Progressive Sentences: Combining the Benefits of Word and Sentence Learning

Nuwan Janaka
nuwanj@u.nus.edu
Smart Systems Institute; Synteraction Lab
National University of Singapore
Singapore

Shengdong Zhao*
shengdong.zhao@cityu.edu.hk
Synteraction Lab, School of Creative Media & Department of Computer Science
City University of Hong Kong
Hong Kong, China

Ashwin Ram†
ashwinram10@gmail.com
Saarland University,
Saarland Informatics Campus
Saarbrücken, Germany

Ruoxin Sun
sun.ruoxin@u.nus.edu
School of Computing; Synteraction Lab
National University of Singapore
Singapore

Sherisse Tan Jing Wen
sherisse_tjw@u.nus.edu
School of Computing
National University of Singapore
Singapore

Danae Li
amethyst_li@outlook.com
MusicX Lab
SKYWORK AI PTE. LTD.
Beijing, China

David Hsu
dyhsu@comp.nus.edu.sg
School of Computing; Smart Systems Institute
National University of Singapore
Singapore

Abstract

The rapid evolution of lightweight consumer augmented reality (AR) smart glasses (a.k.a. optical see-through head-mounted displays) offers novel opportunities for learning, particularly through their unique capability to deliver multimodal information in just-in-time, micro-learning scenarios. This research investigates how such devices can support mobile second-language acquisition by presenting progressive sentence structures in multimodal formats. In contrast to the commonly used vocabulary (i.e., word) learning approach for novice learners, we present a "progressive presentation" method that combines both word and sentence learning by sequentially displaying sentence components (subject, verb, object) while retaining prior context. Pilot and formal studies revealed that progressive presentation enhances recall, particularly in mobile scenarios such as walking. Additionally, incorporating timed gaps between word presentations further improved learning effectiveness under multitasking conditions. Our findings demonstrate the utility of progressive presentation and provide usage guidelines for educational applications—even during brief, on-the-go learning moments.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; • **Applied computing** → **Education**.

Keywords

learning, second-language, multimodal presentation, progressive presentation, sentence, illustration, composition, timing

ACM Reference Format:
Nuwan Janaka, Shengdong Zhao, Ashwin Ram, Ruoxin Sun, Sherisse Tan Jing Wen, Danae Li, and David Hsu. 2025. Progressive Sentences: Combining the Benefits of Word and Sentence Learning. In *Adjunct Proceedings of the 27th International Conference on Mobile Human-Computer Interaction (MobileHCI '25 Adjunct), September 22–25, 2025, Sharm El-Sheikh, Egypt*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3737821.3749564

1 Introduction

The rise of mobile learning reflects growing demand for educational tools that accommodate fast-paced, multitasking lifestyles—particularly in language learning, where consistent practice is essential [7, 36]. However, many learners struggle to maintain regular study due to frequent interruptions, cognitive load fluctuations, and physical limitations such as small mobile screens [12, 32, 43]. These challenges call for lightweight, cognitively efficient solutions that integrate seamlessly into daily routines.

Advances in digital technology offer promising avenues to meet these needs. In particular, consumer-grade augmented reality (AR)

*Corresponding Author.

†Research conducted while at Synteraction Lab, University of Singapore.

MobileHCI '25 Adjunct, Sharm El-Sheikh, Egypt

ACM ISBN 979-8-4007-1970-7/2025/09
https://doi.org/10.1145/3737821.3749564

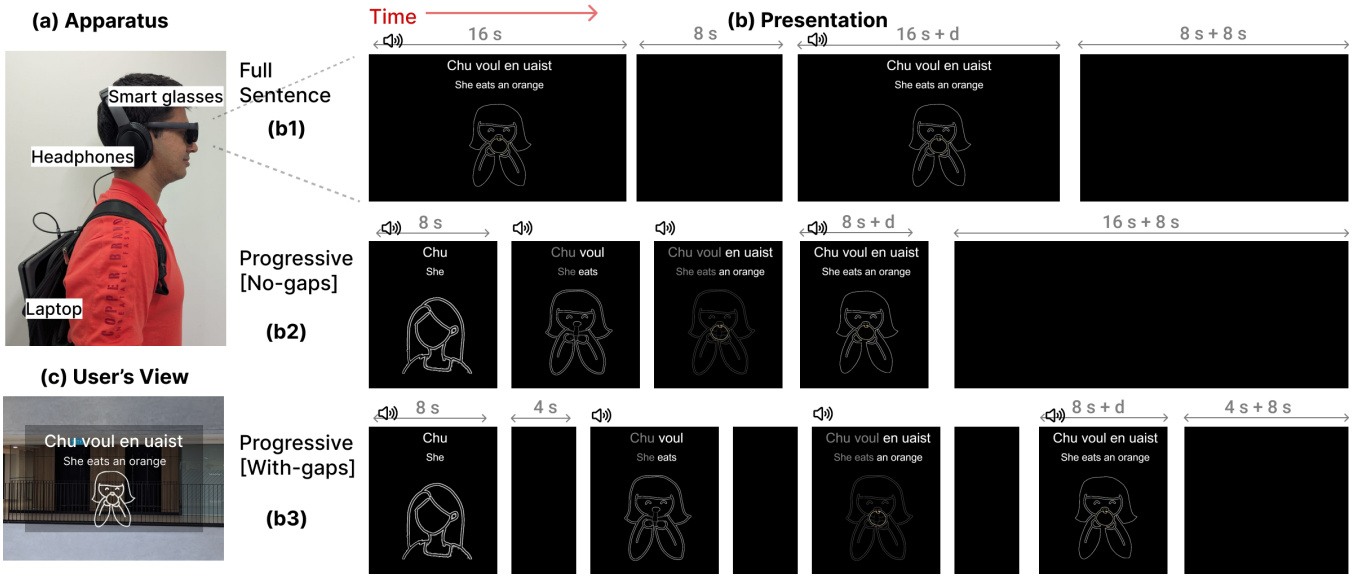


Figure 1: Apparatus and presentation formats used in the study. (a) A participant wearing AR smart glasses and headphones while walking. (b) The participant views learning content through the smart glasses, illustrating the *Full* (b1) and *Progressive* presentation formats, including the *NoGap* (b2) and *Gap* (b3) conditions. The durations for each modality (L1/L2 text, L1/L2 audio, images) and no-content gaps were kept consistent. An extra 8 seconds was added between sentences, and an additional $d = 3$ seconds was included for the final full sentence to ensure adequate viewing time. Black regions in the smart glasses display represent *transparent* areas. (c) Simulated user's view showing how digital content is superimposed on the physical environment. Due to the tinting of smart glasses, the content appears darker than normal. In reality, users do not notice this effect as the frame of the smart glasses blocks peripheral vision.

devices—lightweight Optical See-Through Head-Mounted Displays (OST-HMDs or AR smart glasses)—are emerging as viable tools for everyday use [5, 20]. These assisted reality devices, such as the XReal Air, Rokid, and TCL RayNeo 2, overlay digital content in the user's field of view, enabling quick-glance, context-independent interactions [35]. Unlike immersive AR systems that spatially anchor content, these streamlined devices support subtle, ambient engagement, making them well-suited for opportunistic mobile learning [21].

Building on this potential, we investigate how lightweight AR glasses can support mobile language learning. A key question arises: *how can multimodal information be composed and presented effectively within the constraints of these devices?* Rather than developing complex, context-aware applications, we focus on optimizing content presentation to suit device limitations. This approach addresses both technical constraints of OST-HMDs and real-world usage—typically brief, multitasked interactions.

Grounded in Cognitive Theory of Multimedia Learning [30, 31], our study examines how combinations of visual (e.g., text, images) and auditory (e.g., pronunciation) elements enhance language learning while managing cognitive load. We aim to develop presentation methods that are lightweight yet effective, improving outcomes within the limits of current OST-HMDs.

To this end, we conducted pilot and formal experiments evaluating sentence presentation techniques using lightweight OST-HMDs. We tested different multimodal combinations—text, audio, images—to assess their effect on learning. We also examined how multitasking influences learning, acknowledging varied usage contexts from stationary to mobile settings.

Our findings show that a progressive sentence presentation strategy supports language learning during mobile multitasking. This method incrementally introduces sentence components before presenting the full sentence, helping learners grasp both individual elements and overall structure. Additionally, inserting strategic word gaps enhances recall in mobile contexts.

This work provides preliminary empirical evidence on AR-assisted language learning and offers design insights for educational applications. Our findings highlight the potential of lightweight OST-HMDs to support natural, engaging information acquisition.

2 Related Work

2.1 Mobile Language Learning and Augmented Reality

Mobile microlearning leverages short, intermittent periods to deliver bite-sized language content (e.g., vocabulary), supporting acquisition during everyday tasks [7, 36]. It has gained traction due to its accessibility and convenience [13–15]. To enhance retention, prior work has explored situated contexts, such as location-based microlearning [15]. Recently, OST-HMDs or AR smart glasses have emerged as promising platforms for context-aware learning, offering hands-free, heads-up access to digital content superimposed in the user's view (i.e., physical world) [20]. Systems like *ARbis Pictus* [19] and *VocabulARy* [41] improve recall over non-AR approaches

(e.g., desktop flashcards or tablet annotations) by embedding visual annotations in situated physical contexts. While most prior work has focused on vocabulary learning in both AR and non-AR settings, emerging evidence suggests that embedding vocabulary within broader semantic structures—such as sentence-level presentations or narratives—enhances comprehension and retention in stationary, non-AR environments [9, 27, 39]. Our work extends this by introducing a progressive sentence disclosure method on AR smart glasses, integrating isolated vocabulary with contextual sentence presentation under mobile multitasking conditions.

2.2 Multimodal Presentation and Cognitive Load

Mayer's Cognitive Theory of Multimedia Learning (CTML), along with the Multimedia Principle, posits that effective learning combines visual, auditory, and textual modalities to support richer encoding and improved retention [30, 31]. However, poorly designed multimodal presentations can increase cognitive load, especially for novice learners [16, 40]. This aligns with Sweller's Cognitive Load Theory, which distinguishes between *intrinsic load* (material complexity), *extraneous load* (load from poor design), and *germane load* (resources allocated to schema construction) [31, 40]. These loads interact dynamically: reducing extraneous load frees cognitive capacity for germane processing. In language learning, intrinsic load is high due to unfamiliar vocabulary and grammar, making it essential to minimize extraneous load through effective presentations/interfaces while promoting germane load. Strategies such as the segmenting principle—breaking content into smaller, progressive units—help manage extraneous load by enabling incremental processing [26, 31]. Our approach builds on these principles by employing progressive disclosure: sequentially revealing sentence components with synchronized multimodal cues (text, images, audio). This design reduces extraneous load and supports multitasked learning [21, 34], while potentially enhancing germane load by reinforcing semantic links between words in context.

3 Study Overview

This research investigates effective sentence presentation styles for language learning in mobile contexts using assisted reality [35]. While word-level learning supports beginners [18], sentence-based approaches provide richer context for deeper learning [2, 9, 27]. To explore this, we conducted informal pilot studies to refine 1) modalities for sentence presentation (*Pilot 1*, N=5), and 2) chunking styles that support mobile learning (*Pilot 2*, N=6). These led to a progressive sentence disclosure method that decomposes sentences into subject-verb-object structures. While promising, the pilots also revealed the potential influence of timing gaps between word displays (*Pilot 3*, N=2), prompting a formal investigation (*Study 1*, N=12). A follow-up pilot (*Pilot 4*, N=4) evaluated the method's ecological validity.

4 Common Setting

All pilot (*Pilot 1-4*) and formal (*Study 1*) studies were conducted in a controlled lab environment to minimize confounding factors (e.g., noise, lighting) and shared the following common elements.

4.1 Participants

Participants (total of 27) were adult learners (age 19–26) from the university community, self-reporting full professional fluency in English as their first language (L1) and willingness to learn a second language (L2). Demographics, language background, and fluency are detailed in each study section. All had normal or corrected-to-normal vision and hearing, with no reported impairments.

All procedures were IRB-approved. Participants gave informed consent and were compensated at $7.25/hour. **No** participant took part in more than one study.

4.2 Apparatus

As shown in Figure 1(a), participants wore XREAL Air glasses (1920×1080 px, 46° FoV, 60Hz) and Bose QuietComfort Bluetooth noise-canceling headphones for audio and to secure the glasses. The AR glasses mirrored a laptop's display via USB[1]. The learning content presentations were controlled using a Python backend and VueJS frontend on a MacBook Pro (14-inch, M3 Pro). During walking conditions, the laptop was carried in a light backpack. The laptop also shared its screen via Zoom with the experimenter for monitoring. See https://github.com/Synteraction-Lab/ProgressiveSentences for implementation details.

4.3 Task

We simulated two everyday AR usage scenarios:

***Walking* [Multitasked Learning]**: Mobile multitasking scenario simulating real-world contexts like commuting, where users' attention is divided between non-learning (e.g., walking, checking the surroundings) and learning tasks [43]. Participants assembled a 6-piece Duplo Lego block to match a color/shape sample, transporting pieces along a 5-meter path, simulating hands-busy multitasking. Block colors varied to ensure consistent difficulty.

***Sitting* [Focused Learning]**: Seated language learning to simulate stationary usage (e.g., focused learning while sitting).

4.4 Materials

Participants studied L2 words or sentences with L1 meanings, depending on the condition.

4.4.1 L2 Language. We used Wuggy pseudowords [24] as L2, commonly applied in HCI and linguistics (e.g., [29, 37]). This controlled for prior knowledge, avoided cognates, and ensured consistent complexity [8].

4.4.2 L1 Sentences. Focusing on A1 fluency (beginner level) [3], we created 40 English (L1) statements using basic Subject-Verb-Object grammar. Nouns were drawn from Ogden's Basic English [33], with common pronouns and function words (articles/prepositions). Each sentence had 2–3 unique L1 words (with 4-5 total words).

4.4.3 Sentence List. Each L1 word was mapped to a Wuggy L2 pseudoword. Three co-authors reviewed all pairs to ensure no cognates. L2 sentences were created via word-to-word mappings (see Table 1).

[1]XREAL Air projects a 130-inch equivalent screen at 4 meters

Table 1: Sample L1-L2 words and L1-L2 sentence list

L1 Word (English)	L2 word (Wuggy)
he	sa
she	chu
reads	soan

L1 Sentence (English)	L2 Sentence (Wuggy)	L2 Sentence (Norwegian)
He reads a book	Sa soan en beal	Han leser en bok
She eats an orange	Chu voul en uaist	Hun spiser en appelsin
The boy kicks a ball	Snu bist yills en bune	Gutten sparker en ball

4.4.4 Multimodal Representations. Audio (L1 and L2 pronunciations) was generated via OpenAI TTS API (tts-1, voice: nova, speed: 0.9) and manually verified. Mismatched L2 phrases were corrected by concatenating word-level audio. Audio duration ranged from 0.1–3 seconds.

Google Image Search was used to retrieve outline-style monochrome icons for each word, phrase, and sentence, minimizing visual distraction [16, 38]. Three co-authors iteratively rated and refined images for adherence to the selected 'Seven C' principles (concise, concrete, coherent, comprehensible, correspondent) [1].

4.4.5 Presentation Layout and Style. As shown in Figure 1(b), L2 text, L1 text, and images appeared top-center, minimizing visual obstruction. Transparent-outline images preserved central visibility and enabled quick recognition [23, 44]. All visuals were rendered in white for consistency [38].

Following prior work [18, 37] and informal testing, L2 text appeared before its L1 meaning. Audio was synchronized with text; images appeared with L2 text. Words were displayed for at least 6 seconds, phrases (consisting of 2 or more words) for 8 seconds, and sentences for 12 seconds. Each item was shown twice to enable review. We avoided spaced repetition [14] to equalize exposure time. All modalities and gaps were time-matched across conditions (i.e., equal total exposure time). Refer to individual study sections for presentation timing details.

4.5 Study Design

All studies (pilot and formal) employed a within-subject, repeated-measures design. Experimental conditions were counterbalanced using a Latin square. To control learning material complexity, 3–4 sentences were grouped per condition with matched sentence lengths and L2 word counts. Learning material groups were presented in a fixed order, while experimental conditions were counterbalanced, to minimize biases across groups.

4.6 Measures

Table 2 lists all common measures. We evaluated 'remembering' and 'understanding' based on Bloom's Revised Taxonomy [25]. Primary metrics were free recall (L2–L1 word pairs, no cues) and cued recall (L2 to L1; *Recall*) [16, 25]. *Word-Recall* and *Seen-Sentence-Recall* assessed *remembering* for studied content, while *Unseen-Sentence-Recall* tested *understanding* via novel L2 sentences combining known words [40]. For example, if a participant saw 'She eats an orange' and 'The boy kicks a ball,' an unseen sentence might be 'The boy eats a ball,' which, although potentially unrealistic, would be grammatically correct. A questionnaire collected all four *Recall* types in this order: *Free-Recall*, *Word-Recall*, *Seen-Sentence-Recall*, *Unseen-Sentence-Recall*. Scoring followed prior work [19, 41], allowing partial credit for correct structure and meaning (see Appendix A.1).

Learning cognitive load ratings were collected via 7-point Likert scales per condition, based on validated measures [11, 28].

We also collected preference rankings (*Preference*) and qualitative feedback through semi-structured interviews on strategies, challenges, and suggestions.

Analysis. Data were analyzed using repeated-measures or factorial ANOVA. If assumptions were violated, we used Friedman tests or ART-based ANOVA [42]. Normality and sphericity were checked using Shapiro-Wilk and Mauchly's tests; post-hoc tests used paired t-tests, ART contrasts, or Wilcoxon tests with Bonferroni correction (see Appendix A.2 for details).

Transcribed interview recordings and open-ended responses in questionnaires were analyzed using thematic analysis as described in Braun and Clarke's methodology [6].

4.7 Procedure

After informed consent, participants were introduced to tasks, presentation styles, and sample questions. They could practice aloud but not take notes or rehearse outside sessions. Participants completed all conditions in a set order. After each, they completed a questionnaire. Partial answers for immediate recall were allowed, but guessing was discouraged. A 2-minute break followed each condition.

At the end, participants ranked their preferred styles and completed a 5–10 minute semi-structured interview. Sessions lasted around 1 hour, including training. A delayed recall questionnaire was sent online after 7 days to assess retention [34, 37]. Participants were instructed to rely solely on memory.

5 *Pilot 1*: Identifying Suitable Modalities for Sentences during *Walking*

In this pilot, we compared five common modality combinations (i.e., *NoImageNoAudio:* L2 Text → L1 text; *NoImageWithAudio:* L2 text + L2 audio → L1 text; *ImageWithAudio:* L2 text + L2 audio + Image → L1 audio; *ImageWithText:* L2 text + L2 audio + Image → L1 text; *ImageWithTextAudio:* L2 text + L2 audio + Image → L1 text + L1 audio) [31] with 5 (3F, 2M) participants, to identify complementary modalities for sentence-based learning during *Walking* (Sec 4).

While there were no statistically significant differences, the text-only condition (i.e., *NoImageNoAudio*) had the lowest recall scores, whereas the text+image+audio combination (i.e., *ImageWithTextAudio*) yielded the highest, aligning with the *Multimedia Principle* [31]. Specifically, all participants preferred the combination of *L2 text + L2 audio + image, followed by L1 text + L1 audio.* Audio helped with

Table 2: Measures used in each study

Measure	Sub-measure	Definition/Operationalization
Recall	*Free-Recall*	The total score for writing the correct L1-L2 words pairs without any cues
	Word-Recall	The total score for writing the correct L1 word for the seen (given) L2 word
	Seen-Sentence-Recall	The total score for writing the correct L1 sentence for the seen L2 sentence
	Unseen-Sentence-Recall	The total score for writing the correct L1 sentence for the unseen L2 sentence
Cognitive Load (Learning [11, 28])	*DifficultToUnderstand*	"*I find the learning content during this session/style difficult to understand.*"
	Confusing	"*The learning content presented by this learning session/style is confusing for me.*"
	EasyToRemember	"*This learning session/style allows me to easily remember most of the learning content.*"
	AbsorbedInLearning	"*I was absorbed in using this session/style to learn the new language.*"
	Enjoyable	"*It is enjoyable to learn a new language with this session/style.*"
Preference	*Preference*	The overall preference ranking

pronunciation (L2) and meaning (L1) without drawing visual attention from the mobile multitasking. However, participants noted difficulty understanding full sentences due to limited support for individual word meanings.

6 *Pilot 2*: Evaluating *Progressive Disclosure* during *Walking*

Progressive Presentation of Sentences. To address word-level comprehension within sentences, we developed a *progressive disclosure presentation* (*Progressive*, Figure 1(b)) based on the *Segmenting Principle* [31]. Sentences were broken into subject, verb, and object components, presented sequentially to support stepwise understanding. Previously shown words remained visible, and new ones were highlighted using transparency cues following the *Signaling Principle* [31]. Once all components were introduced, the full sentence was displayed.

Pilot 2. We compared *Progressive* with full sentences (*Full*) and individual words presented randomly (*Word*, no sentence structure), with 6 (4F, 2M) participants during *Walking*. Although differences were not statistically significant, *Progressive* yielded the highest recall. All participants preferred it, noting improved memory for both individual words and sentence structure. It also reduced cognitive load compared to *Full* and helped link L2 and L1 words.

However, three participants (3/6) suggested adding pauses between words in *Progressive* to aid memory consolidation.

7 *Pilot 3*: Determining Optimal Word Gaps for *Progressive* during *Walking*

This pilot tested gap durations (2s, 4s, 6s, 8s) with two (1F, 1M) participants during *Walking*. The findings suggested that the gap should be neither too short—making it unnoticeable—nor too long, which could cause participants to forget the connection between words in a sentence. A 4-second gap was identified as a balanced option and selected for further study. While the optimal duration warrants further investigation, this finding is consistent with prior research on notification resumption gaps, which support the acquisition of secondary information during mobile multitasking [22].

8 *Study 1*: Understanding the Effect of *Mobility* and *Word-Gap* on *Progressive* Presentation

Participants: Twelve volunteers (7F, 5M; age $M = 22.3$, $SD = 3.3$) participated in the study. Ten were native English speakers; two had professional fluency. All but one had used mobile language apps; two had limited prior AR smart glasses exposure.

Design and Procedure: Following Sec 4.7, the study used a 2×2 design: *Mobility* (*Sitting*, *Walking*) × *Word-Gap* (*NoGap* = 0s, *Gap* = 4s). Participants learned 12 sentences (3 per condition) and 32 new L2 words (8 per condition).

Results: Figure 2 shows the significant results (refer to Appendix B: Figure 3, Table 3 for details). For *Walking*, all participants successfully completed the Lego assembly.

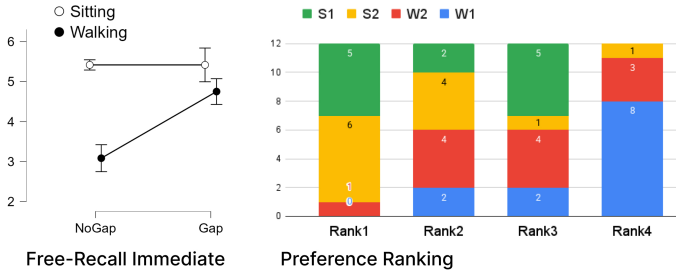


Figure 2: *Free-Recall* (Immediate) interactions and *Preference* rankings for the *Study 1* (N=12). Here, S = *Sitting*, W = *Walking*, 1 = *NoGap*, and 2 = *Gap*. For example, S1 represents, *Sitting* with *NoGap*.

Recall: Repeated-measures ANOVA with ART [42] showed significant main effects of *Mobility* ($p < 0.05$).

In **immediate** *Recall*, *Sitting* yielded significantly higher *Free-Recall*, *Word-Recall*, and *Seen-Sentence-Recall* scores ($p < 0.05$). *Free-Recall* also showed significant main and interaction effects of *Word-Gap* ($F_{1,33} = 5.62$, $p = 0.02$; $F_{1,33} = 5.96$, $p = 0.02$). Post-hoc results (Figure 2) revealed *Gap* ($M = 4.9$, $SD = 2.0$) outperformed *NoGap* ($M = 4.3$, $SD = 1.7$; $p_{bonf} < 0.05$, $d = 0.44$). In *Walking*, *Gap* ($M = 4.7$, $SD = 2.2$) significantly exceeded *NoGap* ($M = 3.1$, $SD = 1.8$); no such effect was found in *Sitting*.

In **delayed (7-day)** *Recall*, *Sitting* again led to significantly better scores for *Word-Recall* ($F_{1,33} = 13.76$, $p < 0.001$) and *Seen-Sentence-Recall* ($F_{1,33} = 5.75$, $p = 0.02$). No other significant effects were found.

Learning Cognitive Load: All subjective scales showed significant main effects of *Mobility* ($p < 0.05$) via ART-based ANOVA. *Sitting* was rated lower in *DifficultToUnderstand* and *Confusing*, and higher in *EasyToRemember*, *AbsorbedInLearning*, and *Enjoyable*. Participants attributed this to divided attention during multitasking, which diminished the capacity for remembering learning content.

Preference results (Figure 2): 11/12 preferred *Sitting* for better focus. One preferred *Walking* to "*keep the mind active*". Word gaps did not affect preferences during *Sitting*, though two noted increased mind-wandering. During *Walking*, the majority (9/12) preferred having word gaps as it "*gave time to practice and solidify the words in memory*" and "*one without any gaps is [felt] too fast.*" Some of them (6/9) also mentioned that the gaps between words "*gave more chances to review previous words and prepare for upcoming words while attention was split.*" The rest (3/12) did not notice the gaps.

Subjective Feedback: When asked to compare *Progressive* and *Full* presentations, most participants (10/12) preferred *Progressive*, citing its one-to-one word mapping, which helped them understand individual meanings and how words connect. In contrast, *Full* sentences would require more cognitive effort to link L2 and L1 words, leaving less time for review. Similarly, all participants favored *Progressive* over random individual words, highlighting the importance of sentence context in supporting comprehension.

9 *Pilot 4*: Preliminary Ecological Validation of the *Progressive* Presentation

To assess generalizability beyond artificial corpora, we used Norwegian (Bokmål) as L2. This pilot followed the *Study 1* procedure (Sec 8) with differences in materials and a pre-test (see Appendix C).

Four volunteers (1F, 3M; age $M = 24.3$, $SD = 2.1$), without prior knowledge of Norwegian or related languages (e.g., German or Norse), participated in the study. Three were native English speakers; one had professional fluency. All had experience with mobile language apps.

Results: Each participant learned 12 sentences (3 new sentences per condition) and 28 new vocabulary words (7 new vocabulary words per condition) across four conditions. See details in Appendix C.3 (Table 4, Figure 4).

Sitting resulted in significantly higher **immediate** *Free-Recall* than *Walking* ($F_{1,3} = 32.0$, $p = 0.01$, $\eta_p^2 = 0.91$); other differences were not significant. The trends observed were consistent with those in *Study 1*. Regarding preference and interview feedback, all participants preferred the *Sitting* over *Walking* (with no impact of *Word-Gaps*), and *Gap* over *NoGap* during *Walking*. Similar to *Study 1*, participants reported that word gaps during *Walking* helped them consolidate memory before the next word appeared, whereas during *Sitting*, the gaps had minimal perceived impact. As one noted, "*I felt that in the second walking [without gaps], the words went too fast, but all were ok for sitting.*" These findings support the applicability of *Progressive* for real languages with similar grammar.

10 Discussion

Why is progressive presentation effective for multitasked learning? Our findings suggest that *Progressive* presentation in AR smart glasses—sequentially revealing information while retaining prior content and highlighting new elements—supports language learning by facilitating word familiarity and preserving sentence context. This method aids learners in understanding how words connect meaningfully. These results align with Layered Serial Visual Presentation (LSVP) [34], which found that persistent, sequential delivery enhances recall in mobile video-learning contexts by improving attention control, consistent with Perceptual Load Theory (PLT) [26].

We also found that inserting *word gaps* (i.e., no-content intervals) improves short-term retention during multitasking. Although no significant effects were observed for long-term retention—likely due to small sample size—our findings align with prior work emphasizing system pacing that matches users' cognitive load and processing capacity [17]. While word gaps had minimal benefit in stationary settings, they were helpful under multitasking, giving learners time to consolidate new information. Although we used a 4-second default, participant feedback suggests that the optimal gap duration may vary based on task demands, highlighting the value of adaptive timing to balance focus between primary and secondary tasks—similar to notification resumption gaps during multitasking [22].

When and how should progressive presentation be used? User feedback indicates that *Progressive* is most effective as a scaffolding technique for novice learners encountering unfamiliar vocabulary. It enables gradual exposure while maintaining sentence context, supporting both word- and sentence-level comprehension. Once learners develop familiarity, full-sentence presentation may be more efficient, allowing them to infer meanings without stepwise support.

In this study, sentence segmentation and image selection were done manually to ensure coherence. To scale this approach, large language models (LLMs) could automatically segment sentences, while text-to-image models [4] could generate relevant visuals. Post-processing methods could evaluate image-sentence alignment and assemble semantically coherent sequences [9, 10].

10.1 Limitations and Future Work

While this study offers preliminary evidence, several limitations affect its generalizability. First, the L2 used had grammatical structures similar to L1, limiting insights into structurally different languages. Second, the participant pool was skewed toward tech-savvy individuals. Third, the small sample size and controlled lab settings limit ecological validity. Due to the sample size, no statistically significant effects were observed for long-term recall. Additionally, due to our scoping, the study did not explore context-dependent AR presentations (e.g., 3D-anchored content), and mobile multitasking was not studied independently from walking.

Future work should validate these findings through longitudinal studies involving more diverse populations, broader language sets, and various AR smart glasses in real-world environments. This includes testing under different conditions (e.g., lighting, distractions) to better assess robustness and generalizability. A power analysis

using the current results as priors could help determine the minimum required sample size for future studies. Additionally, further research is needed to examine whether *Progressive* presentation can be effectively extended to other device platforms in multitasking contexts.

Acknowledgments

This research is supported by the National Research Foundation Singapore and DSO National Laboratories under the AI Singapore Programme (Award Number: AISG2-RP-2020-016). The CityU Start-up Grant (No. 9610677) also provides partial support. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not reflect the views of the National Research Foundation, Singapore. We extend our gratitude to all members of the Synteraction Lab and participants for their help in completing this project. We also thank anonymous reviewers for their valuable feedback.

References

[1] 1998. The Role of Illustrations in Text Comprehension: What, When, for Whom, and Why? In *The Construction of Mental Representations During Reading*, Susan R. Goldman and Herre van Oostendorp (Eds.). Psychology Press. https://psycnet.apa.org/record/1998-06740-008

[2] 2023. Introducing Duolingo Max, a learning experience powered by GPT-4. https://blog.duolingo.com/duolingo-max/

[3] 2024. CEFR descriptors search - Common European Framework of Reference for Languages (CEFR) - www.coe.int. https://www.coe.int/en/web/common-european-framework-reference-languages/cefr-descriptors-search

[4] Nuwan T. Attygalle, Matjaž Kljun, Aaron Quigley, Klen Čopič Pucihar, Jens Grubert, Verena Biener, Luis A. Leiva, Juri Yoneyama, Alice Toniolo, Angela Miguel, Hirokazu Kato, and Maheshya Weerasinghe. 2025. Text-to-Image Generation for Vocabulary Learning Using the Keyword Method. In *Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25)*. Association for Computing Machinery, New York, NY, USA, 1381–1397. doi:10.1145/3708359.3712073

[5] Ronald T. Azuma. 2019. The road to ubiquitous consumer augmented reality systems. 1, 1 (2019), 26–32. doi:10.1002/hbe2.113

[6] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. doi:10.1191/1478088706qp063oa

[7] Carrie J. Cai, Anji Ren, and Robert C. Miller. 2017. WaitSuite: Productive Use of Diverse Waiting Moments. *ACM Transactions on Computer-Human Interaction* 24, 1 (March 2017), 7:1–7:41. doi:10.1145/3044534

[8] Ronald Carter. 1998. *Vocabulary: Applied Linguistic Perspectives*. Psychology Press. Google-Books-ID: G_Rdpfzfy8kC.

[9] Qiaoyi Chen, Siyu Liu, Kaihui Huang, Xingbo Wang, Xiaojuan Ma, Junkai Zhu, and Zhenhui Peng. 2024. RetAssist: Facilitating Vocabulary Learners with Generative Images in Story Retelling Practices. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference (DIS '24)*. Association for Computing Machinery, New York, NY, USA, 2019–2036. doi:10.1145/3643834.3661581

[10] Xu Chen and Di Wu. 2024. Automatic Generation of Multimedia Teaching Materials Based on Generative AI: Taking Tang Poetry as an Example. *IEEE Transactions on Learning Technologies* 17 (2024), 1353–1366. doi:10.1109/TLT.2024.3378279 Conference Name: IEEE Transactions on Learning Technologies.

[11] Kai-Yi Chin, Huai-Ling Chang, and Ching-Sheng Wang. 2024. Applying a wearable MR-based mobile learning system on museum learning activities for university students. *Interactive Learning Environments* 32, 9 (Oct. 2024), 5744–5765. doi:10.1080/10494820.2023.2228843 Publisher: Routledge _eprint: https://doi.org/10.1080/10494820.2023.2228843.

[12] Brita Curum and Kavi Kumar Khedo. 2021. Cognitive load management in mobile learning systems: principles and theories. *Journal of Computers in Education* 8, 1 (March 2021), 109–136. doi:10.1007/s40692-020-00173-6

[13] Tilman Dingler, Dominik Weber, Martin Pielot, Jennifer Cooper, Chung-Cheng Chang, and Niels Henze. 2017. Language learning on-the-go: opportune moments and design of mobile microlearning sessions. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '17* (Vienna, Austria, 2017). ACM Press, 1–12. doi:10.1145/3098279.3098565

[14] Darren Edge, Stephen Fitchett, Michael Whitney, and James Landay. 2012. MemReflex: adaptive flashcards for mobile microlearning. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services (MobileHCI '12)*. Association for Computing Machinery, New York, NY, USA, 431–440. doi:10.1145/2371574.2371641

[15] Darren Edge, Elly Searle, Kevin Chiu, Jing Zhao, and James A. Landay. 2011. MicroMandarin: mobile language learning in context. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Vancouver BC Canada, 3169–3178. doi:10.1145/1978942.1979413

[16] Cassondra M. Eng, Karrie E. Godwin, and Anna V. Fisher. 2020. Keep it simple: streamlining book illustrations improves attention and comprehension in beginning readers. *npj Science of Learning* 5, 1 (Sept. 2020), 1–10. doi:10.1038/s41539-020-00073-5 Number: 1 Publisher: Nature Publishing Group.

[17] Alix Goguey, Carl Gutwin, Zhe Chen, Pang Suwanaposee, and Andy Cockburn. 2021. Interaction Pace and User Preferences. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3411764.3445772

[18] Yongqi Gu and Robert Keith Johnson. 1996. Vocabulary Learning Strategies and Language Learning Outcomes. *Language Learning* 46, 4 (Dec. 1996), 643–679. doi:10.1111/j.1467-1770.1996.tb01355.x 00000.

[19] Adam Ibrahim, Brandon Huynh, Jonathan Downey, Tobias Hollerer, Dorothy Chun, and John O'Donovan. 2018. ARbis Pictus: A Study of Vocabulary Learning with Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* 24, 11 (Nov. 2018), 2867–2874. doi:10.1109/TVCG.2018.2868568

[20] Yuta Itoh, Tobias Langlotz, Jonathan Sutton, and Alexander Plopski. 2021. Towards Indistinguishable Augmented Reality: A Survey on Optical See-through Head-mounted Displays. *Comput. Surveys* 54, 6 (July 2021), 120:1–120:36. doi:10.1145/3453157

[21] Nuwan Janaka, Xinke Wu, Shan Zhang, Shengdong Zhao, and Petr Slovak. 2022. Visual Behaviors and Mobile Information Acquisition. doi:10.48550/arXiv.2202.02748

[22] Nuwan Janaka, Shengdong Zhao, and Samantha Chan. 2023. NotiFade: Minimizing OHMD Notification Distractions Using Fading. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, 1–9. doi:10.1145/3544549.3585784

[23] Nuwan Janaka, Shengdong Zhao, and Shardul Sapkota. 2023. Can Icons Outperform Text? Understanding the Role of Pictograms in OHMD Notifications. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 575, 23 pages. doi:10.1145/3544548.3580891

[24] Emmanuel Keuleers and Marc Brysbaert. 2010. Wuggy: A multilingual pseudoword generator. *Behavior Research Methods* 42, 3 (Aug. 2010), 627–633. doi:10.3758/BRM.42.3.627

[25] David R. Krathwohl. 2002. A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice* 41, 4 (Nov. 2002), 212–218. doi:10.1207/s15430421tip4104_2 06999.

[26] Nilli Lavie, Aleksandra Hirst, Jan W. de Fockert, and Essi Viding. 2004. Load Theory of Selective Attention and Cognitive Control. *Journal of Experimental Psychology: General* 133, 3 (2004), 339–354. doi:10.1037/0096-3445.133.3.339 Place: US Publisher: American Psychological Association.

[27] Joanne Leong, Pat Pataranutaporn, Valdemar Danry, Florian Perteneder, Yaoli Mao, and Pattie Maes. 2024. Putting Things into Context: Generative AI-Enabled Context Personalization for Vocabulary Learning Improves Learning Motivation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3613904.3642393

[28] Jimmie Leppink, Fred Paas, Cees P. M. Van der Vleuten, Tamara Van Gog, and Jeroen J. G. Van Merriënboer. 2013. Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods* 45, 4 (Dec. 2013), 1058–1072. doi:10.3758/s13428-013-0334-1

[29] Manuela Macedonia and Thomas R. Knösche. 2011. Body in Mind: How Gestures Empower Foreign Language Learning. *Mind, Brain, and Education* 5, 4 (2011), 196–211. doi:10.1111/j.1751-228X.2011.01129.x

[30] Richard E. Mayer. 2002. Multimedia learning. In *Psychology of Learning and Motivation*. Vol. 41. Academic Press, 85–139. doi:10.1016/S0079-7421(02)80005-6

[31] Richard E. Mayer. 2014. Cognitive Theory of Multimedia Learning. In *The Cambridge Handbook of Multimedia Learning* (2 ed.), Richard E. Mayer (Ed.). Cambridge University Press, Cambridge, 43–71. doi:10.1017/CBO9781139547369.005

[32] Eunil Park, Jinyoung Han, Ki Joon Kim, Yongwoo Cho, and Angel P. del Pobil. 2018. Effects of Screen Size in Mobile Learning Over Time. In *Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication (IMCOM '18)*. Association for Computing Machinery, New York, NY, USA, 1–5. doi:10.1145/3164541.3164625

[33] Angela Ralli. 1991. Review of Vocabulary: Applied Linguistic Perspectives. *Machine Translation* 6, 1 (1991), 51–54. https://www.jstor.org/stable/40006901 Publisher: Springer.

[34] Ashwin Ram and Shengdong Zhao. 2021. LSVP: Towards Effective On-the-go Video Learning Using Optical Head-Mounted Displays. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 1 (March 2021), 30:1–30:27. doi:10.1145/3448118

[35] Philipp A. Rauschnabel, Reto Felix, Chris Hinsch, Hamza Shahab, and Florian Alt. 2022. What is XR? Towards a Framework for Augmented and Virtual Reality. *Computers in Human Behavior* 133 (Aug. 2022). doi:10.1016/j.chb.2022.107289
[36] Mrigank S Shail. 2019. Using Micro-learning on Mobile Applications to Increase Knowledge Retention and Work Performance: A Review of Literature. *Cureus* 11, 8 (2019), e5307. doi:10.7759/cureus.5307
[37] Smitha Sheshadri, Shengdong Zhao, Yang Chen, and Morten Fjeld. 2020. Learn with Haptics: Improving Vocabulary Recall with Free-form Digital Annotation on Touchscreen Mobiles. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376272
[38] Felicia Fang-Yi Tan, Peisen Xu, Ashwin Ram, Wei Zhen Suen, Shengdong Zhao, Yun Huang, and Christophe Hurter. 2024. AudioXtend: Assisted Reality Visual Accompaniments for Audiobook Storytelling During Everyday Routine Tasks. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–22. doi:10.1145/3613904.3642514
[39] Mark Feng Teng. 2023. The effectiveness of multimedia input on vocabulary learning and retention. *Innovation in Language Learning and Teaching* 17, 3 (May 2023), 738–754. doi:10.1080/17501229.2022.2131791
[40] Susannah Torcasio and John Sweller. 2010. The use of illustrations when learning to read: A cognitive load theory approach. *Applied Cognitive Psychology* 24, 5 (2010), 659–672. doi:10.1002/acp.1577
[41] Maheshya Weerasinghe, Verena Biener, Jens Grubert, Aaron Quigley, Alice Toniolo, Klen Copic Pucihar, and Matjaz Kljun. 2022. VocabulARy: Learning Vocabulary in AR Supported by Keyword Visualisations. *IEEE Transactions on Visualization and Computer Graphics* 28, 11 (Nov. 2022), 3748–3758. doi:10.1109/TVCG.2022.3203116 Publisher: IEEE Computer Society.
[42] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. doi:10.1145/1978942.1978963
[43] Yue Zhao, Tarmo Robal, Christoph Lofi, and Claudia Hauff. 2018. Stationary vs. Non-stationary Mobile Learning in MOOCs. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization - UMAP '18*. ACM Press, Singapore, Singapore, 299–303. doi:10.1145/3213586.3225241
[44] Chen Zhou, Katherine Fennedy, Felicia Fang-Yi Tan, Shengdong Zhao, and Yurui Shao. 2023. Not All Spacings are Created Equal: The Effect of Text Spacings in On-the-go Reading Using Optical See-Through Head-Mounted Displays. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–19. doi:10.1145/3544548.3581430

A Common Setting

A.1 Scoring

Following previous work [19, 37, 41] and our objectives, for each correct recall of an individual L2 word, 1 point was awarded. Variations in the plurality of nouns or verbs, verb tenses, and minor typos in the provided L1 words did not result in a loss of marks (e.g., if the correct L1 word is 'reads' and the given L1 word is 'read', it was considered correct). A similar scheme was applied to sentence scoring, where each correct identification of an L1 sentence resulted in either 2 or 3 points, depending on the number of unique L2 words in the sentence. Partial marks were awarded if the participant recognized the meaning of L2 words and placed them in the correct location within the sentence. The scoring did not consider pronouns, articles, prepositions, and common verbs (i.e., is/are) that appeared in more than one sentence. For example, if the correct L1 sentence is 'He reads a book', but the participant provided 'He x book', 1 point was awarded. If the participant provided 'The book', 0 points were given. If the participant provided, 'He read the book', the full score of 2 points was awarded. For each condition, the total score was calculated by summing the individual scores of words/sentences.

A.2 Analysis

A one-way repeated measures ANOVA (for a single factor with multiple levels) or a factorial ANOVA (for two factors with multiple levels) was used to analyze the quantitative data. When the assumptions of ANOVA were violated, the Friedman test or factorial repeated measures ANOVA after Aligned Rank Transform (ART [42]) was employed. The normality of the data was tested using the Shapiro-Wilk test, and sphericity was tested using Mauchly's test. Paired-sample t-tests, ART contrasts[2], or Wilcoxon signed-rank tests were used as post-hoc tests, with Bonferroni correction applied for multiple comparisons.

B *Study 1*

Table 3 and Figure 3 indicate the performance of the participants (N=12).

C *Pilot 4*

C.1 Materials

We selected Norwegian (L2) because its grammar system is straightforward for English speakers to learn, but its vocabulary, derived from Old Norse and influenced by German, is largely distinct from English[3] (see Table 1 for examples). This choice was made to avoid confounding factors such as cognates. In cases where close cognates existed, those sentences were excluded from the study (e.g., L1: "*book*" vs L2: "*bok*", L1: "*ball*" vs L2: "*ball*").

Audio pronunciation for the L2 texts was generated using the Google Cloud TextToSpeech API (language: 'nb-NO', voice: NEUTRAL, speaking rate: 0.9, format: .mp3) and was manually verified for accuracy.

C.2 Design and Procedure

The study design was the same as the *Study 1* (Sec 8). The procedure was similar to the common setting (Sec 4.7), with the addition of a pre-test to assess participants' familiarity with the selected L2 words. All participants scored 0 on the pre-test, allowing the post-test scores to be used directly to measure recall accuracy.

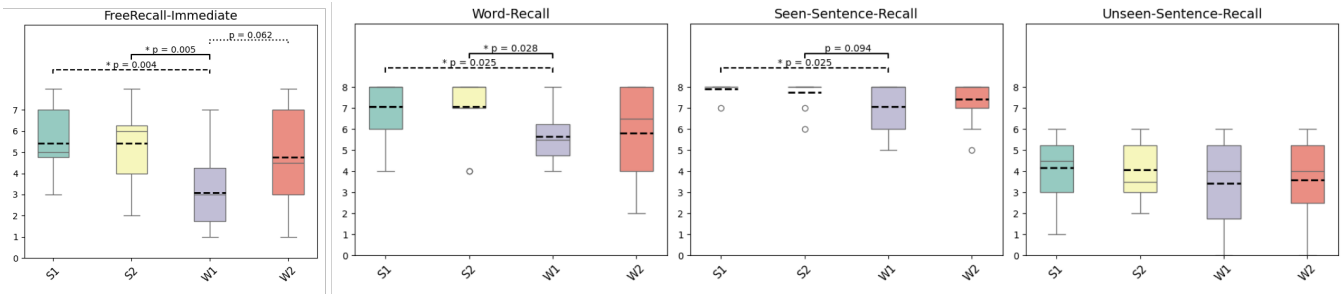
C.3 Results

Table 4 and Figure 4 indicate the performance of the participants (N=4).

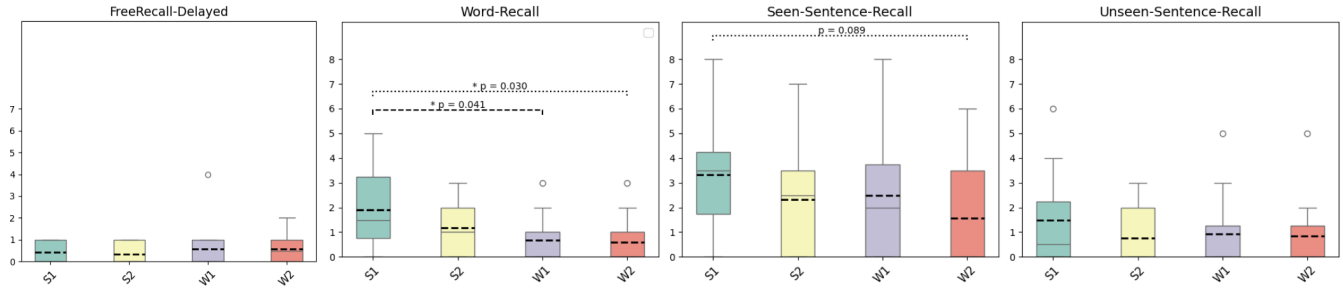
[2]Using art.con(), https://cran.r-project.org/web/packages/ARTool/vignettes/art-contrasts.html
[3]https://thelanguages.com/norwegian/grammar-rules-compared-to-english/

Table 3: Performance and User Ratings with 12 participants in *Study 1*. Here, S = *Sitting*, W = *Walking*, 1 = *NoGap*, and 2 = *Gap*. For example, S1 represents *Sitting* with *NoGap*. The orange color highlight corresponds to the best (mean) performance for *Sitting* while the green color is for *Walking*.

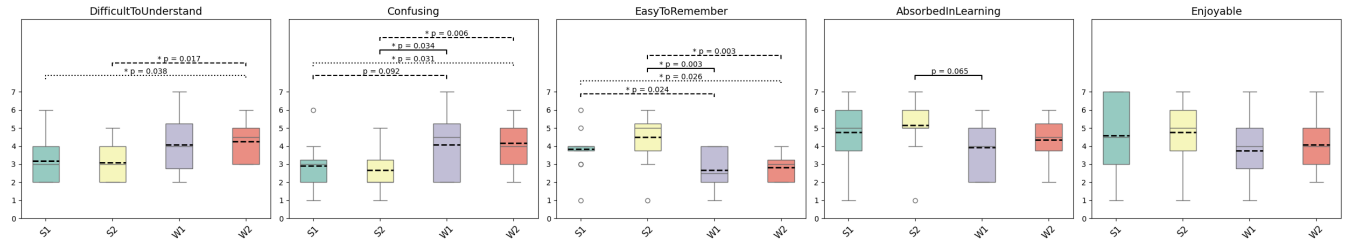
Condition	S1		S2		W1		W2	
Measure	M	SD	M	SD	M	SD	M	SD
Immediate Recall Measures								
Free-Recall	5.417	1.621	5.417	1.782	3.083	1.881	4.750	2.261
Word-Recall	7.083	1.443	7.083	1.505	5.667	1.435	5.833	2.368
Seen-Sentence-Recall	7.917	0.289	7.750	0.622	7.083	1.165	7.417	0.996
Unseen-Sentence-Recall	4.167	1.642	4.083	1.443	3.417	2.234	3.583	2.234
Delayed Recall Measures								
Free-Recall	0.417	0.515	0.333	0.492	0.583	1.165	0.583	0.669
Word-Recall	1.917	1.730	1.167	1.193	0.667	0.985	0.583	0.996
Seen-Sentence-Recall	3.333	2.387	2.333	2.425	2.500	2.812	1.583	2.429
Unseen-Sentence-Recall	1.500	1.977	0.750	1.138	0.917	1.621	0.833	1.528
User Ratings								
DifficultToUnderstand	3.167	1.337	3.083	1.165	4.083	1.730	4.250	1.055
Confusing	2.917	1.379	2.667	1.155	4.083	1.832	4.167	1.267
EasyToRemember	3.833	1.193	4.500	1.508	2.667	1.155	2.833	0.835
AbsorbedInLearning	4.750	1.712	5.167	1.586	3.917	1.564	4.333	1.435
Enjoyable	4.583	2.109	4.750	1.913	3.750	1.765	4.083	1.443



(a) Immediate *Recall* scores. The maximum score for each measure is 8, except for *Unseen-Sentence-Recall*, which has a maximum of 6.



(b) Delayed (7-day) *Recall* scores. The maximum score for each measure is 8, except for *Unseen-Sentence-Recall*, which has a maximum of 6.

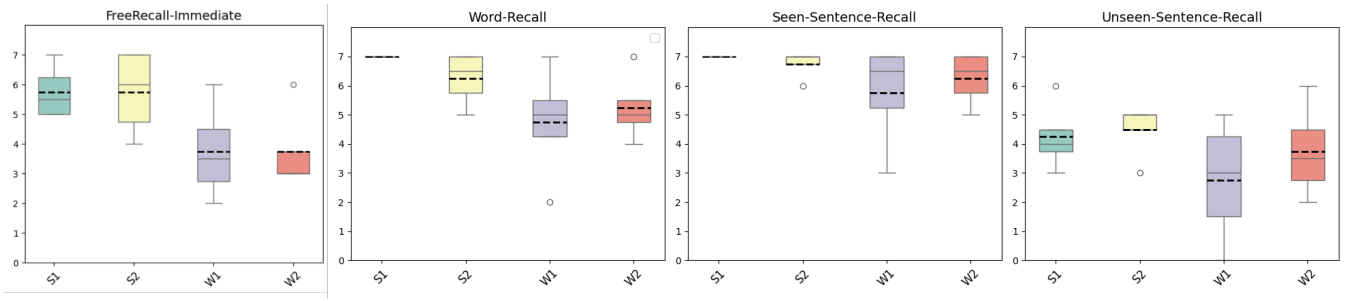


(c) Subjective ratings on a 7-point Likert scale, where 1 = Strongly Disagree and 7 = Strongly Agree.

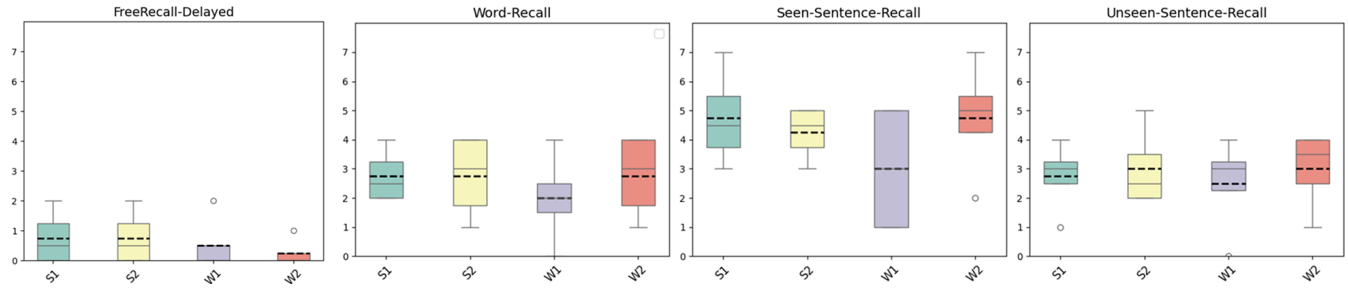
Figure 3: Measures on recall and perception with 12 participants in *Study 1*. Here, S = *Sitting*, W = *Walking*, 1 = *NoGap*, and 2 = *Gap*. For example, S1 represents, *Sitting* with *NoGap*. Dashed lines inside the box plots indicate the mean values.

Table 4: Performance and User Ratings with 4 participants in *Pilot 4*. Here, S = *Sitting*, W = *Walking*, 1 = *NoGap*, and 2 = *Gap*. For example, S1 represents, *Sitting* with *NoGap*. The orange color highlight corresponds to the best (mean) performance for *Sitting* while the green color is for *Walking*.

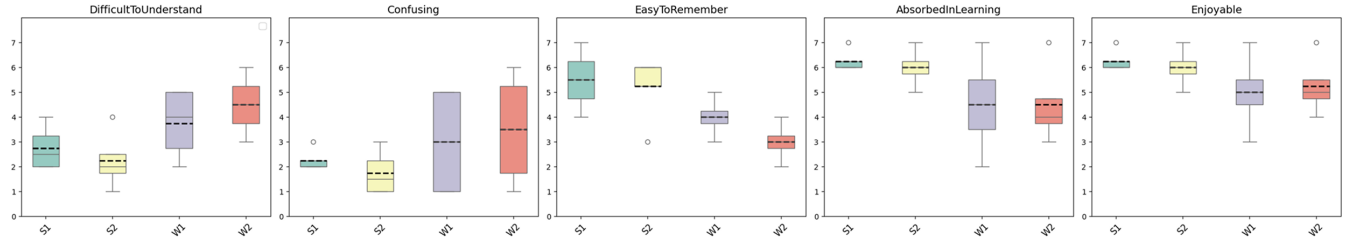
Condition	S1		S2		W1		W2	
Measure	M	SD	M	SD	M	SD	M	SD
Immediate Recall Measures								
Free-Recall	5.750	0.957	5.750	1.500	3.750	1.708	3.750	1.500
Word-Recall	7.000	0.000	6.250	0.957	4.750	2.062	5.250	1.258
Seen-Sentence-Recall	7.000	0.000	6.750	0.500	5.750	1.893	6.250	0.957
Unseen-Sentence-Recall	4.250	1.258	4.500	1.000	2.750	2.217	3.750	1.708
Delayed Recall Measures								
Free-Recall	0.750	0.957	0.750	0.957	0.500	1.000	0.250	0.500
Word-Recall	2.750	0.957	2.750	1.500	2.000	1.633	2.750	1.500
Seen-Sentence-Recall	4.750	1.708	4.250	0.957	3.000	2.309	4.750	2.062
Unseen-Sentence-Recall	2.750	1.258	3.000	1.414	2.500	1.732	3.000	1.414
User Ratings								
DifficultToUnderstand	2.750	0.957	2.250	1.258	3.750	1.500	4.500	1.291
Confusing	2.250	0.500	1.750	0.957	3.000	2.309	3.500	2.380
EasyToRemember	5.500	1.291	5.250	1.500	4.000	0.816	3.000	0.816
AbsorbedInLearning	6.250	0.500	6.000	0.816	4.500	2.082	4.500	1.732
Enjoyable	6.250	0.500	6.000	0.816	5.000	1.633	5.250	1.258



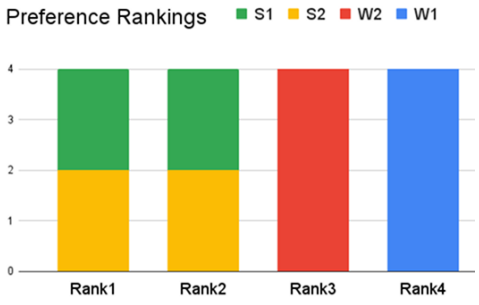
(a) Immediate *Recall* scores. The maximum score for each measure is 7, except for *Unseen-Sentence-Recall*, which has a maximum of 6.



(b) Delayed (7-day) *Recall* scores. The maximum score for each measure is 8, except for *Unseen-Sentence-Recall*, which has a maximum of 6.



(c) Subjective ratings on a 7-point Likert scale, where 1 = Strongly Disagree and 7 = Strongly Agree.



(d) Preference ranking.

Figure 4: Measures on recall and perception with 4 participants in *Pilot 4*. Here, S = *Sitting*, W = *Walking*, 1 = *NoGap*, and 2 = *Gap*. For example, S1 represents, *Sitting* with *NoGap*. Dashed lines inside the box plots indicate the mean values.