LangEye: Toward ‘Anytime’ Learner-Driven Vocabulary Learning From Real-World Objects

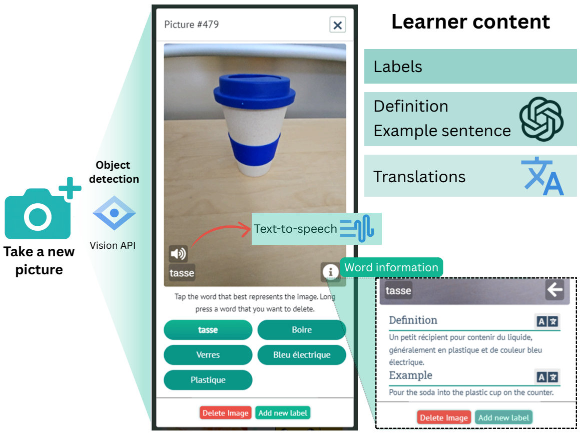
Mariana Shimabukuro, Deval Panchal, and Christopher Collins
Ontario Tech University, Oshawa, ON, Canada
{mariana.shimabukuro, christopher.collins}@ontariotechu.ca

Abstract

We present LangEye, a mobile application for contextual vocabulary learning that combines learner-curated content with generative NLP. Learners use their smartphone camera to capture real-world objects and create personalized “memories” enriched with definitions, example sentences, and pronunciations generated via object recognition, large language models, and machine translation. LangEye features a three-phase review system — progressing from picture recognition to sentence completion and free recall. In a one-week exploratory study with 20 French (L2) learners, the learner-curated group reported higher engagement and motivation than those using pre-curated materials. Participants valued the app’s personalization and contextual relevance. This study highlights the potential of integrating generative NLP with situated, learner-driven interaction. We identify design opportunities for adaptive review difficulty, improved content generation, and better support for language-specific features. LangEye points toward scalable, personalized vocabulary learning grounded in real-world contexts.

1 Introduction

Creating contextual learning opportunities remains a major challenge in second language (L2) acquisition, particularly for learners situated in non-native environments. Immersive experiences, such as studying abroad or participating in language-rich communities, are often inaccessible due to financial, geographic, or logistical barriers (Galloway and Ruegg, 2020). Mobile-Assisted Language Learning (MALL) addresses this by leveraging the ubiquity and portability of smartphones to support “anytime” micro-learning and situated learning approaches (Arakawa et al., 2022; Byrne, 2019; Tran et al., 2023). Yet, many current MALL systems provide limited flexibility in adapting dynamically to learners’ immediate context, personal interests, and cognitive availability, and often rely exclusively on pre-curated, static content.



Tap the word that best represents the image. Long press a word that you want to delete.

Un petit récipient pour contenir du liquide, généralement en plastique et de couleur bleu électrique.

Pour the soda into the plastic cup on the counter.

Figure 1: System diagram of LangEye, illustrating the flow from learner image capture through API-based vocabulary enrichment. The application integrates Google Cloud Vision, Cloud Translation, and Text-to-Speech APIs, along with OpenAI’s GPT model, to generate personalized vocabulary “memories” enriched with labels, definitions, translations, and pronunciation.

We introduce LangEye[1], a mobile application for vocabulary learning that turns real-world objects into interactive “memories” through a learner-curated workflow. Using smartphone cameras and NLP services — including computer vision, large language models, and machine translation — LangEye generates personalized lexical entries with definitions, example sentences, and pronunciation. Learners engage with this content through a structured review system that supports progressive recall and production in the target language. An overview of the system architecture and API integration is shown in Figure 1.

Designed specifically to empower self-directed learners, LangEye supports short, personalized learning interactions directly tied to learners’ physi-

[1] Public demo and repository to be made available at https://vialab.ca/langeye.

cal environments and motivations. Crucially, all review sessions are initiated by learners and grounded in their uniquely captured contexts, promoting deeper personalization, engagement, and learner autonomy. However, due to this highly personalized and learner-curated design, traditional standardized assessments of vocabulary learning outcomes — such as standardized pre- and post-tests — are challenging, as vocabulary items vary greatly across individuals.

To explore the feasibility, learner acceptance, and design implications of LangEye, we conducted an initial one-week exploratory study with 20 French L2 learners, comparing a *camera group* (using LangEye to generate personalized vocabulary entries) and a *control group* (using pre-curated vocabulary). Preliminary findings highlight the motivational and engagement benefits of integrating learner-curated AI-generated content, while also revealing limitations associated with computer vision accuracy and AI-generated contextual sentences. These insights lay the groundwork for our planned longitudinal evaluation, which will rigorously measure personalized vocabulary acquisition and retention over extended use periods. Additionally, future iterations of LangEye will incorporate advanced object detection methods (e.g., YOLO-E) and more dynamic, interactive scenarios such as gamified object treasure hunts, further enhancing contextual vocabulary learning through data-driven methods.

2 Background and Related Work

2.1 Learner-curated Vocabulary with MALL Applications

Compared to Computer-Assisted Language Learning (CALL), MALL excels in accessibility and context-driven learning, making it effective for vocabulary acquisition (Alhuwaydi, 2022; Klimova, 2021). Micro-learning involves short, targeted activities (e.g., 5–15 minutes) (Leong et al., 2020). For instance, MiniHongo (Tran et al., 2023) integrates location and activity data to deliver contextual vocabulary lessons, demonstrating the efficacy of location-relevant micro-learning. Similarly, VocaBura (Hautasaari et al., 2019) utilizes audio and location-based prompts to teach vocabulary during real-world interactions.

VocabEncounter (Arakawa et al., 2022), a CALL application, applies contextual and micro-learning by integrating target vocabulary into web content via natural language processing (NLP) and machine translation (MT) techniques. Comparable techniques embed vocabulary into audiovisual content through automatic glossing and lexical simplification (Alm, 2021; Fievez et al., 2023). VocabNomad (Tsourounis and Demmans Epp, 2016) provides a highly personalized MALL experience with progress tracking, contextual recommendations, and learner-curated entries. By allowing users to add vocabulary, record pronunciations, and browse visual collections, it fosters situated and personalized learning.

These studies highlight the importance of integrating relevant and contextual vocabulary learning into daily life, leveraging micro-learning and personalized approaches. However, most rely on static content or fixed corpora, with limited opportunities for learners to drive content creation based on their immediate environment.

2.2 AI-Enabled Context Personalization for Vocabulary Learning

The advent of large language models (LLMs), beginning with ChatGPT[2], has enabled more dynamic natural language generation, allowing for real-time synthesis of definitions, example sentences, and explanations. While generative AI presents challenges such as ethical concerns and content accuracy (Campolo and Crawford, 2020), it has opened new possibilities in personalized educational applications, particularly in language learning.

Applications like Storyfier (Peng et al., 2023) leverage generative AI to create vocabulary-rich narratives based on learner input. Although they showed limited learning gains, users appreciated the contextualization and narrative integration. Similarly, Leong et al. (2024) found that AI-generated personalized prompts enhanced learner motivation, despite modest measurable gains in vocabulary retention.

Recent systems also incorporate generative AI into mixed-reality environments. WordSense (Vazquez et al., 2017) pioneered contextual vocabulary learning through object recognition linked to dynamically generated content. More recently, FluencyAR (Hollingworth and Willett, 2023) integrated augmented reality (AR) with generative feedback for self-talk, and CuriosityXR (Vaze et al., 2024) allowed educators to create multi-modal, contextual mini-lessons. These

[2] https://openai.com/research/overview. Accessed April 2025

works emphasize engagement and curiosity, often powered by NLP-driven interfaces.

LangEye extends MALL, integrating micro-, situated-, and contextual-learning with modern NLP technologies, including object recognition, large language models for content generation, machine translation, and text-to-speech. Unlike prior systems that personalize content using static corpora or predefined curricula, LangEye allows learners to initiate the content pipeline through real-world object interactions, enabling highly contextualized and self-directed vocabulary acquisition. This learner-driven approach aims to promote both personalization and autonomy, but it also challenges traditional evaluation methods, as vocabulary exposure varies widely across individuals. As such, LangEye raises important questions around how to evaluate open-ended, NLP-enhanced learning systems, where learner agency and environmental context shape the learning trajectory.

3 LangEye Design: Create and Review Memories

LangEye's core interaction is structured around learner-generated *memories* — vocabulary entries tied to real-world images captured by the learner. These memories are enriched using NLP services to provide multilingual definitions, contextual sentences, and audio pronunciation, supporting both vocabulary learning and retention. In this case, the vocabulary items are associated with the pictures taken by the learners. Figure 2 illustrates the learner taking a picture of a cup (*tasse* in French) and interacting with the generated memory's word definition and example sentence. Therefore, the learned words are tied to a familiar object, which is more effective when compared to unfamiliar or no pictures for vocabulary learning (Hwang et al., 2014; Kang, 1995; Saidbakhramovna et al., 2021).

The app creates situational learning opportunities by allowing the learner to interact with objects around them in three ways: (1) take pictures of objects which they can interact with in-situ via editing or exploring the picture (*memory*); (2) take pictures of objects now, but choose to edit or explore the picture (*memory*) later; and (3) start a *Review Memories* session with a desired length — 3 to N words for up to 3 phases per session. Figure 2 (c, d, and e) shows examples of review activities for each phase.

3.1 Creating Memories

In *picture mode*, the learner can aim their device's camera at an object they wish to interact with in the target language (TL) and take a picture of it. As shown in Figure 2, in response, the app returns a list of five likely labels (names) for the object, sorted by probability (most likely object name as the top result). The learner can explore the definition and a sample sentence for each label, and optionally select a different top label based on that information. Alternatively, they can confirm the default suggestion. On the same screen, learners also have the option to listen to the pronunciation of the object name. At this point, they may choose to take another picture to explore more objects or end their study session. The app saves all objects, pictures, and names to the learner's *memories* for further review.

These definitions and sample sentences are dynamically generated using OpenAI's GPT-3.5 API and then translated into the TL via Google Translate.

Prompt template. LangEye uses a standardized prompt designed for beginner learners to generate consistent and level-appropriate definitions and examples:

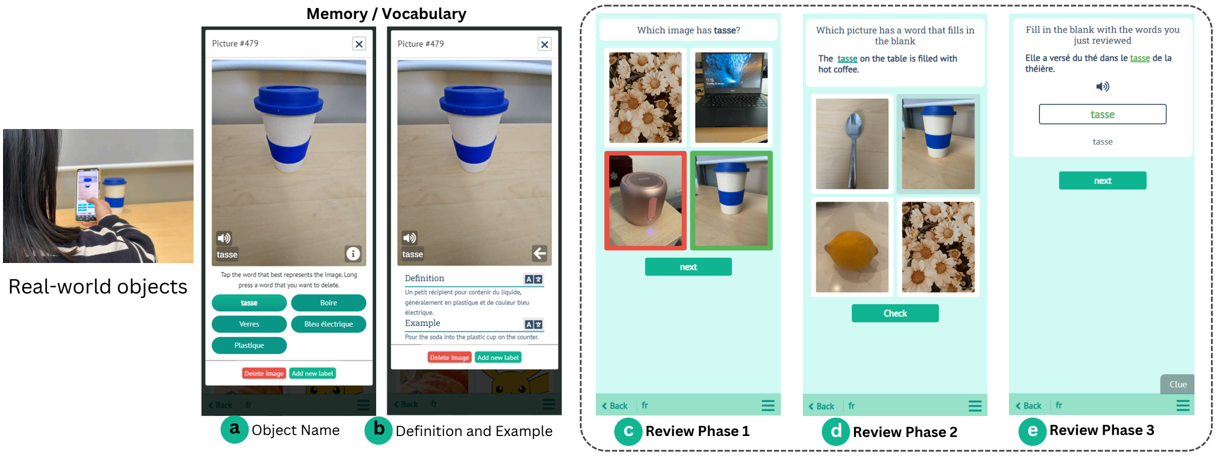
> You are a language tutor for beginner French learners. Given the following word: `<object-label>`, provide: (1) a clear, beginner-friendly English definition of the object, and (2) a short example sentence using the object in everyday context. Make both responses simple and age-appropriate for learners at A1–A2 level.

See Table 1 for an output example: **`cup`** → ***tasse***.

Table 1: Example of a vocabulary memory generated for `cup`.

Label	`cup`
Definition	A cup is a small container used for drinking.
Sentence	She drank tea from her favourite cup.
French	*Une tasse est un petit récipient utilisé pour boire.* *Elle a bu du thé dans sa tasse préférée.*

This structure supports flexible and learner-driven study sessions, allowing learners to create and engage with memories at their own pace, based on what they encounter in their daily environments. By combining visual, textual, and auditory modalities, LangEye supports deeper encoding of vocabulary through multiple channels of reinforcement. This two-stage generation pipeline balances personalization and control: GPT-3.5 generates beginner-



Tap the word that best represents the image. Long press a word that you want to delete.

Un petit récipient pour contenir du liquide, généralement en plastique et de couleur bleu électrique.

Pour the soda into the plastic cup on the counter.

Which image has **tasse**?

Which picture has a word that fills in the blank

The tasse on the table is filled with hot coffee.

Fill in the blank with the words you just reviewed

Elle a versé du thé dans le tasse de la théière.

Figure 2: Memories are created from real-world objects. (a) A memory contains the name of the detected object (cup: *tasse* in French) and its pronunciation. (b) Additional information, including a definition and a sample sentence, can be displayed. This content is dynamically generated using OpenAI's GPT-3.5 API and translated into the learner's target language (here, French) via machine translation. Learners can practice their vocabulary through the *Memory Review* structure, which includes three progressively challenging phases: (c) **Phase 1: Picture Recognition** prompts learners to identify the correct image; feedback is immediate, highlighting the chosen image in green if correct or red if incorrect. (d) **Phase 2: Sentence Completion** requires selecting an image to fill in a blank within a sentence presented in French. (e) **Phase 3: Free Recall** displays a sentence in French and prompts learners to type the object's name; small typos are accepted, and hints are available.

friendly English definitions and sentences, while Google Translate handles multilingual output. This modular design supports better quality control, simplifies debugging, and ensures broader language support, particularly for low-resource languages where LLMs may struggle with robust translation performance. At the time of system development, GPT-3.5 was the most stable and accessible option for generating consistent content.

Editing labels. To mitigate possible computer vision and translation errors, learners can add their own label to the picture using the *add label* button — see Figure 2 (a, b). This allows learners to input a text label they find more appropriate if the app's suggestions are insufficient. Learners type the label in their source language, and the app provides the TL translation — eliminating the need to know the TL term. These custom labels appear alongside their generated definition and sentence — see Figure 2 (b). If a label is deemed irrelevant, learners may long-press it to delete it. In the example shown in Figure 2 (a), the label *"Bleu Électrique"* (Electric Blue) refers to the colour of the cup; the learner might find it unrelated and choose to remove it.

3.2 Languages Supported

LangEye currently supports the following source or target languages: English, French, Spanish, and Portuguese. The *source language* (SL) is the language the learner is familiar with or learning from, and the *target language* (TL) is the language to be learned. The interface elements, such as the menu and activity instructions, can be set to any of the listed languages as the SLs. Likewise, the names of objects and learning content (i.e., definitions and example sentences) can be displayed in any listed language as the TL. Learners can define their SL and TL in the *Settings* menu option. This is enabled using machine translation.

3.3 Reviewing Memories

This feature is a classic quiz-style review of the collected memories (vocabulary words). The *Memory Review* has three phases or types of quiz questions, each increasing in difficulty and reducing support. These phases are (1) Picture Recognition, (2) Sentence Completion, and (3) Free Recall. The system chooses N memories (words/objects) to review during a Memory Review session. For each phase, learners are tested on those same N words. Learners must complete earlier phases to unlock later ones, but they may choose to end the session between phases. This progressive design aims to gradually increase cognitive load and promote long-term retention by reinforcing vocabulary through multiple retrieval formats.

Phase 1: Picture Recognition prompts learners to identify an object by selecting the correct picture from four gallery options (Figure 2 (c)). Feedback is immediate, with correct choices highlighted in green and incorrect in red. Learners have up to four attempts per word, ensuring they review all three target words before moving to the next phase.

In **Phase 2: Sentence Completion**, learners answer "Fill in the blank" questions by selecting a picture that completes an example sentence (Figure 2 (d)). The chosen picture's word fills the blank, allowing reflection before submission. Incorrect answers reveal the correct choice, and sentences are shown in the SL to support beginners. However, this design may limit advanced learners who prefer tasks entirely in the TL.

Phase 3: Free Recall introduces open-ended vocabulary production. Learners type the names of the three target words without visual cues (Figure 2 (e)). A *Clue* button provides definitions if needed, and the system tolerates minor typos, while still highlighting the correct spelling for feedback. Sentences are now in the TL, catering to advanced learners and promoting grammar understanding.

This phased approach bridges the gap between beginner and advanced learners, enabling gradual mastery of vocabulary and TL proficiency. See Table 2 for a feature and language comparison of all three phases.

3.4 Tracking Vocabulary Learning Progress

The *Achievements* feature in LangEye tracks learners' Memory Review history and accuracy. For Phase 1, it records the average number of guesses, while for Phases 2 and 3, it calculates accuracy. Learners can sort words by accuracy, with TL initials (e.g., "fr" for French) displayed for context. While these metrics provide insight into learner behaviour and memory usage, they do not directly measure vocabulary acquisition or retention — a challenge we revisit in our discussion of evaluation.

4 User Study

To explore LangEye's potential as a personalized vocabulary learning tool, we conducted a one-week exploratory study with 20 French (L2) learners. This formative evaluation investigated how learner-curated content and NLP-driven interactions support engagement and vocabulary study in real-world contexts. The study was reviewed and approved by our institution's Research Ethics Board.

Participants were randomly assigned to two groups:

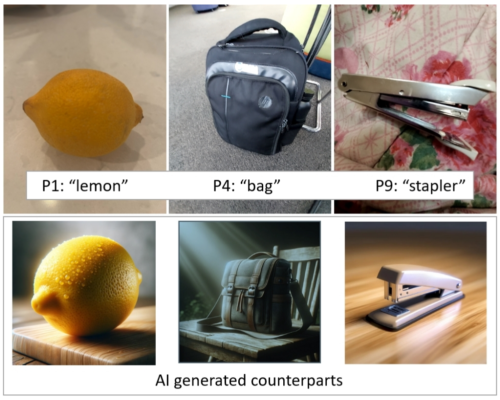


Figure 3: Sample images taken by participants and their AI-generated counterparts used in the study for the control group.

Control group ($N = 10$)**:** used a version of the app with pre-defined vocabulary and AI-generated content based on pre-curated images; and **Camera group** ($N = 10$)**:** used the full app, including features for taking and uploading images and dynamically generating content through integrated NLP services.

Study Design. The study comprised two sessions: **Session 1:** in-lab training, background survey, and post-session usability feedback. **Session 2:** online exit interview after using the app for at least five days. **Between Sessions:** participants were instructed to use the app daily for five days, completing short usability surveys after each use. Reminder emails were sent daily. Each 50-minute session included surveys, session recordings, and app usage data. Photos taken between study sessions were also collected for the camera group. See Appendix A for detailed information on the study sessions and materials; the semi-structured interview questions are included in the Supp. Material. Figure 3 shows a sample of the generated images used by the control group participants.

Recruitment. Participants were 20 French learners, evenly split into control and camera groups. Participants received $10 CAD for Session 1 and $20 CAD for Session 2, recruited through posters in high-traffic campus areas.

Room Setup. The room held eight household objects (an apple, cup, fork, paper, scissors, spoon, sunglasses, and watch) for the camera group to explore and photograph. The control group experienced the same room with objects, but they interacted exclusively with pre-curated memories.

Control Group Memory Curation. Control group memories were curated using data from the cam-

Table 2: Feature and language support comparison across the three Memory Review phases: **Phase 1: Picture Recognition**, **Phase 2: Sentence Completion**, and **Phase 3: Free Recall**. As learners progress through the phases, visual support (e.g., images and multiple-choice options) is gradually reduced, while the use of the target language (TL) increases. This design makes later phases more cognitively demanding. Learners can choose to save and exit the review session at any point between phases.

		Phase 1	**Phase 2**	**Phase 3**
Task type	Identify picture		—	—
	Fill in the blank	—		
Answer support	Picture			—
	Multiple choice			—
	Type/Spell	—	—	
	Clue	—	—	Word definition
Answer and corrective feedback	Instantaneous check		—	—
	Submit and check	—		
	Corrective feedback			
Cognitive task	Word recall	Picture	Picture	Spell
	Word collocation	—		
Language *Source: Source Language* *TL: Target Language*	UI elements	Source	Source	Source
	Vocabulary word	TL	TL	TL
	Sample sentences	—	Source	TL
	Word definition	—	—	Source

era group to ensure comparability between groups. Camera group participants collected an average of 11 pictures ($Median = 9; min = 5; max = 19$), resulting in 24 unique objects.

To ensure uniformity and preserve privacy, we generated realistic images of the objects using OpenAI's Dall-E 3 (full list in Supp. Material). The curated vocabulary ensured consistency without introducing additional biases or privacy concerns. This initial study focused on usability, motivation, and learner perceptions, rather than directly measuring vocabulary acquisition, which is addressed in planned longitudinal follow-ups.

4.1 Study Results: Engagement and Usability

This section presents the user study results, including the pre-session, post-session, exit interview, and software usage data. *Qualitative data analysis.* The questionnaire and the semi-structured interview open-ended questions were coded into categories following commonly mentioned themes in the participants' answers.

4.1.1 Pre-session background questionnaire

Participants Background. Participants (N=20) were aged 18–24 (N=15) and 25–30 (N=5). All were fluent in English, though only 6 identified it as their first language. Most participants (N=15) self-reported as beginners (A1/A2), with 9 in the camera group. The control group included most advanced learners (B1/B2; N=4), who reported studying French for over a year. See Appendix B for details. Participants spoke diverse languages, including Tamil (N=5), Hindi (N=4), and Urdu (N=3). Additional languages learned alongside French included Arabic (N=2), Spanish, and Italian, while 11 participants were not learning another language.

Technology for Language Learning. Duolingo was the most commonly used app (N=12), followed by platforms like Udemy and Memrise. Eight participants, mostly in the control group (N=5), reported not using any apps. Only one camera group participant used apps daily, with most others engaging less frequently (once or 3–6 times a week). Mobile devices were the preferred learning platform (N=14), followed by desktop (N=5) and a single participant choosing "either." Preferences were evenly distributed across groups.

Language Learning Goals. The primary motivation for learning French was career-related (N=11; 7 camera, 4 control), followed by leisure (N=4) and travel (N=3). Other reasons included academic and family goals (each N=1).

4.1.2 Post-Session 1 Feedback and Exit Interview Results

Usability. Ratings for Memories, Review Memories, Achievements, and Picture Mode were measured on a 10-point scale. Participants rated their experience with LangEye after Session 1 (first im-

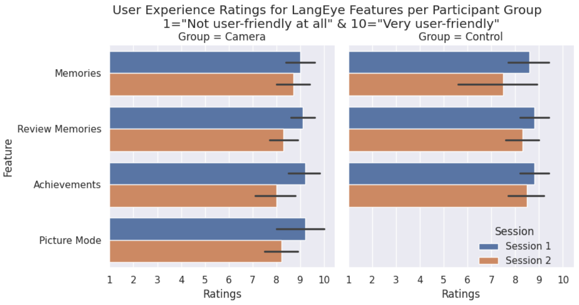


Figure 4: Using a 10-point user experience scale where (1) "Not user-friendly at all" and (10) "Very user-friendly" participants provided ratings on each of four LangEye features: *Memories*, *Review Memories*, *Achievements*, and *Picture Mode* — for camera group only. The chart on the left represents the camera groups and the control group is shown on the right. Overall, Session 1 had higher ratings than Session 2, and the camera group had higher ratings than the control group.

pressions) and Session 2 (after five days of use). Review Memories yielded an average of 8.7 for the camera group and 8.6 for the control group. Session 1 scores were higher for both groups compared to Session 2. Overall, ratings dropped from Session 1 to 2 for all features/groups (Figure 4), with the Memories (9.0 to 8.7 and 8.6 to 7.5) and Achievements (9.2 to 8 and 8.8 to 8.5) features showing the most decline for the camera and control groups, respectively. Picture Mode ratings, available only for the camera group, averaged 8.7, with Session 1 scoring 9.2 and Session 2 scoring 8.2.

Comfort and Control. Participants expressed comfort using LangEye for both vocabulary review and learning, based on 5-point Likert scale ratings. The camera group (4.3 and 4.2) reported similar comfort levels for both activities, while the control group showed lower comfort (4.0 and 4.2) when learning new vocabulary, likely due to their pre-curated and limited vocabulary set. Four control group participants requested options to expand pre-curated content. Self-efficacy ratings over learning were consistent across groups, with both reporting a mean score of 4.3 and a range of 3—5.

Most and Least Favourite Features. The camera group's most liked features were Picture Mode (4), Review Memories (3), and Memories (3). Participants appreciated the personalization offered by Picture Mode: *"[It] allows real-time learning with objects around me"* (P8, A2). The control group preferred Review Memories (8), with Phase 1: Picture Recognition praised for its simplicity. The least liked features for the camera group included Phase 3: Free Recall (3), Phase 2: Sentence Completion (2), and Achievements (2), with some noting confusing sentences in Phases 2 and 3. The control group disliked Achievements (4), citing low interactivity.

Motivation. Self-reported motivation levels, however, showed divergence. The camera group maintained a steady mean of 4.3 across sessions, while the control group's mean dropped from 4.1 in Session 1 to 3.8 in Session 2. This decline may reflect lower engagement with pre-curated content. These findings underscore the benefits of learner-curated content in enhancing comfort, control, and sustained motivation. They also suggest that giving learners agency to drive the content creation process—supported by generative NLP — can foster deeper engagement compared to static, pre-defined content.

Learner Perceptions Compared to Other Tools. *Learning with Pictures.* Participants valued LangEye's use of pictures for vocabulary learning, citing improved memorization and contextual association compared to dictionaries: *"Images make it easier to memorize and associate vocab with objects"* (P15, B1). *Self-curated Memories.* Camera group participants praised the personalization and relevance of self-curated content: *"[LangEye] uses my own pictures, making vocabulary more memorable"* (P3, A1). They suggested combining pre-populated and user-generated content for flexibility. *Multiple Labels per Image.* LangEye's ability to associate multiple concepts with a single image was seen as helpful for intermediate learners but confusing for beginners: *"Pictures have more context and words, good for intermediate learners"* (P14, A1).

4.2 Thematic Learner Feedback

App reminders and gamification. While some participants appreciated the absence of in-app reminders (P12), others requested daily notifications to encourage engagement (P1, P2). *Aesthetics improvements.* Participants (9/20) recommended a more colourful interface, sound effects for feedback, and larger buttons for easier interaction.

Content customization. Participants valued the use of personalized images, citing improved memory and relevance. A camera group participant noted, *"This app adds personal attachment to the picture, making it easier to remember"* (P8).

AI-generated content and robustness. Participants reported object detection errors and overly technical definitions. Cluttered backgrounds and multiple objects caused incorrect labels, while some sentences had mismatched vocabulary contexts. For instance, "wood" (noun) was replaced with "wooden" (adjective). Participants suggested cropping tools and improved AI prompts to reduce errors.

Review memories design. Beginners (A1) preferred Phase 1: Picture Recognition and Phase 2: Sentence Completion but found Phase 3: Free Recall overwhelming, while advanced learners preferred Phase 3 in the target language (TL). Participants generally praised the multi-phase system for its progressive difficulty. P9 said, *"I like the three phases, but the second phase in English was not helpful due to gender issues."*

Language-specific considerations. Participants highlighted issues with gendered nouns in French during Phase 2: Sentence Completion. Gender information was lost in English translations, causing confusion, especially for A2–B2 learners. Suggestions included displaying gender indicators (e.g., P1, P8, P9).

4.3 Daily Usage and Technical Issues

Figure 5 visualizes the decline in daily feedback form submissions across the five study days. The control group maintained more stable participation, while the camera group showed a sharper drop-off, despite initially similar engagement levels. This suggests that while learner-curated content may drive early motivation, maintaining sustained engagement over time remains a challenge. Overall ratings for ease of use, engagement, and vocabulary learning were mostly positive, leaning toward "Strongly agree" or "Neutral." However, control group ratings for learning new words were lower, likely due to limited pre-curated vocabulary. *Technical Issues* 26% of submissions reported technical difficulties, including delays in label loading over mobile networks and incorrect object labels. Sentence quality was another concern, as participants noted that some sentences were contextually incorrect or mismatched vocabulary. The detailed data is available in B.1.

5 Discussion

This exploratory study demonstrates the promise and challenges of combining learner-curated content with generative AI to support vocabulary learning in mobile contexts. While our goal was not to directly measure vocabulary gains, the findings offer formative insights into learner experience, system usability, and the design trade-offs inherent to NLP-enhanced educational tools. Below, we reflect on key lessons learned and identify opportunities for future improvement.

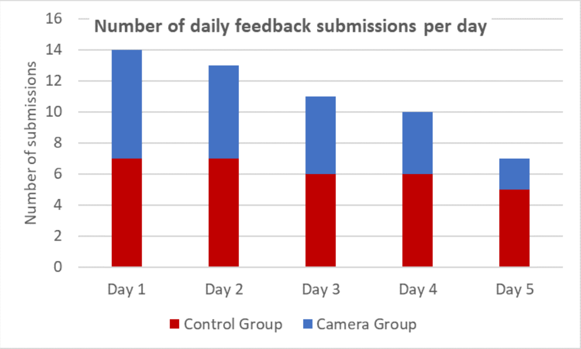


Figure 5: Number of daily feedback form submissions across the five-day study. While total submissions began at 14 on Day 1, they declined to 7 by Day 5. The control group's submissions remained relatively stable (7 to 5), whereas the camera group showed a sharper decline (7 to 2). Participant ratings per day are available in B.1.

Evaluation and Research Implications. Future work will incorporate longitudinal vocabulary tracking and explore adaptive evaluation strategies aligned with learner-curated content. Because LangEye supports open-ended, learner-defined content creation, traditional pre- and post-testing are difficult to apply consistently. Even usage-based metrics, such as phase completions or accuracy scores, are complicated by the variability in content difficulty and prior learner knowledge. These challenges reflect broader tensions in evaluating personalized, generative learning systems and call for alternative strategies such as learner modelling or adaptive diagnostics.

Balancing Pre- and Self-Curated Content. Learner-curated vocabulary fosters autonomy and engagement but also introduces variability in vocabulary scope and difficulty. A hybrid approach—integrating structured, pre-curated content alongside learner-generated memories—may better support novice learners while preserving personalization for advanced users. This balance also enhances scalability across languages without requiring expert-authored corpora.

Tensions in AI-Generated Content. Although generative AI enabled dynamic and personalized vocabulary entries, participants frequently encountered issues such as overly technical definitions, inappropriate word senses, and context mismatches.

For example, “wood” was rendered as “wooden” (Table 3), and cluttered images led to irrelevant labels. These issues reflect broader limitations of prompt-based generation in educational contexts. Future iterations will incorporate prompt tuning, simpler output targets, and human-in-the-loop validation to improve robustness and learner alignment.

Table 3: Example of a vocabulary memory with a word sense mismatch due to ambiguous object labelling.

Label	wood
Definition	Wooden means made of wood.
Sentence	I sat on the wooden chair.
Issue	Learner expected a noun definition for wood, but GPT returned the adjective form wooden.

Language-Specific Considerations. LangEye supports multiple target languages via machine translation; however, users of gendered languages (e.g., French) have noted grammatical issues, particularly in Phase 2, where translations often lack gender agreement. This suggests the need for grammar-aware translation strategies and visual indicators for noun gender, especially in beginner-focused review phases.

Our use of a hybrid generation pipeline, employing GPT-3.5 for English definitions and Google Translate for multilingual output, was driven by a need for modularity, consistency, and broad language coverage. This approach provided control over linguistic complexity in the initial prompt while leveraging production-grade translation tools for low-resource languages, where LLM performance remains less benchmarked. This modular architecture proved essential for supporting LangEye’s multilingual scope but also contributed to mismatches and errors in translated content, underscoring the importance of future refinement in prompt tuning and translation alignment.

Learner Engagement and Personalization. Learners consistently emphasized the motivational value of interacting with vocabulary grounded in their own environment. This supports situated learning theory and highlights how self-curated images can improve recall by reinforcing personal relevance. However, engagement declined over time—particularly in the camera group—suggesting a need for better pacing, reminders, or gamified retention mechanisms to sustain interest.

Review System Calibration. Participants appreciated the phased review design, but feedback suggests the need for difficulty calibration. Phase 3: Free Recall was overwhelming for beginners, while some advanced learners desired more TL immersion earlier. Dynamically adapting review complexity based on learner level and behaviour (e.g., accuracy, completion history) may improve retention and reduce frustration.

Toward Context-Aware Learning Scenarios. LangEye’s current design centers on object-driven vocabulary. Future iterations could support more dynamic interactions, such as context-aware prompts, adaptive content sequencing, and gamified activities (e.g., real-world “treasure hunts”). These enhancements—combined with more accurate object detection (e.g., YOLO-E)—could transform LangEye into a broader platform for situated, task-based language learning.

6 Conclusion

This paper presented LangEye, a mobile language learning application that leverages generative NLP and learner-curated content to support contextual vocabulary acquisition. By combining object recognition, machine translation, and dynamic content generation, LangEye enables self-directed learners to engage in personalized, real-world language practice. Findings from our exploratory study highlight the system’s usability, motivational benefits, and learner preference for personalized visual content.

While this formative evaluation did not assess vocabulary acquisition directly, the results inform design implications for learner-driven, AI-enhanced educational tools. Future work will include longitudinal studies to track learning outcomes, and expand LangEye’s capabilities through adaptive review difficulty and improved language-specific support. Additionally, we envision incorporating more accurate computer vision models (e.g., YOLO-E) to enable dynamic, context-aware interactions such as real-world object “treasure hunts” or live situational vocabulary tasks, further bridging the gap between everyday experiences and language learning.

Limitations

This work has several limitations that inform the scope of its findings and highlight directions for future research.

First, this was an exploratory and short-term study focused on learner engagement and usability. While participants interacted with generative NLP features and learner-curated content, we did not directly assess vocabulary acquisition or retention through pre- and post-testing. Future studies with longer durations and individualized baseline assessments are necessary to evaluate learning outcomes rigorously.

Second, the evaluation was constrained by the personalized nature of the learner-curated content. Since learners selected their own vocabulary items, it was not feasible to apply a standardized test or compare vocabulary gains across participants. While this personalization is central to LangEye's design, it introduces challenges for controlled, quantitative evaluation.

Third, the generative NLP components (e.g., definitions, sample sentences) sometimes produced inconsistent or overly complex outputs. This was especially problematic for beginner learners, who occasionally found definitions too advanced or mismatched in word sense. Our system relies on prompt-based content generation, which can be brittle without careful tuning and contextual awareness. While we did not run expert benchmarking of the AI-generated content in this pilot, this remains an important step for future work, especially for language education applications."

Finally, although LangEye supports multiple languages, our study only examined English–French learners. Language-specific features—such as grammatical gender—presented challenges in the translation pipeline and feedback design, limiting generalizability across linguistic contexts. Further studies should explore broader language pairs and adapt the system to handle grammar-sensitive features more effectively.

Ethical Considerations

This study was reviewed and approved by our institution's Research Ethics Board (REB). All participants provided informed consent prior to participation and were compensated for their time. Data collected during the study, including app usage logs and participant feedback, was anonymized prior to analysis.

To protect participant privacy, especially in the camera group, no personally identifying photos were stored or analyzed. For the control group, object images were generated using OpenAI's DALL·E 3 to avoid the use of participant-provided media.

LangEye integrates generative AI tools (e.g., GPT-3.5, Google Translate) to produce multilingual learning content. While this automation enables scalability, care was taken to limit content generation to isolated vocabulary contexts, and the system does not store user data beyond local app sessions. Limitations of AI output— such as occasional mismatches in word sense — were disclosed to participants, and learners had full control over which content to save and review.

Acknowledgments

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

Abdulhameed A Alhuwaydi. 2022. A review on vocabulary learning-designed mall applications in the efl context. *Theory and Practice in Language Studies*, 12(10):2191–2200.

Antonie Alm. 2021. Language learning with netflix: extending out-of-class l2 viewing. In *2021 International Conference on Advanced Learning Technologies (ICALT)*, pages 260–262. IEEE.

Riku Arakawa, Hiromu Yakura, and Sosuke Kobayashi. 2022. Vocabencounter: Nmt-powered vocabulary learning by presenting computer-generated usages of foreign words into users' daily lives. In *Proceedings of the 2022 CHI conference on human factors in computing systems*, pages 1–21.

Jason Byrne. 2019. Anytime autonomous english mall app engagement. *International Journal of Emerging Technologies in Learning (iJET)*, 14(18):145–163.

Alexander Campolo and Kate Crawford. 2020. Enchanted determinism: Power without responsibility in artificial intelligence. *Engaging Science, Technology, and Society*.

Isabeau Fievez, Maribel Montero Perez, Frederik Cornillie, and Piet Desmet. 2023. Promoting incidental vocabulary learning through watching a french netflix series with glossed captions. *Computer Assisted Language Learning*, 36(1-2):26–51.

Nicola Galloway and Rachael Ruegg. 2020. The provision of student support on english medium instruction programmes in japan and china. *Journal of English for Academic Purposes*, 45:100846.

Ari Hautasaari, Takeo Hamada, Kuntaro Ishiyama, and Shogo Fukushima. 2019. Vocabura: A method for supporting second language vocabulary learning while walking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(4):1–23.

Shanna Li Ching Hollingworth and Wesley Willett. 2023. Fluencyar: Augmented reality language immersion. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–3.

Wu-Yuin Hwang, Holly SL Chen, Rustam Shadiev, Ray Yueh-Min Huang, and Chia-Yu Chen. 2014. Improving english as a foreign language writing in elementary schools using mobile devices in familiar situational contexts. *Computer assisted language learning*, 27(5):359–378.

Sook-Hi Kang. 1995. The effects of a context-embedded approach to second-language vocabulary learning. *System*, 23(1):43–55.

Blanka Klimova. 2021. Evaluating impact of mobile applications on efl university learners' vocabulary learning–a review study. *Procedia Computer Science*, 184:859–864.

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, New York, NY, USA. Association for Computing Machinery.

Kelvin Leong, Anna Sung, David Au, and Claire Blanchard. 2020. A review of the trend of microlearning. *Journal of Work-Applied Management*, 13(1):88–102.

Zhenhui Peng, Xingbo Wang, Qiushi Han, Junkai Zhu, Xiaojuan Ma, and Huamin Qu. 2023. Storyfier: Exploring vocabulary learning support with text generation models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–16.

Azizova Fotimakhon Saidbakhramovna, Rahmatova Nargiza Valijonvna, and Kurbanbayeva Dilnoza Sharofidinovna. 2021. The method of educating vocabulary in a foreign language or target language. *Linguistics and Culture Review*, 5(S1):1649–1658.

Nguyen Tran, Shogo Kajimura, and Yu Shibuya. 2023. Location-and physical-activity-based application for japanese vocabulary acquisition for non-japanese speakers. *Multimodal Technologies and Interaction*, 7(3):29.

Stephen Tsourounis and C Demmans Epp. 2016. Learning dashboards and gamification in mall: Design guidelines in practice. *The international handbook of mobile-assisted language learning*, pages 370–398.

Aaditya Vaze, Alexis Morris, and Ian Clarke. 2024. Curiosityxr: Context-aware education experiences with mixed reality and conversation ai. In *2024 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR)*, pages 41–49. IEEE.

Christian David Vazquez, Afika Ayanda Nyati, Alexander Luh, Megan Fu, Takako Aikawa, and Pattie Maes. 2017. Serendipitous language learning in mixed reality. In *Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems*, pages 2172–2179.

A Research Methods

A.1 Control Group Memory Curation

One of the main challenges in comparing the two group of participants is to curate the control group's memories. As discussed in section 4, our approach to this problem was to run both session of the study with the camera group first. This allowed us to use the collection of memories created by that camera group as the control group's memories. On average, the camera group collected 11 pictures ($Median = 9; min = 5; max = 19$). The aggregation of duplicates of the camera group memories resulted in a total of 24 objects (memories) listed below:

apple
bag
cup
dishware
drinkware
eyewear
floor
flooring
food
fork
glasses
hand
lemon
paper
peripheral
scissors
serveware
slipper
speaker
spoon
stapler
tableware
watch
wood

Recalling that, to create uniformity in the images, avoid bias toward the quality of images taken by the camera group, and preserve the participants' privacy, OpenAI's Dall-E 3 was used to create the images for each of these memories. The prompt included the object name and instructions for the illustration to be "realistic" to mimic a photo taken of the object — sample shown in Figure 3. The pairs of all AI-generated images and labels can be found in Supplemental Materials. This approach was used to present a similar curation of vocabulary while not adding words that might not have been added using a smartphone camera.

A.2 Detailed Study Sessions

A.2.1 Study Session 1: Introducing LangEye

Session 1 was conducted in a controlled lab environment with separate setups for the camera and control groups.

Room Setup. The room featured a collection of eight household objects (see Figure 6) for the camera group to explore and photograph. The control group experienced the same room setup but interacted exclusively with pre-curated memories.

Pre-Session Questionnaire. Participants completed a brief background survey about their French language learning experience and use of language learning apps.

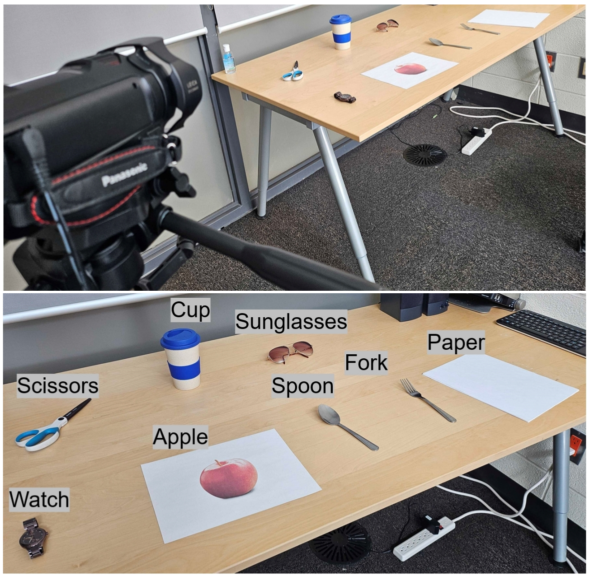


Figure 6: Top: Room setup with video recording. Bottom: Objects available for the camera group to explore and create memories.

Training Tasks. Participants were introduced to the app's features through a demonstration and a printed tutorial. Both groups explored the app's main features, with the control group focusing on editing pre-curated memories and the camera group using the camera mode to create their own. Participants could ask questions during the session and were required to interact with each feature before proceeding to the post-session survey.

Post-Session Questionnaire. Participants evaluated LangEye's usability and practicality, providing feedback on the app's usefulness for language learning.

A.2.2 Between Sessions

Participants were instructed to use LangEye daily for five days between Sessions 1 and 2. Daily reminder emails prompted them to complete a short feedback form covering usability, error reporting, and general app impressions. The second session was scheduled 5–10 days after the first.

A.2.3 Study Session 2: Exit Semi-Structured Interview

In Session 2, participants reflected on their experiences with LangEye, discussing usability, vocabulary acquisition, and the accuracy of object recognition and labeling. The camera group shared insights on creating memories, while the control group focused on pre-curated content. Interviews were recorded using Google Meet, capturing video, audio, and transcripts. Transcripts were reviewed

Table 4: Summary of participants' background information per study group: camera and control.

Attribute	Camera	Control	Total
Age			
18–24 years old	6	9	15
25–30 years old	4	1	5
L1			
Other	8	6	14
English	2	4	6
French level			
A1	5	5	10
A2	4	1	5
B1	–	4	4
B2	1	–	1

for accuracy and used alongside structured session notes for qualitative analysis of participant responses.

B Results Data Visualizations

Visual representations of some of the results are available in this section. Table 4 shows the tabulated participants demographics information. Table 5 shows the tabulated data on participants' French lerning background.

B.1 Daily Feedback Submissions

Participants were asked to submit a daily feedback form after using the app in between sessions. While the number of daily submissions (Figure 5) remained somewhat stable for the control group (from 7 to 5), the camera group had the most decline (from 7 to 2). When aggregating both groups, at Day 1 there were 14 submissions, which was reduced to 7 at Day 5. The charts in Figures 7 and 8 show the participants' ratings (5-Point Likert scale for agreement) per day. The difference in the volume of submissions makes it difficult to compare across groups, but overall, the ratings lean toward "Strongly agree" to "Neutral" throughout the study days. Here are the statement items participants were asked to rate:

- "Overall, this app is easy to use."
- "I'm having fun using this app."
- "I have learned new French words using this app."
- "I feel more in control of my French vocabulary learning progress and content since using this app."

Table 5: Table shows participants' main method for learning French and the duration of their studies. The study was run in Canada, a bilingual country (English and French are official languages). Thus, French immersion schools are commonly available in Canadian education. Courses for French ("Course at school") as a foreign language are also common in Anglophone schools. And other "French course" or classes are easily accessible in language institutes. Participants who indicated "None" were never enrolled in a course or followed a specific method.

French Study		Camera	Control	Total
Duration	**Method**			
1 week or less	None	3	1	4
less than 6 months	French immersion	–	1	1
	Online course or resource	1	–	1
	None	1	–	1
1 year+	Course at school	–	1	1
	Online course or resource	1	1	2
5 years+	Course at school	1	5	6
	French course	1	–	1
10 years+	French course	1	–	1
	French immersion	1	1	2

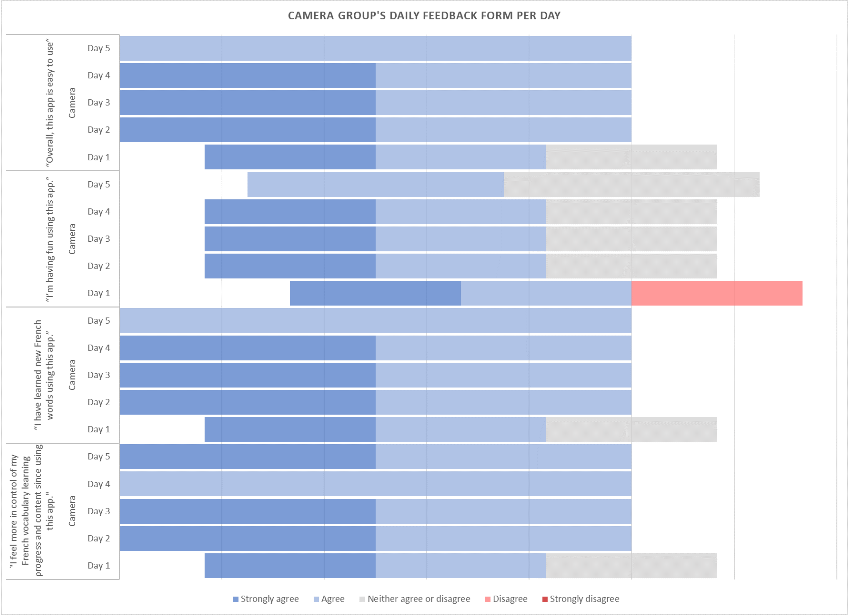


Figure 7: Camera group participants' ratings for the daily feedback form per day. The Figure 5 shows the number of responses per day. While at Day 1 there were 7 submissions, that number declines along the days. This chart shows the distribution of the respondents' answers to the 5-Point Likert scale agreement statement items at each day, from bottom (Day 1) to top (Day 5) at each item. Although there is a shift to "Strongly agree"/"Neutral" as days pass the number of responses are reduced.

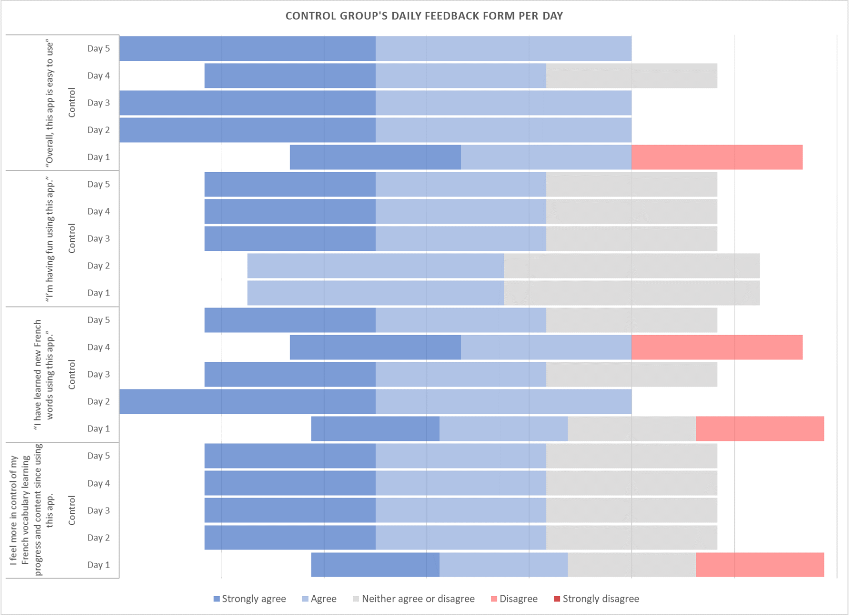


Figure 8: Control group participants' ratings for the daily feedback form per day. The Figure 5 shows the number of responses per day. While at days 1 and 2 there were 7 submissions, that number declines to 5 at Day 5; which is higher than the camera group's Day ($N = 2$). This chart shows the distribution of the respondents' answers to the 5-Point Likert scale agreement statement items at each day, from bottom (Day 1) to top (Day 5) at each item. Although there is a shift to "Strongly agree"/"Neutral" as days pass the number of responses are reduced.