ARText: Vocabulary Learning with Personalised and Predefined Keyword-Associations in Augmented Reality

Maheshya Weerasinghe[1*], Matjaž Kljun[1,2*], Nuwan T. Attygalle[1,3], Aaron Quigley[4], Jens Grubert[5], Verena Biener[6], Juri Yoneyama[7], Hirokazu Kato[8], Klen Čopič Pucihar[1,2*]

[1*]University of Primorska, Koper, Slovenia.
[2]Stellenbosch University, Stellenbosch, South Africa.
[3]Université catholique de Louvain, Louvain-la-Neuve, Belgium.
[4]CSIRO Data61, Eveleigh, Australia.
[5]Coburg University of Applied Sciences, Coburg, Germany.
[6]University of Stuttgart, Stuttgart, Germany.
[7]Université de Rennes, Rennes, France.
[8]Nara Institute of Science and Technology, Nara, Japan.

*Corresponding author(s). E-mail(s):
maheshya.weerasinghe@famnit.upr.si; matjaz.kljun@upr.si;
klen.copic@famnit.upr.si;
Contributing authors: nuwan.attygalle@famnit.upr.si;
aaron.quigley@csiro.au; jens.grubert@hs-coburg.de;
verena.biener@visus.uni-stuttgart.de; juri.yoneyama@inria.fr;
kato@is.naist.jp;

Abstract

The "keyword method" is a mnemonic technique that can be used for learning foreign vocabulary by associating a word's meaning with a phonetically similar keyword. For example, the Japanese word for "tree" is "ki", which sounds like "key" (keyword) so one might imagine "a tree with key-shaped leaves" (association). Research in non-contextualised settings (e.g., on paper or screen) shows that personalised keyword-associations improve retention when learners create and visualise their own associations. Studies also suggest that externalising these associations through images enhances recall, while Augmented Reality (AR)

further strengthens retention by anchoring words to real-world objects. However, existing AR studies have only explored predefined keyword-associations with expert-designed visuals, leaving a gap in understanding how personalised and predefined approaches compare in contextualised AR learning. To explore this, we developed ARText, an AR system that visually annotates real-world objects with (i) their corresponding words in both English and the target language, (ii) keywords, and (iii) visual representations of associations, generated using text-to-image synthesis. Participants experienced keyword-associations in both PERSONALISED condition (keyword-associations created by users) and PREDEFINED condition (keyword-associations designed by experts). The findings indicate that participants preferred PREDEFINED keyword-associations. This condition also facilitated faster and more efficient word recall. In this paper, we discuss possible reasons for these outcomes and explore their implications for designing future AR-based vocabulary learning systems.

Keywords: Keyword Method, Text-to-image Synthesis, Augmented Reality, Contextualised Vocabulary Learning

1 Introduction

Expanding vocabulary is an important aspect of learning a foreign language, and several techniques exist to support learners in this process [1]. Each technique employs a distinct mechanism to aid retention. For example, (i) formal learning, such as reading and rehearsing words, relies on repetition only, (ii) flashcards, with a foreign word on one side and its translation or image on the other, rely on creating a mental link between what is on the front and back side of the card through card flipping, (iii) physically labelling objects in the environment, such as placing post-it notes on them, rely on the environmental context, (iv) and mnemonic strategies, such as the keyword method, enhance memorisation by linking new vocabulary to prior knowledge through elaborate encoding techniques using visual, spatial, semantic, or acoustic cues.

Several of these techniques rely on visual stimuli, either through external images/objects (e.g. flash cards, labelled objects), or mental visualisations. The latter is used in the keyword method, which requires learners to mentally visualise an association between a foreign word's meaning to a phonetically similar keyword [2, 3]. For example, a Japanese word for a "postcard" is "hagaki" (written in rōmaji), which sounds similar to "hug a key" (keyword), and the imagined scene between the two can be "a hand hugging a key on the postcard" (association). When recalling the word, reconstructing this mental image aids memory, which has been shown this method to be an effective vocabulary learning strategy [4].

Learning foreign words in real-world context, such as by labelling objects in one's surrounding, has also been shown to enhance vocabulary learning [1]. Augmented Reality (AR) supports such learning by overlaying digital elements onto physical objects, combining real-world scenes with augmented visuals [5, 6]. Recent studies has demonstrated the effects of integrating AR and the keyword method by augmenting real objects with expert generated PREDEFINED keyword-associations and corresponding

images of associations [7, 8]. However, PREDEFINED keyword-associations have limitations: (i) they are restricted to a predetermined set of objects, making it difficult to expand to new words, and (ii) they do not allow users to come up with their own personalised keyword-associations. Personalised keyword associations promote deeper cognitive processing by linking new information to prior knowledge, fostering meaningful connections [9], and improving long-term retention and recall [10, 11].

Externalising mental visualisations for personalised keyword-associations was very difficult or close to impossible, until the recent advent of text-to-image generators, such as OpenAI DALL-E 2 [1] [12] or Stability-AI Stable Diffusion[2] [13]. These tools can generate images with semantic coherence based on natural language inputs, such as "a hand hugging a key on the postcard". Recent studies have shown that externalising personalised associations with images enhances retention compared to traditional mental visualisations [14]. Building on this, we integrated a text-to-image generator into an AR system called ARText, that combines personalised keyword-associations with visualisations and contextualised learning of vocabulary—a novel approach implemented for the first time in our prototype, ARText as seen in Figure 1.

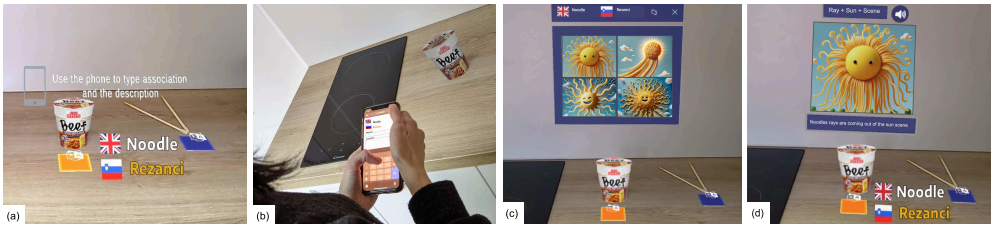


Fig. 1 ARText vocabulary learning system: (a) Initial AR annotations of the object with the English word (Noodle) and the Slovenian word (Rezanci); (b) The smartphone application for users to come up with and enter the personalised keyword and description of the association for a given word; (c) The selection menu presenting four different images generated using DALL-E 2 text-to-image synthesis with the description of the association as a prompt; and (d) AR view showing the same scene as in (a) together with the selected image, the keyword, and the description of the association.

ARText system offers two modes: (i) PREDEFINED keyword-associations and their visualisations designed by experts using phonetically similar English words, or (ii) PERSONALISED keyword-associations, where users create their own and generate images in real-time. The system then dynamically annotates real-world objects in AR with keyword-associations and their visualisations. Using our system we aimed to compare PREDEFINED and PERSONALISED keyword-associations in contextualised AR environments in terms of learning outcomes (e.g., retention and learning efficiency) and user preference, as well as what design considerations should be taken into account to optimise the use of keyword method in vocabulary learning AR systems?

Our results show that PREDEFINED keyword-associations outperform PERSONALISED ones in memory recall (immediately after the study and delayed after 7 days), task completion time, and learning efficiency (immediate and delayed). Additionally,

[1] https://openai.com/index/dall-e-2/
[2] https://stability.ai/

PREDEFINED condition required significantly less mental effort and was preferred by most participants, who found it helped them retrieve words more quickly and easily. In our discussion, we explore possible reasons for these findings and their implications.

2 Research Background

In this section we identify and analyse the relevant literature on keyword method, text-to-image synthesis, and context-based vocabulary learning in AR.

2.1 Mnemonics and the Keyword Method

Research on memory and learning suggests that comprehension and recall depend on instructional methods that influence how information is processed and stored [15]. In language learning, incidental vocabulary acquisition (e.g., through reading) is less effective without systematic vocabulary learning techniques that reinforce memorisation [16, 17]. Mnemonics are one such technique [18–20], enhancing recall by linking new information to prior knowledge using visual or acoustic cues. Various mnemonic techniques exist, including phonetic systems, keywords, rhyming words, or acronyms [18, 19, 21–25].

One of the mnemonic technique used in vocabulary learning is the so-called "keyword method", which requires learners to mentally visualise an association between a foreign word's meaning and a phonetically similar keyword [26]. Paivio et al. [27, 28] proposed that forming mental images aids learning. According to their Dual Coding Theory (DCT), mental visualisation and verbal cues [27, 29] are functionally independent and processed differently and along distinct channels in human mind when a person encodes information about a particular concept. In vocabulary learning, a concept like "tree" can be stored both as a word and an image, allowing retrieval from either or both channels, making the keyword and similar methods an effective learning strategy [1, 20].

The vocabulary learning methods based on mental imagery work best for words with high degree of "imageability" (e.g. moon vs. truth) [30], or for word pairs (the foreign word and the familiarly-sounding word) between which the learner can form some kind of semantic link [31]. Nevertheless, they provide a powerful tool for quick vocabulary acquisition of such words, given that memorable link clearly relates to the thing being remembered Besides mental imagery, actual pictures are often used in vocabulary learning [32] as depicted by a plethora of picture dictionaries combining words and images (e.g. [33]). Automatically creating images that semantically match a user-provided text description from the mental imagery is a challenging problem, which has proved elusive until the advent of Deep Learning, and more concretely until the advent of deep generative models. In the last five years, different deep generative models have been devised to synthesise images from text, including Generative Adversarial Networks (GANs) [34], Variational Auto-Encoders (VAEs) [35], flow-based models [36], autoregressive models [37], and more recently diffusion Models [13]. They are nowadays considered state-of-the-art approaches to text-to-image generation.

A recent study explored externalising mental visualisations in the keyword method with text-to-image generators in a non-contextualised settings on a desktop

computer [14]. It investigated challenges in externalising personalised keyword-associations, compared text-to-image models for image quality, and assessed their impact on vocabulary retention. The findings showed that externalised personalised keyword-associations enhance recall and reduce mental effort, with DALL-E 2 producing the highest-quality images. However, the researchers did not compare predefined and personalised keyword-associations in terms of learning outcomes, user preference, and efficiency—an important gap we aim to address in this paper for designing future vocabulary learning systems in contextualised settings of AR.

2.2 Vocabulary Learning in AR

It has been known that learners are more motivated when they see the content as relevant to their situation or find it engaging [38]. For example, being in a café while travelling abroad may naturally encourage the desire to learn phrases for ordering coffee. Such contextual cues help create associations that improve recall in similar settings [5, 6, 39, 40]. This means that words related to the learning environment are more likely to be remembered than unrelated ones [41]. Augmented Reality (AR) can enhance learning by providing interactive, context-specific information and extensive research has demonstrated AR's potential to support real-world learning, including vocabulary acquisition. In addition, studies have shown that AR technology for language learning can contribute to increased motivation and enjoyment [42, 43]. Some AR systems display labels for objects with corresponding foreign words [5, 44], while others create imaginary settings to describe and enhance the characteristics of physical items [45, 46].

Several studies have explored AR applications for vocabulary learning using handheld devices. VocaBura [47], a smartphone application for learning vocabulary, presents vocabulary related to the current location via audio. A study found that participants using AR recalled significantly more words after seven days than those using an audio-only method. Another handheld AR system displayed text, images, animation and sound next to real-world objects [5]. While initial recall favoured a non-AR flashcard method, long-term retention showed no difference.

Early AR-based vocabulary learning on AR head-mounted devices (HMDs) used markers to link virtual content to physical objects. For example, the Wordsense platform [40] enhanced real-world objects with language-learning content such as words, sentences, and multimedia, but no formal user study evaluated its effectiveness. Another system, ARbis Pictus [6], labelled objects with target-language vocabulary and was compared to a flashcard-based method. Results showed AR improved recall and enjoyment both immediately and after four days. However, differences in presentation–such as AR displaying words alongside objects while flashcards separated text and images—-make it unclear whether AR itself or the learning approach led to better outcomes.

Recent research highlights the benefits of integrating the keyword method with context-based vocabulary learning in AR [7, 8, 48]. VocabulARy system [7] has demonstrated that AR condition resulted in better recall, efficiency, motivation, engagement, and user experience compared to tablet-based learning. However, the system used

PREDEFINED keyword-associations with visualisations, rather than allowing the users to come up with their own PERSONALISED keywords and associations.

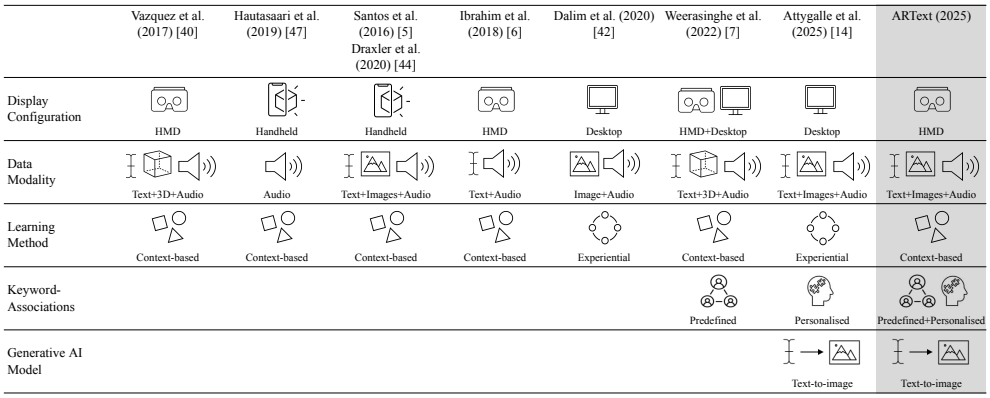


Fig. 2 Situating our work (the rightmost system ARText) within vocabulary learning of either AR or keyword method systems based on: Display Configuration (handheld devices, desktop setups, HMDs), Data Modality (text, images, audio, 3D elements), Learning method (context-based, experiential learning), Keyword-Associations (predefined, personalised), Generative AI type.

To compare the aforementioned existing approaches we provide a classification of the design space of either AR-based vocabulary learning systems or the keyword method systems in Figure 2. Our work on the right side expands the existing approaches by comparing predefined with personalised keyword-associations within AR. Thus, our work aims to investigate:

(i) How do PREDEFINED and PERSONALISED keyword-associations conditions compare in terms of learning outcomes (e.g., retention and learning efficiency, completion time, mental effort) in contextualised AR environments?
(ii) Which method—PREDEFINED or PERSONALISED—do learners prefer, find more efficient, and easier to use in AR?
(iii) What design factors should be considered to enhance the effectiveness of the keyword method in AR-based vocabulary learning systems?

To answer these questions, we developed ARText—an AR system that provides keywords and visual annotations of associations, generated using text-to-image synthesis. These keyword-associations are either PERSONALISED, created by users based on their prior knowledge and native/familiar language, or PREDEFINED, designed by experts using phonetically similar English words as a reference.

3 Visualisation of Keyword-Associations in AR with Text-to-Image Synthesis

In this section, we describe the ARText prototype, study conditions, study design, study procedure, participants' sampling, data collection and analysis.

3.1 The ARText Prototype

The prototype was developed as an AR head mounted display (HMD) application for Microsoft HoloLens 2[3] using Unity3D game development environment[4] and the MRTK Mixed Reality tool kit[5]. For text-to-image generator we selected DALL-E 2, as it has been shown adequate for this type of task [14] and it was the latest one released and publicly available at the time of the study.

The system allows the user to observe their physical environment where certain objects are labelled with an AR button indicating that their translation is available. Upon clicking on a button, the English and Slovenian words (the target language) appear (Figure 1(a)). In addition, an audio pronunciation of the foreign word is played.

In the PERSONALISED keyword-associations condition, a smartphone (i.e. Google Pixel 3) was used as the text input device. Upon clicking on the AR button a notification is sent to the smartphone application asking users to type a keyword and the association (see Figure 1(b)). Users then generate images by clicking on the "Generate Image" button on the smartphone application. The association is used as a prompt for generating images. Once the images are generated using the text-to-image generator, a menu containing four (4) images pops up on the AR HMD next to the corresponding object (see Figure 1(c)). Users need to select the image that best represents their mental visualisation, and once selected, the image appears in AR alongside the object, accompanied by the English and Slovenian words, the keyword, and the association (see Figure 1(d). Participants kept the HMD visor down while interacting with the smartphone, gazing downward to align the screen with their direct line of sight. An informal pilot study assessed this interaction method, revealing no usability concerns for brief tasks like typing short phrases.

In the PREDEFINED keyword-associations version of the system, the same information (i.e., the keyword, association, and related visualisations) is displayed. However, these are pre-generated, and the user does not go through the process of creating or selecting them.

The HoloLens' built-in tracking system was used for camera pose tracking. To initialise augmentation positions, we used Vuforia [49] along with custom image markers, which were removed after initialisation. These markers ensured reliable and accurate detection of physical objects linked to the vocabulary set. While object recognition techniques [50–52] could enable identification and localisation without prior setup, allowing broader system implementation, this was not the current aim of the study.

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

Fig. 3 A user wearing the Hololense HMD with ARText prototype and holding the mobile device for text input.

Careful attention was given to the selection of annotation and image size. Research suggests that image size influences memory retention during naturalistic exploration [53]. However, in such studies, participants freely examine images without specific instructions before recalling details. This approach does not ensure equal distribution of visual attention across the image, and as images shrink, key details become smaller and easier to overlook.

To prevent participants from missing key information, the application displays AR information only for one object at a time, preventing scene clutter and information overload. Furthermore, we strategically placed AR buttons on the physical surface at close proximity to objects. This guided users to the appropriate physical location from which AR visualisations are clearly visible as the corresponding annotation and image sizes were appropriated for such viewing.

3.2 Study Design

We conducted a within-subjects study in two conditions across two vocabulary learning scenarios. The first scenario involved ten kitchen-related objects on the kitchen counter, while the second involved ten office-related objects on the office desk. Each scenario was paired with either the PREDEFINED or PERSONALISED condition. In the PREDEFINED condition, participants were provided with keyword-associations created by an expert (e.g., a teacher) along with a corresponding visualisation. In the PERSONALISED condition, participants had to create their own keyword-associations, generate the visualisation, and select an image from those generated by a text-to-image generator. The order of the conditions as well as the order of the learning scenarios (the kitchen and the office environments) were counterbalanced.

3.3 Participants

The study was completed by 26 university graduate students (Masters', Doctoral and Post doctoral) who volunteered and consented to the study. None had prior knowledge

of Slovenian language. This was confirmed through a short competency test where they hat to identify the meanings of 10 Slovenian nouns after hearing their pronunciations and select the correct one from three options. The selection criterion was a lack of familiarity with any of the words, which all participants met. All the participants had advanced proficiency in English (TOEIC $>$ 785, working proficiency plus able to communicate effectively in any situation[6], $\overline{x} = 897$ and $SD = 68.4$). The sample included ten (10) female participants (38.5%). Participants were aged between 23 to 38 (mean of $\overline{x} = 27.4$ and $SD = 4.1$) and were randomly assigned to one of the two conditions.

3.4 Procedure

Participants were randomly assigned to start with either the PERSONALISED or PREDEFINED condition, followed by a randomly chosen learning scenario (kitchen or office environment), with counterbalancing applied After signing a consent form and receiving a study briefing, participants could ask questions and were informed of their right to withdraw at any time.

Before the main task, they completed a five-minute training session to familiarise themselves with the system. They then proceeded with the first learning scenario (10 words), followed by NASA-TLX (mental effort), immediate recall (remembering 10 words), system usability questionnaire (SUS) [54] and a user experience questionnaire (UEQ) [55]. After a five minute break, they completed the second scenario (10 words) and repeated the same evaluations. To avoid the fatigue of wearing the HMD we kept the study under 30 minutes long.

After the experiment, participants completed a questionnaire assessing their views on retrieval efficiency, ease, and method preference, along with questions about demographics, AR experience, and their vision. This was followed by a short interview that explored usability challenges of interacting with a smartphone while wearing an HMD, as well as gathered feedback about the system and/or experiment. The entire study lasted 45–60 minutes.

After one week, participants were again asked to answer the same recall questionnaires to assess their delayed recall performance as in prior work [7, 47].

3.5 Data Collection and Analysis

To measure the task completion time (the duration participants spent learning the words), the time stamp data (start time and end time of the learning task) were logged by the system. The NASA Task Load Index (NASA-TLX) [56, 57] was used to measure participants' perceived *mental effort*.

In the recall questionnaires, participants were asked to recall the words they had learnt, both immediately after using the prototype (*immediate recall*) and one week later (*delayed recall*). *Learning efficiency* was calculated as the ratio of performance (*immediate* or *delayed recall*) to task difficulty (*mental effort*), as proposed in [58]. Performance and task difficulty data were then standardised using $z = (r - M)/\sigma$, where z = Z-score, r = Raw data score, M = Population mean, and σ = Standard

[6] https://toeic-testpro.com/blog/

deviation. Next, the *learning efficiency* was assessed using $E = (z_P - z_M)/\sqrt{2}$, where E = Learning efficiency, z_P = Average performance in Z-scores, and z_M = Average task difficulty in Z-scores [58–60].

System usability was assessed using the System Usability Scale (SUS) [61]. User experience was measured with the short version of the User Experience Questionnaire (UEQ-S) [55, 62] with eight items/questions, reported on a 7-point Likert scale. The first four represent pragmatic qualities (Perspicuity, Efficiency and Dependability) and the last four hedonic qualities (Stimulation and Novelty) [62].

The Shapiro–Wilk test [63] was used to assess normality, confirming that all data sets were normally distributed. Statistical analyses were conducted with a significance level o $p - value > 0.05$ and a 95% confidence interval (CI). A Paired Samples t-test [64] was used for immediate recall, delayed recall, mental effort, task completion time, and learning efficiency, while the Wilcoxon signed-rank test [65] assessed system usability. Statistical significance is indicated in tables using asterisks (ns: $p > 0.05$, *: $p < 0.05$, **: $p < 0.01$, and ***: $p < 0.001$).

To asses the reliability of recall questionnaires, we performed a Kuder-Richardson 20 test [66]. The $KR = 0.87 > 0.5$ value indicates that the reliability is acceptable. We used a power analysis to validate the study's results. Effect sizes (Cohen's d) [67] were calculated, with the selected minimum effect size ($d = 0.73$) and estimated statistical power ($1-\beta$), to check whether the type II error probability (β) is within an acceptable range. Given a sample size of $n = 26$ per group and a significance level of $\alpha = 0.05$, the estimated power of 0.94 confirms a $> 90\%$ probability of correctly rejecting the null hypothesis.

4 Results

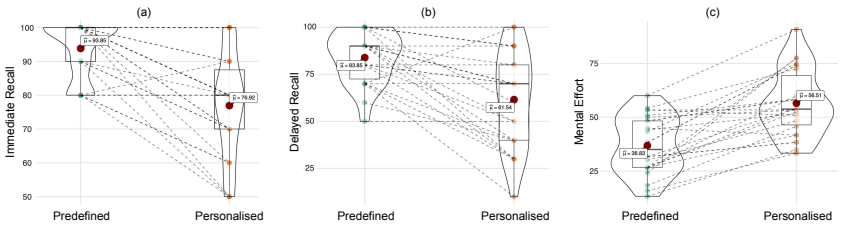
The following subsections describe the effect of conditions (PREDEFINED and PERSONALISED) on learning performance (immediate and delayed recall), mental effort, task completion time, and learning efficiency (immediate and delayed).

4.1 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 condition on *immediate recall* is shown in Figure 4(a) (top left). The results of Paired Samples t test show that the effect of condition on *immediate recall* is statistically significant ($df = 25.0$, $p < .001$). The violin graph indicates that the *immediate recall* in the PREDEFINED condition ($\overline{x} = 93.85.0\%$, $SD = 8.52$) is significantly better compared to the PERSONALISED condition ($\overline{x} = 76.92\%$, $SD = 15.43$).



		Immediate Recall			Delayed Recall			Mental Effort		
		Statistics	df	p	Statistics	df	p	Statistics	df	p
Predefined	**Personlaised**	5.91	25	< .001	4.83	25	< .001	-6.59	25	< .001

Fig. 4 The results for (a) Immediate recall in percentage of correctly remembered words, (b) Delayed recall in percentage of correctly remembered words, (c) Overall mental effort invested during the task. The tables include results of Paired Samples t tests over dependent variables.

4.1.2 Delayed recall

The effect of condition on *delayed recall* is shown in Figure 4(b). The effect of condition on *delayed recall* is also statistically significant ($df = 25.0$, $p < .001$), where the *delayed recall* in the PREDEFINED condition ($\overline{x} = 83.85\%$, $SD = 13.59$) is significantly better compared to the PERSONALISED condition ($\overline{x} = 61.54\%$, $SD = 25.09$).

4.2 Mental Effort

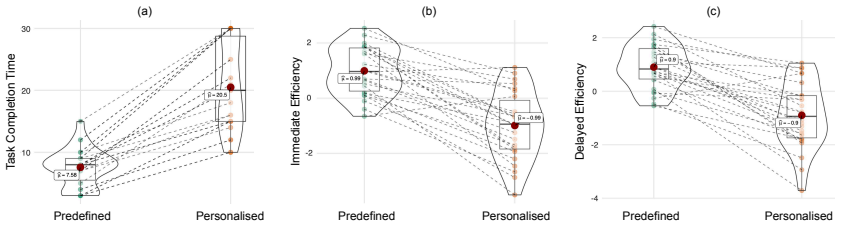
The effect of condition on *mental effort* is shown in Figure 4(c). The results indicate that the effect of condition on *mental effort* is statistically significant ($df = 25.0$, $p < .001$). Participants' *mental effort* in the PREDEFINED condition ($\overline{x} = 36.83$, $SD = 13.15$) is significantly lower compared to the PERSONALISED condition ($\overline{x} = 56.51$, $SD = 15.08$).

4.3 Task Completion Time

The Paired Samples t test results for the *task completion time* presented in Figure 5(a) indicate that the condition has a statistically significant effect on the *task completion time* ($df = 25.0$, $p < .001$), where the time in the PREDEFINED condition ($\overline{x} = 7.69$ min, $SD = 2.74$) is significantly lower compared to the PERSONALISED condition ($\overline{x} = 20.50$ min, $SD = 7.08$).

4.4 Learning Efficiency

The effect of condition on *learning efficiency* is shown in Figure 5(b) (immediate) and Figure 5(c) (delayed). The results indicate a statistically significant effect on both immediate ($df = 25.0$, $p < .001$) and delayed ($df = 25.0$, $p < .001$) *efficiency*. The *learning efficiency immediate recall* of the PREDEFINED condition ($\overline{x} = 0.99$ s, $SD = 0.87$) is significantly higher compared to the PERSONALISED condition ($\overline{x} = -0.99$ s, $SD = 1.23$). Also, the *learning efficiency* for *delayed recall* of the PREDEFINED condition ($\overline{x} = 0.90$ s, $SD = 0.81$) is significantly higher compared to the PERSONALISED condition ($\overline{x} = -0.90$ s, $SD = 1.23$).



		Task Completion Time			Immediate Efficiency			Delayed Efficiency		
		Statistics	df	p	Statistics	df	p	Statistics	df	p
Predefined	**Personlaised**	-9.17	25	< .001	8.31	25	< .001	7.81	25	< .001

Fig. 5 The results for (a) *task completion time* in minutes, (b) *learning efficiency* for immediate recall, (c) *learning efficiency* for delayed recall. The tables include results of Paired Samples t tests over dependent variables.

4.5 System Usability and User Experience

We assessed *system usability* using the SUS questionnaire [61]. In the PREDEFINED condition, participants scored 90 on the SUS, whereas in the PERSONALISED condition, they scored 85. The Wilcoxon signed-rank test showed no significant difference between the PREDEFINED and PERSONALISED condition ($df = 25.0$, $p > .05$).

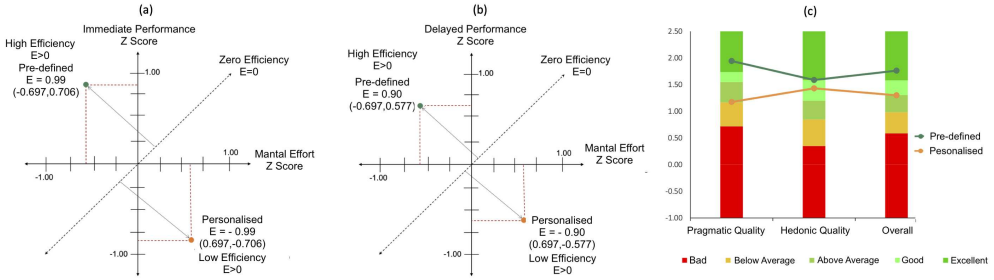


Fig. 6 The results for (a) Immediate recall learning efficiency, (b) Delayed recall learning efficiency, (c) UEQ factors (pragmatic and hedonic) and all items/questions together (overall) with benchmarks for each factor.

To compare user experience between the two conditions, we used the UEQ-s questionnaire [68]. Following the standard calculation method [55, 62], we determined values for *pragmatic quality*, *hedonic quality* and *overall* user experience for both conditions.

As seen in Figure 6(a), in the PREDEFINED condition, participants rated *pragmatic quality* ($\overline{x} = 1.94$), *hedonic quality* ($\overline{x} = 1.59$), and *overall* user experience ($\overline{x} = 1.77$) as excellent (benchmarks: > 1.74 for pragmatic, > 1.59 for hedonic and > 1.58 for overall). In the PERSONALISED condition, participants rated *hedonic quality* ($\overline{x} = 1.43$) as good (benchmark: 1.25 - 1.55), and the *overall* user experience ($\overline{x} = 1.30$) as above

average (benchmark: 1.02 - 1.39). However, *pragmatic quality* ($\overline{x} = 1.10$) fell below average (benchmark: 0.73 - 1.14).

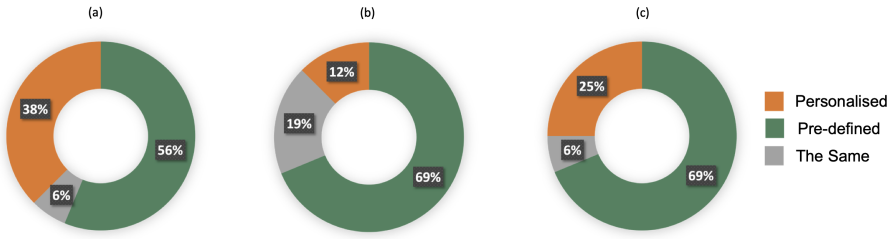


Fig. 7 The participants' preference results for: (a) Method preference, (b) Retrieval efficiency and, (c) Retrieval ease.

4.6 Retrieval Efficiency, Ease and Method Preference

The results from the *method preference*, *retrieval efficiency* and *retrieval ease* are shown in Figure 7. PREDEFINED condition was preferred by 56% participants (Figure 7(a)) and 69% thought that it helped them retrieve words faster (Figure 7(b)) and easier (Figure 7(c)).

5 Discussion and Future Directions

Learning a foreign language involves a range of educational methods, with preferences differing among learners and evolving over time. The "keyword method" is just one such technique specifically used for vocabulary learning. Since building a strong vocabulary is important for improving speaking skills, integrating this method with novel technologies can enhance the learning process.

To explore this, we developed ARText, a novel system that combines the keyword method with text-to-image synthesis and augmented reality (AR) context-based learning. Using our system, we investigated whether learning outcomes differ when using PREDEFINED versus PERSONALISED keyword-associations. This is important not only because personalisation can potentially foster the learning experiences but also because it enables the system to adapt to any learning context, removing the limitation on relying solely on expert-prepared predefined word sets. However, the PERSONALISED version in our experiment did not yield better results, which we investigate further hereafter and explore implications for AR system design.

5.1 System Usability and User Experience

Participants rated the ARText system with good SUS scores in both PREDEFINED (score 90) and PERSONALISED (score 85) condition, clearly above the average (68), with no significant difference between them. Despite the fact that in PERSONALISED version we introduced cross-reality interaction [69, 70] between the smartphone and AR content.

User experience was rated high for hedonic qualities (stimulation and novelty) in both conditions. However, the PERSONALISED condition scored lower in pragmatic qualities (perspicuity, efficiency, and dependability), likely due to (i) some users struggling to generate their own keywords and associations, and (ii) taking three times longer ($\overline{x} = 21.5$ min) to complete the task compared to the PREDEFINED condition ($\overline{x} = 7.8$ min). While prolonged HMD use can affect usability [71–75], SUS scores did not indicate this. Also, the 20-minute duration is relatively short compared to prior studies where users wore HMDs for up to eight hours [75].

It is also important to note that our results were obtained in a controlled environment of a laboratory designed to minimise hardware and software factors that could impact usability and user experience. For example, lighting conditions were controlled, objects were arranged in a way to avoid occlusion, and the experimenter was present to assist users in case of system failures. All this had a positive effect on the usability and user experience. This controlled setup was deliberate and essential to strengthen the validity of the preference and learning results, as the primary objective of this work is to compare the effectiveness of PERSONALISED and PREDEFINED conditions of the keyword method in real-world context of vocabulary learning. This is important because the keyword method can be applied beyond vocabulary learning (e.g., face-name recall, anatomy learning) and this study opens up the new real-world use cases for AR systems–now feasible with recent advances in text-to-image generation.

subsectionLearning Outcomes

Our results show that PREDEFINED support, which includes predefined keywords, associations and visualisations, lead to better learning outcomes compared to PERSONALISED support, where participants had to came up with their own keywords, associations, an select visualisations. This trend was consistent across all metrics, including recall, learning efficiency, mental effort, and task completion time.

Some of these results were anticipated, as coming up with one's own keyword-associations requires greater mental effort, which can negatively impact learning efficiency. Nevertheless, we still expected higher recall in the PERSONALISED condition, given that research has demonstrated personalised mental visualisation improves memory performance, strengthens memory retention, recall, and reduces false memories [76]. For example, it has been shown that mental visualisation of words on a list we are memorising reduces false memories at the time of recall [77], and that forming mental visualisation while reading increases the amount of content remembered over time [78]. However, these tasks required simple visualisations of words.

In contrast, coming up with meaningful keyword-associations can be challenging for new users. Research suggests that people might have difficulties coming up with their own keywords, so that providing predefined keywords may reduce the required effort and lead to better learning outcomes [26]. However, lowering mental effort in some learning scenarios can sometimes result in poorer learning outcomes [79, 80]. In our study, this effect was not observed, likely because participants still had to engage with the task to retain words. Even understanding the PREDEFINED keywords may have required sufficient effort to support memorisation.

Other factors that may have influenced our results include cultural differences in learning styles. Previous research has demonstrated that the effectiveness of the

personalised keyword method varies depending on the language being learned [14]. In a study comparing Japanese and Slovenian, two culturally distinct groups were examined, with Japanese participants reporting higher mental effort when using the personalised keyword method compared to their European counterparts. While differences in learning styles have been observed between Japanese and other cultures [81–83], other factors such as personality and educational major can also shape one's learning style [84]. Users who favour structured learning, guided instruction, and memorisation techniques, might prefer PREDEFINED approach. While users who favour a high degree of independent, creative thinking might prefer PERSONALISED method.

In addition, the study duration was limited to 30 minutes per participant to prevent prolonged use of the HMD, ensuring usability and comfort were not compromised. While this decision helped isolate the impact of learning strategies without the confounding factor of HMD fatigue, it may have restricted the potential benefits of the PERSONALISED approach. Given more time, participants might have been able to create stronger, more memorable associations, potentially improving recall. However, under time constraints, the PREDEFINED condition provided a more efficient and structured learning experience, leading to better overall outcomes. However, additional studies are required to confirm the confounding factors.

5.2 Implications for Design

It has already been shown that AR enhances vocabulary learning by contextualising it within real-world environment, leading to better retention and engagement compared to traditional flat-screen displays [7]. Like flat-screen displays, AR can facilitate dual coding, where learners process information both visually (objects) and verbally (words and their pronunciation, keywords, associations), which is known to enhance memory [28, 85]. However, unlike flat-screen displays, AR allows users to physically interact with their surroundings, making keyword-associations learning more immersive.

While these were the reasons to select AR as learning platform, our study focused on designing pedagogical content and researching its use within AR. The results show that predefined content yielded better results compared to personalised content. However, both PREDEFINED and PERSONALISED approaches have strengths in enhancing memory and recall. A key question in instructional design is determining the optimal level of support provided in a learning environment [86–89]. This issue, known as the "assistance dilemma" [90], explores whether it is more effective to directly present learners with information or to encourage them to construct knowledge independently.

Research suggests that the ideal level of instructional support often lies somewhere in between full guidance and complete autonomy. Wise and O'Neill argue that an intermediate level of support is typically the most effective, with the granularity of assistance—how detailed and specific the support is—playing a crucial role in learning outcomes [91]. Instead of assuming that more guidance always leads to better learning, it is essential to strike a balance where learners receive just enough assistance to enhance comprehension and engagement without diminishing their cognitive involvement [92].

One way to achieve this balance is by progressive disclosure. This is either providing multiple PREDEFINED (i) keyword-associations, allowing users to select and adapt

them, or (ii) keywords only, allowing users to generate associations and visualisations, or (iii) allowing users to come up with their own keyword-associations. The system could start with predefined content and gradually transition toward personalised keyword generation. Such adaptation to different learning styles would cater to learners that benefit from structured, step-by-step guidance, as well as support quick progress for learners that prefer a more exploratory, self-driven approach.

5.3 Limitations and Future Directions

Our participants were postgraduate students (see subsection 3.3) with TOEIC scores above 785, 62% of whom had international professional proficiency (scores between 905-990), while the rest had working proficiency (scores between 785-900). This age group was chosen due to their youth, international travel experience, and exposure to multiple cultures. However, as all participants were non-native English speakers, language could have caused an interaction effect, as they used English prompts for text-to-image generation. Despite their high English proficiency, future studies should control for this by either providing a text-to-image generation system in participants' native language or conducting the experiment with English native speakers.

Further, the keyword method works best for words of high imageability. A number of studies have shown that concrete terms are better remembered than abstract terms [93]. The benefits of AR in context learning will therefore be reduced when abstract terms are considered as it becomes more difficult to make them relevant to the context of users' immediate environment. However, as explained earlier, the method is only one of many and worth exploring for quick vocabulary acquisition.

Future studies could explore the potential of other types of visuals since static 2D content might not be ideal for AR. This includes dynamic visuals that can be generated by text-to-video such as Sora[7] or even more contextually relevant 3D objects that can be generated by text-to-3D DreamFusion[8] and Magic3D[9].

Our prototype was tested for a short time on a limited vocabulary. To further validate findings, the study should be extended and possibly integrate spaced repetition and recall technique. This is particularly important as USED-DEFINED keywords could be affected by a shallow learning curve reaching peak performance only after extensive practice. At this point the outcome might change making PERSONALISED keyword method outperform PREDEFINED keywords. This could be explored with an adaptive system mentioned in previous subsection.

6 Conclusion

The keyword method is a mnemonic techniques for learning vocabulary in a foreign language, where a learner uses the pronunciation of a foreign word, finds a similarly sounding word (or a combination of words) in one's native language (known as keyword), and makes a memorable visual connection between the two (know as association). Research has shown that externalising these personalised connections with

[7] https://openai.com/sora/
[8] https://dreamfusion3d.github.io/
[9] https://research.nvidia.com/labs/dir/magic3d/

text-to-image generated visual cues and encoding them during memorisation process can aid the retention and recall [14]. However, PERSONALISED externalisation has not yet been explored in AR, a technology known to enhance keyword method vocabulary learning of PRE-DEFINED keyword-associations by the experts [48].

To explore the differences between PERSONALISED and PRE-DEFINED keyword-associations in AR, we developed ARText. This is an innovative cross-reality AR system that can generate and visualise these two types of keywords-associations using text-to-image generator and a smartphone for quick text input.

First, our findings indicate that ARText provides a smooth user experience with no significant usability issues, even as interaction shifts between the smartphone and AR content. Second, the results show that PREDEFINED keyword-visualisations outperform PERSONALISED ones in terms of *immediate recall*, *delayed recall*, *mental effort*, *task completion time*, and *learning efficiency*. Furthermore, most participants favoured the PREDEFINED approach as it helped them retrieve foreign words faster and easier.

This might suggest that PREDEFINED keyword-visualisations provide a more efficient learning experience compared to PERSONALISED. However, both have their advantages in improving memory and recall, and more than third of our participants still preferred the PERSONALISED method. The results might also have been influenced by cultural differences in learning styles. This calls for additional studies and solutions using SEMI-PERSONALISED approach, where users are presented with *multiple predefined* keyword-associations and can select, adapt or ignore them to support their learning preferences.

Declarations

- Funding: This research was funded by the Slovenian Research Agency, grant number P5-0433, IO-0035, N2-0354, J5-50155, J7-50096 and BI-JP/20-22-001. This work has also been supported by the research program CogniCom (0013103) at the University of Primorska.

- Conflict of interest/Competing interests: The authors declare that they have no conflict of interest.
- Ethics approval and consent to participate: The study was approved by the Commission of the University of Primorska for Ethics in Human Subjects and all the participants in the study signed an informed declaration of consent.
- Consent for publication: The authors declare that they consent to publication.
- Author contribution: Conceptualisation: [Maheshya Weerasinghe]; Methodology: [Maheshya Weerasinghe, Verena Biener], Software [Maheshya Weerasinghe, Nuwan T. Attygalle]; Formal analysis and investigation: [Maheshya Weerasinghe, Juri Yoneyama], Writing - original draft preparation: [Matjaž Kljun, Maheshya Weerasinghe]; Writing - review and editing: [All authors]; Validation [Klen Čopič Pucihar]; Funding acquisition: [Hirokazu Kato, Matjaž Kljun]; Resources: [Hirokazu Kato, Jens Grubert, Klen Čopič Pucihar]; Supervision: [Matjaž Kljun, Aaron Quigley, Klen Čopič Pucihar].

References

[1] Pressley, M., Levin, J.R., Kuiper, N.A., Bryant, S.L., Michener, S.: Mnemonic versus nonmnemonic vocabulary-learning strategies: Additional comparisons. Journal of Educational Psychology **74**(5), 693–707 (1982) https://doi.org/10.1037/0022-0663.74.5.693

[2] Raugh, M.R., Atkinson, R.C.: A mnemonic method for learning a second-language vocabulary. Journal of Educational Psychology **67**(1), 1–16 (1975) https://doi.org/10.1037/h0078665

[3] Anonthanasap, O., He, C., Takashima, K., Leelanupab, T., Kitamura, Y.: Mnemonic-based interactive interface for second-language vocabulary learning. In: Proceedings of the Human Interface Society (HIS), pp. 14–19 (2014)

[4] Atkinson, R.C.: Mnemotechnics in second-language learning. American Psychologist **30**(8), 821–828 (1975) https://doi.org/10.1037/h0077029

[5] Santos, M.E.C., Lübke, A.i.W., Taketomi, T., Yamamoto, G., Rodrigo, M.M.T., Sandor, C., Kato, H.: Augmented reality as multimedia: the case for situated vocabulary learning. Research and Practice in Technology Enhanced Learning **11**(1), 1–23 (2016) https://doi.org/10.1186/s41039-016-0028-2

[6] Ibrahim, A., Huynh, B., Downey, J., Höllerer, T., Chun, D., O'Donovan, J.: Arbis pictus: A study of vocabulary learning with augmented reality. IEEE Transactions on Visualization and Computer Graphics **24**(11), 2867–2874 (2018) https://doi.org/10.1109/TVCG.2018.2868568

[7] Weerasinghe, M., Biener, V., Grubert, J., Quigley, A., Toniolo, A., Čopič Pucihar, K., Kljun, M.: Vocabulary: Learning vocabulary in ar supported by keyword visualisations. IEEE Transactions on Visualization and Computer Graphics **28**(11), 3748–3758 (2022) https://doi.org/10.1109/TVCG.2022.3203116

[8] Weerasinghe, M., Quigley, A., Čopič Pucihar, K., Toniolo, A., Miguel, A., Kljun, M.: Arigatō: Effects of adaptive guidance on engagement and performance in augmented reality learning environments. IEEE Transactions on Visualization and Computer Graphics **28**(11), 3737–3747 (2022) https://doi.org/10.1109/TVCG.2022.3203088

[9] Dinsmore, D.L., Alexander, P.A.: A critical discussion of deep and surface processing: What it means, how it is measured, the role of context, and model specification. Educational psychology review **24**, 499–567 (2012) https://doi.org/10.1007/s10648-012-9198-7

[10] Craik, F.I., Tulving, E.: Depth of processing and the retention of words in episodic memory. Journal of experimental Psychology: general **104**(3), 268 (1975) https://doi.org/10.1037/0096-3445.104.3.268

[11] Altmeyer, M., Wang, T., Schweizer, K.: On the relationship between the retrieval of information and learning: the influence of deep processing. Psychological Test and Assessment Modeling **59**(3), 343 (2017)

[12] Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., Chen, M.: Hierarchical text-conditional image generation with CLIP latents. arXiv preprint arXiv:2204.06125 (2022) https://doi.org/10.48550/arXiv.2204.06125

[13] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-Resolution Image Synthesis with Latent Diffusion Models . In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10674–10685 (2022). https://doi.org/10.1109/CVPR52688.2022.01042

[14] Attygalle, N.T., Kljun, M., Quigley, A., Pucihar, K., Grubert, J., Biener, V., Leiva, L.A., Yoneyama, J., Toniolo, A., Miguel, A., Kato, H., Weerasinghe, M.: Text-to-image generation for vocabulary learning using the keyword method. In: Proceedings of the 30th International Conference on Intelligent User Interfaces. IUI '25, pp. 1381–1397 (2025). https://doi.org/10.1145/3708359.3712073

[15] Dunlosky, J., Rawson, K.A., Marsh, E.J., Nathan, M.J., Willingham, D.T.: 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) https://doi.org/10.1177/1529100612453266

[16] Koren, S.: Vocabulary instruction through hypertext: Are there advantages over conventional methods of teaching. TESL-EJ **4**(1), 1–18 (1999)

[17] Paribakht, T.S., Wesche, M.: Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. Second language vocabulary acquisition: A rationale for pedagogy **55**(4), 174–200 (1997) https://doi.org/10.1017/CBO9781139524643.013

[18] Putnam, A.L.: Mnemonics in education: Current research and applications. Translational Issues in Psychological Science **1**(2), 130–139 (2015) https://doi.org/10.1037/tps0000023

[19] Mastropieri, M.A., Scruggs, T.E.: Teaching Students Ways to Remember: Strategies for Learning Mnemonically. Brookline Books, Cambridge, MA (1991)

[20] Pressley, M., Levin, J.R., Delaney, H.D.: The mnemonic keyword method. Review of Educational Research **52**(1), 61–91 (1982) https://doi.org/10.3102/00346543052001061

[21] Luria, A.R.: The Mind of a Mnemonist: A Little Book About a Vast Memory. Harvard University Press, Cambridge, MA (1987)

[22] Yates, F.A.: The Art of Memory. Pimlico, London (1992)

[23] Higbee, K.L.: Your Memory: How It Works and How to Improve It. Da Capo Lifelong Books, Boston, MA (2001)

[24] Gardner, M.: Mathematical Puzzles & Diversions. G. Bell and Sons, London (1961)

[25] Herrmann, D.J., Raybeck, D., Gruneberg, M.M.: Improving Memory and Study Skills: Advances in Theory and Practice. Hogrefe & Huber Publishers, Seattle, WA (2002)

[26] Atkinson, R.C.: Mnemotechnics in second-language learning. American Psychologist **30**(8), 821–828 (1975) https://doi.org/10.1037/h0077029

[27] Paivio, A., Desrochers, A.: Mnemonic techniques in second-language learning. Journal of Educational Psychology **73**(6), 780–795 (1981) https://doi.org/10.1037/0022-0663.73.6.780

[28] Paivio, A.: Imagery and Verbal Processes. Holt, Rinehart and Winston, New York (1971)

[29] Wilson, M.: Mrc psycholinguistic database: Machine-usable dictionary, version 2.00. Behavior Research Methods, Instruments, & Computers **20**, 6–10 (1988) https://doi.org/10.3758/BF03202594

[30] Richardson, J.T.E., Jones, A.: Mental Imagery and Human Memory. Macmillan, London (1980)

[31] Ellis, N.C.: The psychology of foreign language vocabulary acquisition: Implications for call. Computer Assisted Language Learning **8**(2-3), 103–128 (1995) https://doi.org/10.1080/0958822950080202

[32] Dolati, R., Richards, C.: Harnessing the use of visual learning aids in the english language classroom. Arab World English Journal **2**(1), 3–17 (2011)

[33] Adelson-Goldstein, J., Shapiro, N.: Oxford Picture Dictionary. Oxford University Press, Oxford (2016)

[34] Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., Lee, H.: Generative adversarial text to image synthesis. In: Proceedings of The 33rd International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 48, pp. 1060–1069 (2016)

[35] Huang, H., Li, Z., He, R., Sun, Z., Tan, T.: Introvae: Introspective variational autoencoders for photographic image synthesis. In: Advances in Neural Information Processing Systems 31 (NeurIPS 2018), pp. 52–63 (2018)

[36] Yang, K., Goldman, S., Jin, W., Lu, A.X., Barzilay, R., Jaakkola, T., Uhler,

C.: Mol2Image: Improved conditional flow models for molecule to image synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6688–6698 (2021). https://doi.org/10.1109/CVPR46437.2021.00662

[37] Essera, P., Rombach, R., Blattmann, A., Ommer, B.: ImageBART: Bidirectional context with multinomial diffusion for autoregressive image synthesis. In: Proc. NeurIPS (2021). https://doi.org/10.48550/arXiv.2108.08827

[38] Pintrich, P.R.: A motivational science perspective on the role of student motivation in learning and teaching contexts. Journal of Educational Psychology **95**(4), 667–686 (2003) https://doi.org/10.1037/0022-0663.95.4.667

[39] Yang, S., Mei, B.: Understanding learners' use of augmented reality in language learning: insights from a case study. Journal of Education for Teaching **44**(4), 511–513 (2018) https://doi.org/10.1080/02607476.2018.1450937

[40] Vazquez, C.D., Nyati, A.A., Luh, A., Fu, M., Aikawa, T., Maes, P.: 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). https://doi.org/10.1145/3027063.3053098

[41] Edge, D., Searle, E., Chiu, K., Zhao, J., Landay, J.A.: Micromandarin: Mobile language learning in context. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 3169–3178 (2011). https://doi.org/10.1145/1978942.1979413

[42] Dalim, C.S.C., Sunar, M.S., Dey, A., Billinghurst, M.: Using augmented reality with speech input for non-native children's language learning. International Journal of Human-Computer Studies **134**, 44–64 (2020) https://doi.org/10.1016/j.ijhcs.2019.10.002

[43] Li, S., Chen, Y., Whittinghill, D.M., Vorvoreanu, M.: A pilot study exploring augmented reality to increase motivation of chinese college students learning english. In: 2014 ASEE Annual Conference & Exposition, pp. 24–85 (2014)

[44] Draxler, F., Labrie, A., Schmidt, A., Chuang, L.L.: 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). https://doi.org/10.1145/3313831.3376537

[45] Ibáñez, M.B., Di Serio, Á., Villarán, D., Kloos, C.D.: Experimenting with electromagnetism using augmented reality: Impact on flow student experience and educational effectiveness. Computers & Education **71**, 1–13 (2014) https://doi.org/10.1016/j.compedu.2013.09.004

[46] Strzys, M.P., Kapp, S., Thees, M., Klein, P., Lukowicz, P., Knierim, P., Schmidt,

A., Kuhn, J.: Physics holo.lab learning experience: using smartglasses for augmented reality labwork to foster the concepts of heat conduction. European Journal of Physics **39**(3), 035703 (2018) https://doi.org/10.1088/1361-6404/aaa8fb

[47] Hautasaari, A., Hamada, T., Ishiyama, K., Fukushima, S.: Vocabura: A method for supporting second language vocabulary learning while walking. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **3**(4), 135–113523 (2019) https://doi.org/10.1145/3369824

[48] Weerasinghe, M., Biener, V., Grubert, J., Deja, J.A., Attygalle, N.T., Trajkovska, K., Kljun, M., Pucihar, K.Č.: Vocabulary replicated: comparing teenagers to young adults. In: 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 283–285 (2022). https://doi.org/10.1109/ISMAR-Adjunct57072.2022.00064

[49] PTC: Vuforia Developer Portal. https://developer.vuforia.com/. Accessed: 2022-07-27

[50] Castle, R.O., Murray, D.W.: Object recognition and localization while tracking and mapping. In: 2009 8th IEEE International Symposium on Mixed and Augmented Reality, pp. 179–180 (2009). https://doi.org/10.1109/ISMAR.2009.5336477

[51] Salas-Moreno, R.F., Newcombe, R.A., Strasdat, H., Kelly, P.H.J., Davison, A.J.: Slam++: Simultaneous localisation and mapping at the level of objects. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1352–1359 (2013). https://doi.org/10.1109/CVPR.2013.178

[52] Salman, M., Pearson, M.J.: Whisker-ratslam applied to 6d object identification and spatial localisation. In: Vouloutsi, V., Halloy, J., Mura, A., Mangan, M., Lepora, N., Prescott, T.J., Verschure, P.F.M.J. (eds.) Biomimetic and Biohybrid Systems, pp. 403–414. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-95972-6_44

[53] Masarwa, S., Kreichman, O., Gilaie-Dotan, S.: Larger images are better remembered during naturalistic encoding. Proceedings of the National Academy of Sciences **119**(4), 2119614119 (2022) https://doi.org/10.1073/pnas.2119614119

[54] Lewis, J.R., Sauro, J.: The factor structure of the system usability scale. In: Proceedings of the International Conference on Human Centered Design, pp. 94–103. Springer, Berlin, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02806-9_12

[55] Schrepp, M., Hinderks, A., Thomaschewski, J.: Design and evaluation of a short version of the user experience questionnaire (ueq-s). International Journal of Interactive Multimedia and Artificial Intelligence **4**(6), 103–108 (2017)

https://doi.org/10.9781/ijimai.2017.09.001

[56] Hart, S.G.: Nasa-task load index (nasa-tlx); 20 years later. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 50, pp. 904–908. Sage Publications, Los Angeles, CA (2006). https://doi.org/10.1177/154193120605000

[57] Hart, S.G., Staveland, L.E.: Development of nasa-tlx (task load index): Results of empirical and theoretical research. In: Human Mental Workload. Advances in Psychology, vol. 52, pp. 139–183 (1988). https://doi.org/10.1016/S0166-4115(08)62386-9

[58] Paas, F.G., Van Merriënboer, J.J.: The efficiency of instructional conditions: An approach to combine mental effort and performance measures. Human factors **35**(4), 737–743 (1993) https://doi.org/10.1177/001872089303500412

[59] Halabi, A.K.: Applying an instructional learning efficiency model to determine the most efficient feedback for teaching introductory accounting. Global Perspectives on Accounting Education **3**(1), 93–113 (2006)

[60] Clark, R.C., Nguyen, F., Sweller, J.: Efficiency in Learning: Evidence-Based Guidelines to Manage Cognitive Load. Pfeiffer, San Francisco, CA (2006)

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

[62] Laugwitz, B., Held, T., Schrepp, M.: Construction and evaluation of a user experience questionnaire. In: Holzinger, A. (ed.) HCI and Usability for Education and Work, pp. 63–76. Springer, Berlin, Heidelberg (2008). https://doi.org/10.1007/978-3-540-89350-9_6

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

[64] Blair, R.C., Higgins, J.J.: Comparison of the power of the paired samples t test to that of wilcoxon's signed-ranks test under various population shapes. Psychological Bulletin **97**(1), 119 (1985) https://doi.org/10.1037/0033-2909.97.1.119

[65] Woolson, R.F.: Wilcoxon signed-rank test. Wiley encyclopedia of clinical trials, 1–3 (2007)

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

[67] Cohen, J.: Statistical power analysis for the behavioral sciences. England: Routledge (1988)

[68] Hinderks, A., Schrepp, M., Thomaschewski, J.: A benchmark for the short version of the user experience questionnaire. In: WEBIST, pp. 373–377 (2018)

[69] Auda, J., Gruenefeld, U., Faltaous, S., Mayer, S., Schneegass, S.: A scoping survey on cross-reality systems. ACM Comput. Surv. **56**(4) (2023) https://doi.org/10.1145/3616536

[70] Simeone, A.L., Khamis, M., Esteves, A., Daiber, F., Kljun, M., Pucihar, K., Isokoski, P., Gugenheimer, J.: International workshop on cross-reality (xr) interaction. In: Companion Proceedings of the 2020 Conference on Interactive Surfaces and Spaces. ISS Companion '20, pp. 111–114. Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3380867.3424551

[71] Guo, J., Weng, D., Zhang, Z., Jiang, H., Liu, Y., Wang, Y., Duh, H.B.-L.: Mixed reality office system based on maslow's hierarchy of needs: Towards the long-term immersion in virtual environments. In: 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 224–235 (2019). https://doi.org/10.1109/ISMAR.2019.00019 . IEEE

[72] Guo, J., Weng, D., Zhang, Z., Liu, Y., Wang, Y.: Evaluation of maslows hierarchy of needs on long-term use of hmds–a case study of office environment. In: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 948–949 (2019). https://doi.org/10.1109/VR.2019.8797972 . IEEE

[73] Shen, R., Weng, D., Chen, S., Guo, J., Fang, H.: Mental fatigue of long-term office tasks in virtual environment. In: 2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 124–127 (2019). https://doi.org/10.1109/ISMAR-Adjunct.2019.00-65 . IEEE

[74] Steinicke, F., Bruder, G.: A self-experimentation report about long-term use of fully-immersive technology. In: Proceedings of the 2nd ACM Symposium on Spatial User Interaction, pp. 66–69 (2014). https://doi.org/10.1145/2659766.2659767

[75] Biener, V., Kalamkar, S., Nouri, N., Ofek, E., Pahud, M., Dudley, J.J., Hu, J., Kristensson, P.O., Weerasinghe, M., Pucihar, K.Č., *et al.*: Quantifying the effects of working in vr for one week. IEEE Transactions on Visualization and Computer Graphics **28**(11), 3810–3820 (2022) https://doi.org/10.1109/TVCG.2022.3203103

[76] Richardson, J.T.: Social class limitations on the efficacy of imagery mnemonic instructions. British Journal of Psychology **78**(1), 65–77 (1987) https://doi.org/10.1111/j.2044-8295.1987.tb02226.x

[77] Oliver, M.C., Bays, R.B., Zabrucky, K.M.: False memories and the drm paradigm: effects of imagery, list, and test type. The Journal of General Psychology **143**(1), 33–48 (2016) https://doi.org/10.1080/00221309.2015.1110558

[78] Kulhavy, R., Swenson, I.: Imagery instructions and the comprehension of text. British Journal of Educational Psychology **45**(1), 47–51 (1975) https://doi.org/10.1111/j.2044-8279.1975.tb02294.x

[79] Kirschner, P.A., Kirschner, F.: Mental effort. Springer (2012). https://doi.org/10.1007/978-1-4419-1428-6_226

[80] Salomon, G.: Television is" easy" and print is" tough": The differential investment of mental effort in learning as a function of perceptions and attributions. Journal of educational psychology **76**(4), 647 (1984) https://doi.org/10.1037/0022-0663.76.4.647

[81] Yamazaki, Y.: Learning styles and typologies of cultural differences: A theoretical and empirical comparison. International Journal of Intercultural Relations **29**(5), 521–548 (2005) https://doi.org/10.1016/j.ijintrel.2005.07.006

[82] Hammond, C.: Culturally responsive teaching in the japanese classroom: A comparative analysis of cultural teaching and learning styles in japan and the united states. Kyoto Gakuen University Faculty of Economics Bulletin **17**(1), 41–50 (2007)

[83] Yamazaki, Y., Toyama, M., Attrapreyangkul, T.: Cross-cultural differences in learning style and learning skills: A comparison of japan, thailand, and the usa. In: Handbook of Research on Cross-Cultural Business Education, pp. 160–182. IGI Global, Pennsylvania, USA (2018). https://doi.org/10.4018/978-1-5225-3776-2.ch008

[84] Kolb, D.A.: Experiential Learning: Experience as the Source of Learning and Development. Prentice Hall, Englewood Cliffs, NJ (1984). https://doi.org/10.1002/job.4030080408

[85] Clark, J.M., Paivio, A.: Dual coding theory and education. Educational psychology review **3**(3), 149–210 (1991) https://doi.org/10.1007/BF01320076

[86] Mayer, R.E.: Constructivism as a theory of learning versus constructivism as a prescription for instruction. Constructivist instruction: Success or failure, 184–200 (2009) https://doi.org/10.4324/9780203878842

[87] Kirschner, P.A., Sweller, J., Clark, R.E.: 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) https://doi.org/10.1207/s15326985ep4102_1

[88] Kuhn, D.: Is direct instruction an answer to the right question? Educational psychologist **42**(2), 109–113 (2007) https://doi.org/10.1080/00461520701263376

[89] Manganello, F.: Constructivist instruction: Success or failure? Journal of Educational Technology & Society **13**(3), 281–284 (2010). Accessed 2025-03-30

[90] Koedinger, K.R., Aleven, V.: Exploring the assistance dilemma in experiments with cognitive tutors. Educational Psychology Review **19**(3), 239–264 (2007) https://doi.org/10.1007/s10648-007-9049-0

[91] Wise, A.F., O'Neill, K.: Beyond more versus less: A reframing of the debate on instructional guidance. (2009) https://doi.org/10.4324/9780203878842

[92] Nadolski, R.J., Kirschner, P.A., Van Merriënboer, J.J.: Optimizing the number of steps in learning tasks for complex skills. British Journal of Educational Psychology **75**(2), 223–237 (2005) https://doi.org/10.1348/000709904X22403

[93] Shepard, R.N.: Recognition memory for words, sentences, and pictures. Journal of verbal Learning and verbal Behavior **6**(1), 156–163 (1967) https://doi.org/10.1016/S0022-5371(67)80067-7