CT −0.208 ns	1				
IR −0.884 ***	0.135 ns	1			
DR −0.618 *	0.432 ns	0.568 *	1		
IE −0.966 ***	0.174 ns	0.975 ***	0.609 *	1	
DE −0.863 ***	0.375 ns	0.777 ***	0.930 ***	0.841 ***	1

*** Correlation is significant at 0.001 level, ($p < 0.001$)
** Correlation is significant at 0.01 level, ($p < 0.01$)
* Correlation is significant at 0.05 level, ($p < 0.05$)
ns Correlation is not significant, ($p > 0.05$)

this group could benefit from an improved language learning system. Nevertheless, the results cannot be generalised over the whole population, and expanding the study to other age groups and exploring the effect of age on the proposed learning system is an important future direction.

Another bias is the self-selection bias because the participants who opted-in are likely to be at ease when learning a new language. However, even if this study attracted only a particular group of language learners, the results still show the benefit of AR guidance with adaptive associations that would most likely benefit also other users in our studied age group. However, this needs to be further investigated. Using questionnaires for measuring certain aspects such as motivation and mental effort can result in social desirability bias as users could answer them to please the researcher. However, we used well established questionnaires together with measurements that can hardly be affected by such bias (i.e., recall and efficiency and task completion time).

The Arigatō prototype covers the first three levels (categories) of the Bloom's taxonomy [1] only: Knowledge (remembering the words, sentences), Comprehension (understanding the structure of sentences), and Application (applying or speaking out the sentences). The expansion to higher levels of the taxonomy (Analysis (analyse), Synthesis, and Evaluation (evaluate), Creation) is also relevant and we aim to take this into consideration in future implementations. In addition, we used the task of learning vocabulary and other language constructs as a use-case to explore the potential of AR in adaptive learning scenarios. In future, this use-case should be expanded to other use-cases to confirm its generalisability.

7 CONCLUSION

In this paper we present the Arigatō (Augmented Reality Instructional Guidance & Tailored Omniverse) prototype – an adaptive guidance augmented reality (AR) system for language learning. AR has been identified as an ideal platform to support simulating various experiences for experiential learning and we explored the AR's design space in the learning context rather than comparing it to other technologies. With our prototype we investigated how the amount of guidance (fixed vs. adaptive-amount) and the type of guidance (fixed vs. adaptive-associations) affects the engagement and consequently the learning outcomes of language learning in an AR environment.

Compared to the adaptive-amount, the fixed-amount of guidance group scored better in the immediate and delayed (after 7 days) recall tests. However, this group also invested a significantly higher mental effort to complete the task. Adaptive-associations group outperformed the fixed-associations group in the immediate, delayed (after 7 days) recall tests, and learning efficiency. The adaptive-associations group also showed significantly lower mental effort and spent less time to complete the task. Both results hint at potential for the archetypal user models in comparable learning scenarios, which might allow users to "transfer motivation" between individual learners or between instructors and learners, and slowly adapt pre-developed personal models. Such approach could achieve the balance between the (adaptive) amount and (adaptive) type to optimise the mental effort need to complete the learning task.

The results also show the potential of AR in adaptive learning scenarios where the learning environment needs to be simulated and where virtual content needs to be added or removed on the fly based on the learning needs. Learning vocabulary and other language constructs as a use-case explored in this study, presents just one of many possibilities that offer venues for future work in AR classrooms as it is likely that the technology will get better and widely accessible in the future [58].

ACKNOWLEDGMENTS

REFERENCES

[1] A taxonomy for learning, teaching, and assessing; a revision of bloom's taxonomy of educational objectives, complete edition, 2001.
[2] *Implicit emotional tagging of multimedia using EEG signals and brain computer interface*, 2009.
[3] *An Adaptive AR Tutor For Cabling a Network Topology.*, 2017.
[4] O. Alrashidi, H. P. Phan, and B. H. Ngu. Academic engagement: An overview of its definitions, dimensions, and major conceptualisations. *International Education Studies*, 9(12):41–52, 2016.
[5] American Institute of Physics. *What levels of guidance promote engaged exploration with interactive simulations?*, vol. 1064, 2008.

[6] M. Amiryousefi and S. Ketabi. Mnemonic instruction: A way to boost vocabulary learning and recall. *Journal of Language Teaching & Research*, 2(1), 2011.

[7] J. R. Anderson, C. F. Boyle, and B. J. Reiser. Intelligent tutoring systems. *Science*, 228(4698):456–462, 1985.

[8] J. J. Appleton, S. L. Christenson, and M. J. Furlong. Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools*, 45(5):369–386, 2008.

[9] P. Arvanitis and P. Krystalli. Digital storytelling and augmented reality-based scenarios for foreign language teaching. In *Digital Pedagogies and the Transformation of Language Education*, pp. 1–22. IGI Global, 2021.

[10] D. Ary, L. C. Jacobs, C. K. S. Irvine, and D. Walker. *Introduction to research in education.* Cengage Learning, 2018.

[11] R. C. Atkinson. Mnemotechnics in second-language learning. *American psychologist*, 30(8):821, 1975.

[12] J. Birt, Z. Stromberga, M. Cowling, and C. Moro. Mobile mixed reality for experiential learning and simulation in medical and health sciences education. *Information*, 9(2):31, 2018.

[13] D. Boud, R. Cohen, and D. Walker. *Using experience for learning.* McGraw-Hill Education (UK), 1993.

[14] J. Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

[15] J. S. Bruner. The act of discovery. *Harvard educational review*, 1961.

[16] A. I. Canhoto and J. Murphy. Learning from simulation design to develop better experiential learning initiatives: An integrative approach. *Journal of Marketing Education*, 38(2):98–106, 2016.

[17] K. J. Carbonneau and S. C. Marley. Instructional guidance and realism of manipulatives influence preschool children's mathematics learning. *The Journal of Experimental Education*, 83(4):495–513, 2015.

[18] R. C. Clark, F. Nguyen, and J. Sweller. *Efficiency in learning: Evidence-based guidelines to manage cognitive load.* John Wiley & Sons, 2011.

[19] R. E. Clark. How much and what type of guidance is optimal for learning from instruction. *Constructivist instruction: Success or failure*, pp. 158–183, 2009.

[20] A. D. Cohen and E. Aphek. Retention of second-language vocabulary overtime: Investigating the role of mnemonic associations. *System*, 8(3):221–235, 1980.

[21] J. Cohen. Statistical power analysis for the behavioral sciences. *England: Routledge*, 1988.

[22] L. J. Cronbach and R. E. Snow. *Aptitudes and instructional methods: A handbook for research on interactions.* Irvington, 1977.

[23] F.-A. Dita. A foreign language learning application using mobile augmented reality. *Informatica Economica*, 20(4):76, 2016.

[24] F. Draxler, A. Labrie, A. Schmidt, and L. L. Chuang. Augmented reality to enable users in learning case grammar from their real-world interactions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2020.

[25] J. Dunlosky, K. A. Rawson, E. J. Marsh, M. J. Nathan, and D. T. Willingham. Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public interest*, 14(1):4–58, 2013.

[26] N. C. Ellis. The psychology of foreign language vocabulary acquisition: Implications for call. *Computer Assisted Language Learning*, 8(2-3):103–128, 1995.

[27] Erasmus+. Factsheets and statistics on Erasmus+. https://erasmus-plus.ec.europa.eu/node/2585. 2022-05-24.

[28] G. Ferrero, F. Bichai, and M. Rusca. Experiential learning through role-playing: Enhancing stakeholder collaboration in water safety plans. *Water*, 10(2):227, 2018.

[29] C. A. Frantzidis, C. Bratsas, C. L. Papadelis, E. Konstantinidis, C. Pappas, and P. D. Bamidis. Toward emotion aware computing: an integrated approach using multichannel neurophysiological recordings and affective visual stimuli. *IEEE transactions on Information Technology in Biomedicine*, 14(3):589–597, 2010.

[30] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris. School engagement: Potential of the concept, state of the evidence. *Review of educational research*, 74(1):59–109, 2004.

[31] P. A. Freund, J.-T. Kuhn, and H. Holling. Measuring current achievement motivation with the qcm: Short form development and investigation of measurement invariance. *Personality and Individual Differences*, 51(5):629–634, 2011.

[32] T. P. Grantcharov, V. B. Kristiansen, J. Bendix, L. Bardram, J. Rosenberg, and P. Funch-Jensen. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *British journal of surgery*, 91(2):146–150, 2004.

[33] J.-Q. Guan, L.-H. Wang, Q. Chen, K. Jin, and G.-J. Hwang. Effects of a virtual reality-based pottery making approach on junior high school students' creativity and learning engagement. *Interactive Learning Environments*, pp. 1–17, 2021.

[34] A. K. Halabi. Applying an instructional learning efficiency model to determine the most efficient feedback for teaching introductory accounting. *Global Perspectives on Accounting Education*, 3(1):6, 2006.

[35] P. Herman and L. M. Gomez. Taking guided learning theory to school: Reconciling the cognitive, motivational, and social contexts of instruction. pp. 74–93, 2009.

[36] G. Huang, X. Qian, T. Wang, F. Patel, M. Sreeram, Y. Cao, K. Ramani, and A. J. Quinn. Adaptutar: An adaptive tutoring system for machine tasks in augmented reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–15, 2021.

[37] T.-C. Huang, C.-C. Chen, and Y.-W. Chou. Animating eco-education: To see, feel, and discover in an augmented reality-based experiential learning environment. *Computers & Education*, 96:72–82, 2016.

[38] A. Ibrahim, B. Huynh, J. Downey, T. Höllerer, D. Chun, and J. O'donovan. Arbis pictus: A study of vocabulary learning with augmented reality. *IEEE transactions on visualization and computer graphics*, 24(11):2867–2874, 2018.

[39] IEEE. *Remarks on emotion recognition from multi-modal bio-potential signals*, vol. 2, 2003.

[40] M. Jantjies, T. Moodley, and R. Maart. Experiential learning through virtual and augmented reality in higher education. In *Proceedings of the 2018 international conference on education technology management*, pp. 42–45, 2018.

[41] L. Jarmon, T. Traphagan, M. Mayrath, and A. Trivedi. Virtual world teaching, experiential learning, and assessment: An interdisciplinary communication course in second life. *Computers & Education*, 53(1):169–182, 2009.

[42] L. Joplin. On defining experiential education. *Journal of experiential education*, 4(1):17–20, 1981.

[43] S. Joseph. *Measuring cognitive load: A comparison of self-report and physiological methods.* PhD thesis, Arizona State University, 2013.

[44] S. Kalyuga, P. Chandler, and J. Sweller. Learner experience and efficiency of instructional guidance. *Educational Psychology*, 21(1):5–23, 2001.

[45] C. D. Katsis, N. Katertsidis, G. Ganiatsas, and D. I. Fotiadis. Toward emotion recognition in car-racing drivers: A biosignal processing approach. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 38(3):502–512, 2008.

[46] G. Kearsley and B. Shneiderman. Engagement theory: A framework for technology-based teaching and learning. *Educational technology*, 38(5):20–23, 1998.

[47] P. A. Kirschner, J. Sweller, and R. E. Clark. Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist*, 41(2):75–86, 2006.

[48] K. R. Koedinger and V. Aleven. Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19(3):239–264, 2007.

[49] R. A. L. Koelstra. Affective and implicit tagging using facial expressions and electroencephalography., 2012.

[50] D. Kolb. *Experiential learning: experience as the source of learning and development.* Prentice Hall, 1984.

[51] D. A. Kolb. Experiential learning theory and the learning style inventory: A reply to freedman and stumpf. *Academy of Management Review*, 6(2):289–296, 1981.

[52] D. A. Kolb, R. E. Boyatzis, C. Mainemelis, et al. Experiential learning theory: Previous research and new directions. *Perspectives on thinking, learning, and cognitive styles*, 1(8):227–247, 2001.

[53] G. F. Kuder and M. W. Richardson. The theory of the estimation of test reliability. *Psychometrika*, 2(3):151–160, 1937.

[54] D. Kuhn. Is direct instruction an answer to the right question? *Educational psychologist*, 42(2):109–113, 2007.

[55] A. W. Lazonder and R. Harmsen. Meta-analysis of inquiry-based learning: Effects of guidance. *Review of educational research*, 86(3):681–718, 2016.

[56] J. R. Lewis and J. Sauro. The factor structure of the system usability scale. In *International conference on human centered design*, pp. 94–103. Springer, 2009.

[57] L. H. Lewis and C. J. Williams. Experiential learning: Past and present. *New directions for adult and continuing education*, 1994(62):5–16, 1994.

[58] R. A. Liono, N. Amanda, A. Pratiwi, and A. A. Gunawan. A systematic literature review: learning with visual by the help of augmented reality helps students learn better. *Procedia Computer Science*, 179:144–152, 2021.

[59] P.-H. E. Liu and M.-K. Tsai. Using augmented-reality-based mobile learning material in efl english composition: An exploratory case study. *British journal of educational technology*, 44(1):E1–E4, 2013.

[60] S.-J. Lu and Y.-C. Liu. Integrating augmented reality technology to enhance children's learning in marine education. *Environmental Education Research*, 21(4):525–541, 2015.

[61] P. Mair and R. Wilcox. Robust statistical methods in r using the wrs2 package. *Behavior research methods*, 52(2):464–488, 2020.

[62] M. A. Mastropieri and T. E. Scruggs. *Teaching students ways to remember: Strategies for learning mnemonically.* Brookline Books, 1991.

[63] C. Mather, T. Barnett, V. Broucek, A. Saunders, D. Grattidge, and W. Huang. Helping hands: using augmented reality to provide remote guidance to health professionals. In *Context Sensitive Health Informatics: Redesigning Healthcare Work*, pp. 57–62. IOS Press, 2017.

[64] R. E. Mayer. Constructivism as a theory of learning versus constructivism as a prescription for instruction. *Constructivist instruction: Success or failure*, pp. 184–200, 2009.

[65] B. Mergel. Instructional design and learning theory, 1998.

[66] M. D. Merrill. First principles of instruction. *Educational technology research and development*, 50(3):43–59, 2002.

[67] N. Moorhouse, T. Jung, et al. Augmented reality to enhance the learning experience in cultural heritage tourism: An experiential learning cycle perspective. *eReview of Tourism Research*, 8, 2017.

[68] N. Moorhouse, M. C. tom Dieck, and T. Jung. An experiential view to children learning in museums with augmented reality. *Museum Management and Curatorship*, 34(4):402–418, 2019.

[69] A. Morrison, A. Oulasvirta, P. Peltonen, S. Lemmela, G. Jacucci, G. Reitmayr, J. Näsänen, and A. Juustila. Like bees around the hive: A comparative study of a mobile augmented reality map. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, p. 1889–1898. Association for Computing Machinery, New York, NY, USA, 2009. doi: 10.1145/1518701.1518991

[70] D. Mueller and J. M. Ferreira. Marvel: A mixed-reality learning environment for vocational training in mechatronics. In *Proceedings of the Technology Enhanced Learning International Conference (TEL 03)*, 2003.

[71] F. G. Paas and J. J. Van Merriënboer. The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human factors*, 35(4):737–743, 1993.

[72] A. Paivio and A. Desrochers. Mnemonic techniques in second-language learning. *Journal of Educational Psychology*, 73(6):780, 1981.

[73] R. L. Plackett. Karl pearson and the chi-squared test. *International Statistical Review/Revue Internationale de Statistique*, pp. 59–72, 1983.

[74] M. Pressley, J. R. Levin, and H. D. Delaney. The mnemonic keyword method. *Review of Educational Research*, 52(1):61–91, 1982.

[75] A. L. Putnam. Mnemonics in education: Current research and applications. *Translational Issues in Psychological Science*, 1(2):130, 2015.

[76] M. R. Raugh and R. C. Atkinson. A mnemonic method for learning a second-language vocabulary. *Journal of Educational Psychology*, 67(1):1, 1975.

[77] J. Reeve and C.-M. Tseng. Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, 36(4):257–267, 2011.

[78] F. Rheinberg, R. Vollmeyer, and B. D. Burns. Fam: Ein fragebogen zur erfassung aktueller motivation in lern-und leistungssituationen (langversion, 2001). *Diagnostica*, 2:57–66, 2001.

[79] J. T. Richardson and A. Jones. *Mental imagery and human memory.* Macmillan International Higher Education, 1980.

[80] R. Rzayev, S. Korbely, M. Maul, A. Schark, V. Schwind, and N. Henze. Effects of position and alignment of notifications on ar glasses during social interaction. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. Association for Computing Machinery, New York, NY, USA, 2020.

[81] SAGE Publications Sage CA: Los Angeles, CA. *Engagement and Competence in VR and non-VR Environments*, vol. 61, 2017.

[82] G. Salomon. The differential investment of mental effort in learning from different sources. *Educational psychologist*, 18(1):42–50, 1983.

[83] W. B. Schaufeli. What is engagement. *Employee engagement in theory and practice*, 15:321, 2013.

[84] M. Schrepp. Ueq-user experience questionnaire, 2019.

[85] M. Schrepp, A. Hinderks, and J. Thomaschewski. Design and evaluation of a short version of the user experience questionnaire (ueq-s). *IJIMAI*, 4(6):103–108, 2017.

[86] H. Schuman, S. Presser, and J. Ludwig. Context effects on survey responses to questions about abortion. *Public Opinion Quarterly*, 45(2):216–223, 1981.

[87] P. Seedhouse, A. Preston, P. Olivier, J. Dan, H. Philip, B. Madeline, A. Rafiev, and M. Kipling. The european digital kitchen project. *Bellaterra Journal of Teaching & Learning Language & Literature*, 7(1):1–16, 2014.

[88] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965.

[89] Springer. *Time-frequency analysis of EEG signals for human emotion detection*, 2008.

[90] L. P. Steffe and J. E. Gale. *Constructivism in education.* Psychology Press, 1995.

[91] G. M. Sullivan and A. R. Artino Jr. Analyzing and interpreting data from likert-type scales. *Journal of graduate medical education*, 5(4):541–542, 2013.

[92] L. Sullivan. Interquartile range (iqr), 2016.

[93] B. Thoravi Kumaravel, F. Anderson, G. Fitzmaurice, B. Hartmann, and T. Grossman. Loki: Facilitating remote instruction of physical tasks using bi-directional mixed-reality telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 161–174, 2019.

[94] S. Tobias. When do instructional methods. *Educational Researcher*, 11(4):4–9, 1982.

[95] S. Tobias and T. M. Duffy. *Constructivist instruction: Success or failure?* Routledge, 2009.

[96] F. B. Topu and Y. Goktas. The effects of guided-unguided learning in 3d virtual environment on students' engagement and achievement. *Computers in Human Behavior*, 92:1–10, 2019.

[97] K. L. Vaughan, R. E. Vaughan, and J. M. Seeley. Experiential learning in soil science: Use of an augmented reality sandbox. *Natural Sciences Education*, 46(1):1–5, 2017.

[98] C. D. Vazquez, A. A. Nyati, A. Luh, M. Fu, T. Aikawa, and P. Maes. Serendipitous language learning in mixed reality. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 2172–2179, 2017.

[99] M. Weber, J. Giacomin, A. Malizia, L. Skrypchuk, V. Gkatzidou, and A. Mouzakitis. Investigation of the dependency of the drivers' emotional experience on different road types and driving conditions. *Transportation research part F: traffic psychology and behaviour*, 65:107–120, 2019.

[100] M. Weerasinghe, A. Quigley, J. Ducasse, K. Č. Pucihar, and M. Kljun. Educational augmented reality games. pp. 3–32, 2020.

[101] X. Wei, D. Weng, Y. Liu, and Y. Wang. A tour guiding system of historical relics based on augmented reality. In *2016 IEEE Virtual Reality (VR)*, pp. 307–308. IEEE, 2016.

[102] G. Westerfield, A. Mitrovic, and M. Billinghurst. Intelligent augmented reality training for motherboard assembly. *International Journal of Artificial Intelligence in Education*, 25(1):157–172, 2015.

[103] C. Weston and P. A. Cranton. Selecting instructional strategies. *The Journal of Higher Education*, 57(3):259–288, 1986.

[104] C. D. Wickens, J. G. Hollands, S. Banbury, and R. Parasuraman. *Engineering psychology and human performance.* Psychology Press, 2015.

[105] A. F. Wise and K. O'Neill. Beyond more versus less: A reframing of the debate on instructional guidance. 2009.

[106] J. B. Worthen and R. R. Hunt. *Mnemonology: Mnemonics for the 21st century.* Psychology Press, 2011.

[107] S. Yang and B. Mei. Understanding learners' use of augmented reality in language learning: insights from a case study. *Journal of Education for Teaching*, 44(4):511–513, 2018.