*informatics*

MDPI

*Article*

# A Context-Based Multimedia Vocabulary Learning System for Mobile Users

**Andrew Vargo [1,*], Kohei Yamaguchi [2], Motoi Iwata [1] and Koichi Kise [1]**

[1] Graduate School of Informatics, Osaka Metropolitan University, Sakai 599-8531, Japan; imotoi@omu.ac.jp (M.I.); kise@omu.ac.jp (K.K.)
[2] Graduate School of Engineering, Osaka Prefecture University, Sakai 599-8531, Japan; aug10_t-spirits0355@outlook.jp
* Correspondence: awv@omu.ac.jp

**Abstract:** Vocabulary acquisition and retention is an essential part of learning a foreign language and many learners use flashcard applications to repetitively increase vocabulary retention. However, it can be difficult for learners to remember new words and phrases without any context. In this paper, we propose a system that allows users to acquire new vocabulary with media which gives context to the words. Theoretically, this use of multimedia context should enable users to practice with interest and increased motivation, which has been shown to enhance the effects of contextual language learning. An experiment with 46 English as foreign language learners showed better retention after two weeks with the proposed system as compared to ordinary flashcards. However, the impact was not universally beneficial to all learners. An analysis of participant attributes that were gathered through surveys and questionnaires shows a link between personality and learning traits and affinity for learning with this system. This result indicates that the proposed system provides a significant advantage in vocabulary retention for some users, while other users should stay with traditional flashcard applications. The implications of this study indicate the need for the development of more personalized learning applications.

**Keywords:** context-based learning; language learning; mobile learning; vocabulary acquisition



**Citation:** Vargo, A.; Yamaguchi, K.; Iwata, M.; Kise, K. A Context-Based Multimedia Vocabulary Learning System for Mobile Users. *Informatics* **2024**, *11*, 1. https://doi.org/10.3390/informatics11010001

Academic Editor: Antony Bryant

Received: 14 October 2023
Revised: 11 December 2023
Accepted: 12 December 2023
Published: 19 December 2023



## 1. Introduction

The way learners engage with educational content has evolved due to the introduction of internet-connected devices. Mobile-ready e-learning allows for access to study and practice without a location or time constraints [1,2]. In the field of second language acquisition, flashcard applications are a dominant form of mobile study due to their ease of construction and straightforward usage scenario for the learners [3,4]. These digital flashcards are often improvements upon their analog counterparts in that they include performance metrics and steps to reduce the forgetting curve. However, a limitation of these types of flashcard applications is that they do not take context-based learning and incidental learning into account. Unknown words need to be pre-defined, rather than be discovered. Developing technologies that support such learning on mobile platforms is important since researchers have shown that diverse and informative contexts will enhance the effects of vocabulary acquisition [5,6]. Additionally, learners gain vocabulary incidentally through reading behavior like extensive reading [7], and such a learning strategy shows better performance compared to intentional learning [8]. In this paper, we develop and test a system where learners interact with video and caption content to discover and then practice learning unknown words. This allows for the creation of a flashcard system that can present vocabulary with context and encourage incidental learning.

Context-based learning allows learners to have better learning outcomes by giving them targeted learning material that is related to their interests. This enhances the learning effects of the material in terms of comprehension and motivation to study [9,10]. Previous

attempts at introducing context-based learning have been carried out on both the desktop format [11] and the mobile format [12]. These implementations of context-based learning are severely limited in the type of content that can be delivered to the learner, as they primarily focus on text-based learning. In this research, we developed a system which broadens the choices of the learning materials. Our system expands the source of the learning materials multimedia so that users can learn through various media such as video captions and lyrics in a fully mobile setting.

This work uses a smartphone application for several reasons. First, mobile learning environments are a growing modality of learning for college-age adults [13], especially since the start of the COVID-19 pandemic [14,15]. In addition, mobile learning allows for sporadic and unplanned learning sessions which can be engaged regardless of location [2]. This is especially beneficial for learning tasks where memorization is key, such as language learning, and is also important for users who may not have stable locations for study [16]. Finally, an application allows us to investigate the usage of the context-based system in a mobile setting and allows users to use the application wherever they want.

In this manuscript, we performed an experiment with 46 non-native English-speaking participants for four weeks to measure the English vocabulary memorization retention impact of the proposed system versus a traditional flash-card system. The goal of this experiment is to understand whether the context-based multimedia application is beneficial for all or some participants with the learning and retention of vocabulary. We found that the proposed system is effective in enhancing retention for many users. However, there are users for whom the proposed system is not advantageous. An analysis of user attributes indicates that there may be some ways to identify and suggest for whom context-based learning is appropriate. These results provide insight into the value of context-based systems for learning vocabulary, and highlight the need for personalization.

The main contributions of this work are as follows:

- We extend a vocabulary flashcard system to include multimedia context-based system features.
- We perform an analysis of a context-based vocabulary system which gives insight into how different learners benefit from such a system.

## 2. Related Work

In this work, we focus on the concept of being able to provide context-based learning in a mobile smartphone-based scenario. That is, learners are able to study whenever and wherever they want. Therefore, we look at two areas of related work: context-based learning for second language acquisition and mobile applications that use flashcards for the learning of vocabulary.

### *2.1. Context-Based Learning*

Previous work has discussed the effectiveness of context when learners acquire vocabulary. For example, Webb et al. [17] have revealed that context by itself does not affect vocabulary acquisition. Researchers have shown that informative sentences improve the power of vocabulary retention [5,18,19]. Additionally, learning from multiple sentences rather than a single sentence enhances the effectiveness of context-based learning [5,6].

There are many factors involved in improving the quality of context-based learning. For example, vocabulary is effectively acquired when learners interact with documents they are interested in [20,21]. In addition, video captions and lyrics are also effective for learning vocabulary [22,23]. Moreover, visual and auditory information help to comprehend contexts and retain vocabulary, which are different aspects of reading documents [24].

Researchers have also explored the combination of intentional learning and incidental learning [25,26]. For example, Laufer [27] revealed that reading itself leads to little gain in vocabulary mastery, but supplemental use of word-focused activity such as using a dictionary will bring vocabulary gain. Schmitt [28] has suggested that intentional learning

and incidental learning each have disadvantages and they compensate and positively require a combination of each other.

### *2.2. Mobile Learning Flashcard Systems for Vocabulary Acquisition*

Mobile devices are suitable for learning not only in terms of convenience but also in terms of learning effectiveness due to the ability to learn anywhere at any time [29]. Nikoopour et al. [30] revealed that the learning effects of mobile flashcards and paper flashcards are superior to that of flashcards on a desktop owing to their portability. This result indicates that the learning effect of flashcards cannot be improved simply by digitizing them but ensured by making them portable. Moreover, Nikoopour et al. also suggested mobile flashcards are the most active way to learn vocabulary. This impact will be more pronounced in the long term. Xodabande et al. [31] have revealed that the amount of vocabulary learned through a year was significantly higher with the mobile flashcards than with paper flashcards, and also found that acquisition of academic words was higher with mobile flashcards in both a controlled setting [32] and in a long term-self-directed learning scenario [33]. Zakian et al. [34] also found that a mobile flashcard system outperformed traditional word lists for long-term learning. The results suggest that mobile flashcards are superior to paper flashcards in terms of a chance to be exposed to more vocabulary.

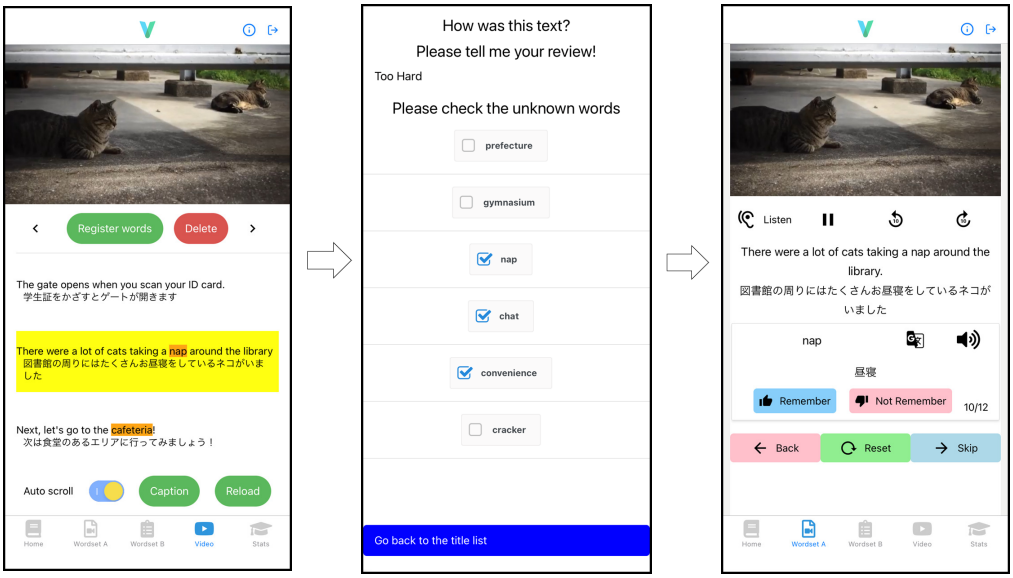
While previous literature indicates that context-based learning is effective, there is no concrete proof that context-based learning is always more effective than a simple repetitive drilling application in a mobile setting. Yamaguchi et al. [12] used a previous version of the system presented in this paper to look at the effectiveness of a context-based system with text-based news articles. In this case, words were selected as unknown from within text-based articles which the participants chose as interesting. The results of this study were limited in duration and sample size but showed that the method was effective overall. However, the power of the effectiveness dropped over a week and was not effective for all participants. This indicates that simple repetition may be more effective for some users when they study on their own.

## 3. Application Architecture

In order to evaluate context-based learning in a mobile setting, we extended a mobile application with context enabled content called the Vocabulometer based on previous web and mobile versions [11,12,35]. In particular, this work extends upon Yamaguchi et al.'s work [12] where a text-only context-based flashcard system was presented and shown to be effective. This application uses the main elements of a flashcard application for learning English and can have its contents personalized to a user's interests with different types of media.

### *3.1. Learning English Using the Proposed System*

Figure 1 shows how a user would enable and use the application. First, the user registers topics they are interested in and answers the vocabulary questions from the San Diego Quick Assessment [36] to initialize the content. The proposed system generates a personalized word list for the user based on their interests and results. After initialization, the user registers their preferred videos. When a user starts to play the video, captions and these translations will display as shown in Figure 1. In the video player screen, the unknown words are highlighted. After watching the videos, the user sends feedback on the subjective difficulty and answers a question form. The question form consists of the words the proposed system estimates as unknown words from the feedback and the user's word list as shown in Figure 1. Users are asked to check whether they already know the word or not. The words displayed on the question forms are stemmed and do not include stopwords. A user's word list is updated every time the question form is answered, thus improving the personalization of the proposed system every time the user watches a video.




(**a**) (**b**) (**c**)

**Figure 1.** The flow of learning with a video in the Proposed System. In (**a**), the yellow represents the audio that is currently being played with English and Japanese captions. The orange represents probable unknown words. In (**b**), the participant can then select unknown words from the predicted set. In (**c**), the participant can review the word in the context of the video. All Japanese text in the subfigures is accompanied by the English translation above it. (**a**) Video and Caption, (**b**) Feedback, (**c**) Review.

We use flashcards for the review system as shown in Figure 1. The flashcards consist of three elements: unknown word, embedded YouTube video, and video caption including the unknown word. The user can watch the part of a video that corresponds to the captions. Words labeled as memorized by the user are filtered out and the proposed system automatically displays words labeled as not memorized. The proposed system displays them again at an interval to encourage vocabulary retention. We apply the Leitner system [37] which is a traditional spaced repetition algorithm. The proposed system sends push notifications to the user when the flashcards are updated and thus motivates the user to use the application.

### *3.2. Learning Material*

The proposed system uses multimedia content for learning materials: documents, video captions, and song lyrics. For text content, we use documents from two sources: the Newsela [38] dataset and news articles. Newsela provides English articles for language learners with five reading levels per article. This allows learners to select the content that matches their skills. We categorized the Newsela articles into seven topics (entertainment, economy, environment, lifestyle, politics, sport, and science) to allow participants to choose interesting content more easily. We also use news articles that are available on Yahoo! News (https://news.yahoo.com/ (accessed on 10 October 2023)). The user can select a news article from it. When the user selects English articles from the Newsela dataset or registers the URL of the news article, the proposed system displays the estimated unknown words in highlights. For audio-visual content, we use videos that are available on Youtube (https://www.youtube.com/ (accessed on 10 October 2023)). When the user registers the URL of their preferred YouTube videos, the proposed system displays the embedded video, captions, and its translation with their estimated unknown words highlighted. Finally, for audio content, we use lyrics from songs available on Spotify (https://www.spotify.com/ (accessed on 10 October 2023)). The user can link their own Spotify account to the proposed system. When they successfully link to their account, the proposed system displays the

lyrics from the playlist they created with their estimated unknown words highlighted. The user can also seek the part he/she wants to watch by tapping the caption.

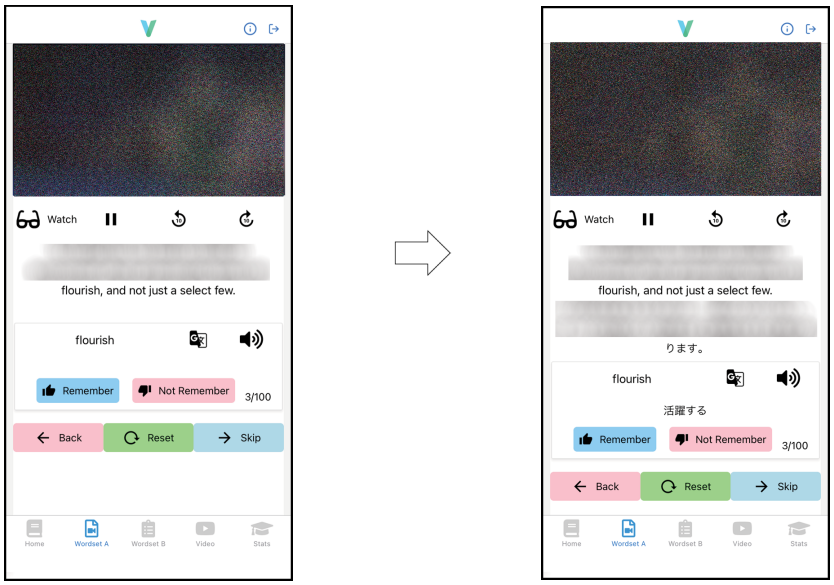
These multi-media functions in the application enable the user to enhance the effects of incidental vocabulary learning in terms of context diversity. Moreover, users acquire the benefits of using videos and lyrics as learning materials that they could not obtain by using text alone.

## 4. Experiment

In order to understand the effectiveness of the proposed system, we conducted an experiment to examine the effectiveness of the proposed system by having participants ($N = 46$, Male: $N = 33$, Female: $N = 13$) learn unknown English vocabulary with both the proposed system and a standard flashcard application. Specifically, we conduct the analysis with a within-subjects approach. Participants used their own smartphones for using the application and subsequent experiment (iOS: 31, Android: 15). 44 of the participants were Japanese university students with an age range of 18 to 25 years, and two participants were not university students (31 years old and 41 years old, respectively). Participants had a wide range of English abilities (TOEIC scores of participants: mean = 686.1 (min: 440, max: 955)) and while some participants studied English daily, some rarely studied English. All participants had exposure to studying English and there were no English-learning beginners. At the end of the study, participants received 1000 JPY per hour for a total sum of 11,000 JPY (approximately 75 USD) for their participation. The study received ethical approval from the Graduate School of Engineering at Osaka Prefecture University (Now: Osaka Metropolitan University) on 27 December 2021.

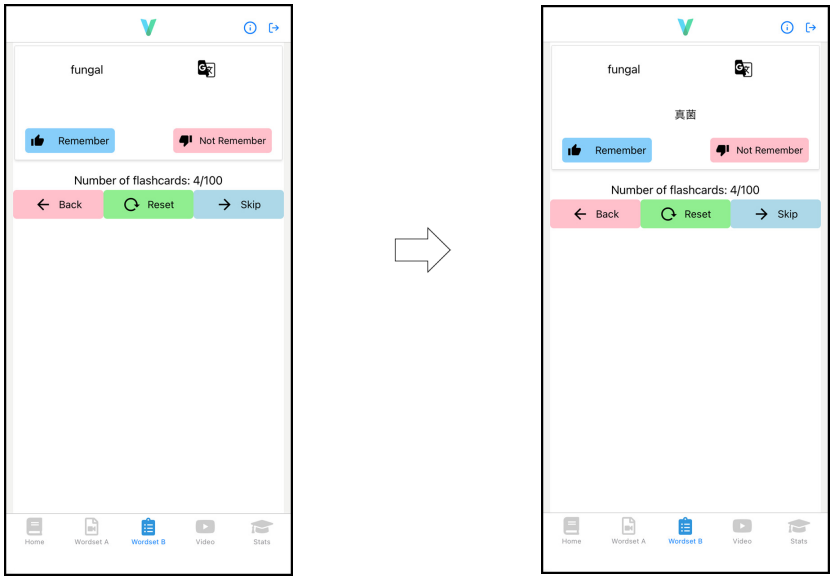
First, we randomly divided participants into two groups (Group X and Group Y). Next, participants selected videos they were interested in and registered them to the proposed system. They then picked out unknown words in each video with a questionnaire form provided by the proposed system. They repeated this process until they had collected 100 unknown words. We call these unknown words the Experimental wordset. We then selected another 100 words that were not used in the Experimental wordset and called them the Control wordset. The Control wordset is adjusted by its word frequency to that of the Experimental wordset to normalize the difficulty of both wordsets. The participants confirmed all of these words as unknown words. After that, the participants attempted to memorize the unknown words. The participants of Group X memorized the Experimental wordset using the proposed system. As shown in Figure 2, the unknown word, the video captions, and its corresponding part of the video are displayed first. Then, the participants check the translation by tapping the translation button. The participants of Group Y memorized the Control wordset using the flashcard function which displays only unknown words and the corresponding translations. As shown in Figure 3, the unknown word is displayed first, then the participants check the translation by tapping the translation button. The participants memorized these 100 words for three days. They were asked to review until they thought they had memorized them "by heart" each day. However, it should be noted that this element was left for the participants themselves to determine.

We conducted three post-tests at intervals of 1, 7, and 14 days after the last day of the memorization part. The post-test of 1 day after the last day of memorization consisted of all 100 words. We randomly divided 100 words into halves to take test effects into account. The post-test of 7 days after the last day of memorization consisted of a randomly selected halves of the 100 words. The post-test of 14 days after the last day of memorization consisted of the other halves. After finishing these post-tests, the participants of Group X were asked to memorize the Control wordset and take post-tests, and the participants of Group Y were asked to memorize the Experimental wordset and take post-tests. We conducted post-experiment questionnaires after all data were collected from the experiment.




(**a**) (**b**)

**Figure 2.** The Proposed System Application for the Experimental Setting of the Experiment. Participants can declare whether they remember the word or not. In addition, they can repeat or skip the word. Content is intentionally blurred. (**a**) Participants encounter an English word without any translation, (**b**) The Japanese translation is made available.




(**a**) (**b**)

**Figure 3.** Flashcard Application for the Control Setting of the Experiment. Participants can declare whether they remember the word or not. In addition, they can repeat or skip the word. (**a**) Participants encounter an English word without any translation, (**b**) The Japanese translation is made available.

## 5. Analysis

Figure 4 shows the difference between the number of correct answers for each experiment mode and the number of people who reached that score range. A positive value on the horizontal axis means the participants could learn effectively by using the proposed system. As shown in Figure 4, 19 out of the 47 participants obtained a better score on the post-test in the Experimental mode than in the Control mode.

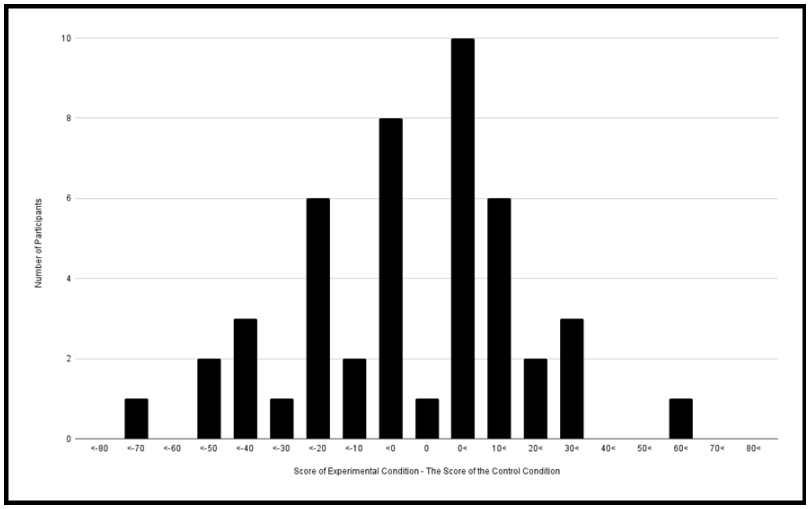



**Figure 4.** Aggregate Test Results for Users: Positive Results Indicate an Aggregate in Favor of Experimental Condition.

In order to investigate whether there is an overall and universal impact, we ran a two-tailed Wilcoxon–Pratt Signed-Rank Test [39]. Table 1 shows the overall statistic values for the Experimental wordset and the Control wordset. Table 2 shows the results for the Wilcoxon–Pratt Signed-Rank Test. The results indicate we cannot conclude that the proposed system is more or less effective for inducing better English vocabulary retention based upon a comparison test.

**Table 1.** Mean and Standard Deviation for the post-test score of the Experimental and Control Conditions for all Days of Testing. The Full score of Day 1 is 100, and the Full score of Day 7 and Day 14 is 50.

| Test Day | Experimental Mean (Score Rate) | Experimental STDEV | Control Mean (Score Rate) | Control STDEV |
|---|---|---|---|---|
| 1 | 79.19 (79.19%) | 17.29 | 81.25 (81.25%) | 21.72 |
| 7 | 35.21 (70.42%) | 9.69 | 37.23 (74.46%) | 10.01 |
| 14 | 28.82 (57.64%) | 8.45 | 27.63 (55.26%) | 10.54 |

**Table 2.** Results of a Wilcoxon-Pratt Signed-Rank Test Between the Experimental and Control Conditions for Each Test Day.

| Test Day | *p*-Value |
|---|---|
| 1 | 0.267 |
| 7 | 0.218 |
| 14 | 0.420 |

However, due to the distributions shown in Figure 4, it is necessary to investigate the impact on each user as well. Figure 5 shows the results for each participant. The results are ordered by the largest difference of decreasing correct answer rates for the Experimental wordset over the Control wordset. White tiles indicate ties, whereas blue tiles indicate the Control wordset scoring over the Experimental wordset. Based on this rule, Figure 5 lists the participants who have a positive effect from the proposed system, and the participants who have a positive effect from the standard flashcards. This means that there are some types of users for which the proposed system works well and others for which traditional methods are better.

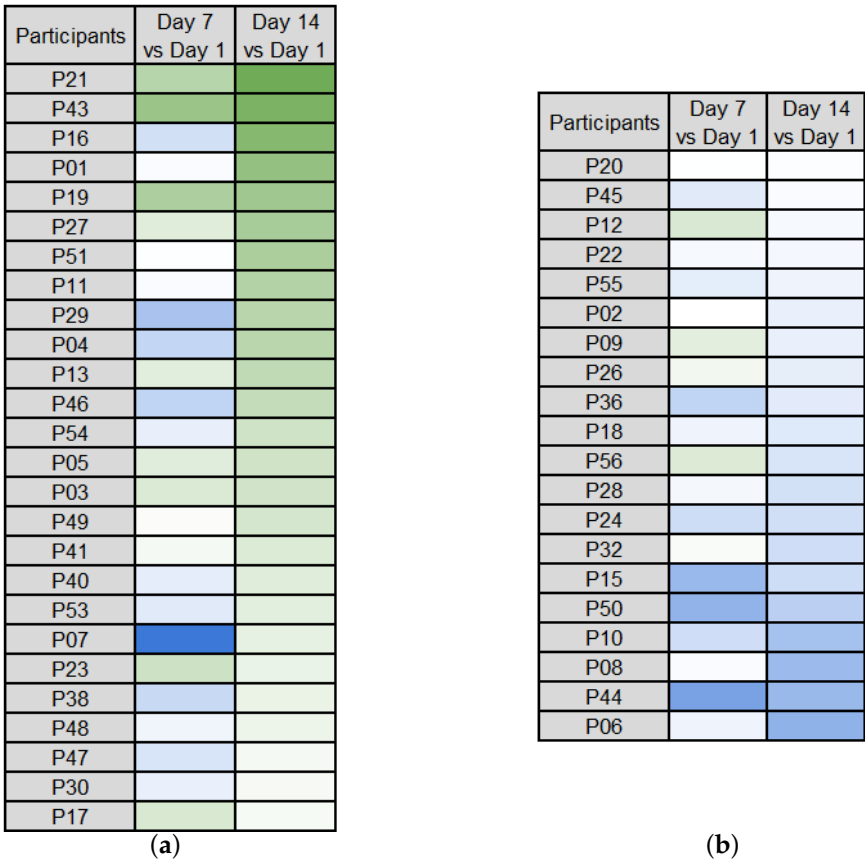



**Figure 5.** Test Day Differences in Memory Between Experimental wordset and Control wordset per User. White = No Difference, Light Green to Dark Green in Ascending Order of a Greater Difference in Favor of Experimental wordset. Light Blue to Dark Blue in Ascending Order of a Greater Difference in Favor of the Control wordset. Users are ordered by Day 14 Results. (**a**) Users where the Experimental Condition is more effective, (**b**) Users where the Control Condition is more effective.

The results provide some implications after two weeks of memorization. Firstly, participants with green colored squares in the second column show that vocabulary learned with the proposed system is less likely to be forgotten, and vocabulary learned from a standard flashcard is rapidly forgotten. Secondly, white squares show that learning with a standard flashcard easily loses its effectiveness. These implications indicate that the proposed system is, overall, effective for vocabulary retention. However, it is not applicable to all participants.

Since the learning effects are not universal, we utilized a post-experiment questionnaire and analyzed the results. To examine whether the proposed system is effective for a learner, we conducted linear regression analysis with variables collected from questionnaires. Table 3 shows variables we used in regression analysis. As shown in Table 3, the variables consist of Big-Five factor [40], Index of learning styles [41], TOEIC score, differences of repetition, and four answers to post-experiment questionnaires. To briefly understand which index of learning affects the proposed system, we divided each ILS score into two elements and normalized it on a continuous scale. Table 4 shows the questions asked in the post-experiment questionnaire. Question 1 asks the participants about their perceived ability to learn vocabulary. Question 2 asks about preference for learning with the proposed system and the traditional learning. Question 3 asks the participants about how they used the available context during the experiment. Question 4 asks the participants about how often they study English.

**Table 3.** Variables used in regression analyses.

| Association | Variables |
|---|---|
| Big-Five factor | Extraversion<br>Agreeableness<br>Conscientiousness<br>Neuroticism<br>Openness |
| Index of learning styles | Active<br>Reflective<br>Sensing<br>Intuitive<br>Visual<br>Verbal<br>Sequential<br>Global |
| English skill | TOEIC score |
| Difference of repetition | Mean<br>Minimum<br>Maximum<br>Median |
| Post-test questionnaire | Q1<br>Q2<br>Q3<br>Q4 |

**Table 4.** Post-test questionnaire.

| No. | Question | Response Method | Options |
|---|---|---|---|
| 1 | Are you good at vocabulary learning? | 5-point Scale | Very Bad - Very Good |
| 2 | Which flashcards do you prefer for daily use and to what extent? | 9-point Scale | Completely Prefer using context flashcards-Completely Prefer traditional flashcards |
| 3 | How did you use videos and captions in context-based flashcards? | 3 Choices | 1. "Complementary Use" 2. "Used to Same Extent as Traditional Memorization" 3. "Primarily Used" |
| 4 | In the past year, how often did you study English? | 6 Choices | About Once a Year-More than Once a Week |

We then constructed a score to compare success between the Experimental Condition and the Control Condition. As part of the experiment, the 100 word sets, were divided randomly into two exclusive sets with 50 words: $s_1$ and $s_2$. These were tested after 7 days and 14 days, respectively.

Let a function $m_x(s)$ be the number of words successfully memorized on Day $x$ from the set $s$. The retention rate is $r_x(s) = m_x(s)/|s|$ and the forgetting rate is $f_x(s) = 1 - r_x(s)$. When we describe the difference of forgetting rate between day $x$ and $y$ as $d(x, y)$, $d(7, 1)$, $d(14, 1)$, and $d(7, 14)$ are defined as follows:

- $d(7, 1) = f_7(s_1) - f_1(s_1)$.
- $d(14, 1) = f_{14}(s_2) - f_1(s_2)$.
- $d(14, 7) = f_{14}(s_2) - f_7(s_1)$.

Let $d(x, y|\mathrm{P})$ be the difference obtained by the proposed system and let $d(x, y|\mathrm{F})$ be the difference obtained by the flashcard. Thus:

- Day 7 vs. Day 1 is defined as: $d(7, 1|\mathrm{P}) - d(7, 1|\mathrm{F})$

- Day 14 vs. Day 1 : $d(14,1|\mathrm{P}) - d(14,1|\mathrm{F})$
- Day 14 vs. Day 7 : $d(14,7|\mathrm{P}) - d(14,7|\mathrm{F})$

After checking the variables for normality, we ran a series of three linear regression models. Table 5 shows the regression analysis result of taking Day 7 vs. Day 1, Day 14 vs. Day 1, and Day 14 vs. Day 7 as a dependent variable. We used the Akaike Information Criterion to refine the stepwise model selection. In the process of model selection for Day 14 vs. Day 1 and Day 14 vs. Day 7, participant P6 was removed as a high-leveraged outlier according to Cook's Distance. All models were satisfied with the assumptions for a linear model and multicollinearity was checked with a conservative threshold of VIF < 2.5. Multiple comparisons were handled with Bonferroni corrections.

**Table 5.** Result of regression analysis. * $p < 0.025$, ** $p < 0.005$.

| Target | Variables | Estimate ($\times 10^{-2}$) | SE ($\times 10^{-2}$) | t-Value | *p*-Value |
|---|---|---|---|---|---|
| Day 7 vs. Day 1 | Agreeableness | −0.715 | 0.395 | −1.810 | 0.0799 |
| | Neuroticism | −0.463 | 0.355 | −1.301 | 0.2027 |
| | Openness | 0.591 | 0.473 | 1.250 | 0.2208 |
| | Sensing | −3.978 | 1.387 | −2.868 | 0.0074 * |
| | Global | 1.947 | 1.361 | 1.430 | 0.1626 |
| | TOEIC | −0.027 | 0.023 | −1.182 | 0.2461 |
| | Q1 | 3.131 | 2.219 | 1.411 | 0.1681 |
| | Q2 | 1.556 | 0.983 | 1.584 | 0.1234 |
| | factor(Q3)2 | −9.808 | 5.205 | −1.884 | 0.0689 |
| | factor(Q3)3 | 1.920 | 5.277 | 0.364 | 0.7184 |
| | Median | −1.115 | 0.742 | −1.501 | 0.1434 |
| Day 14 vs. Day 1 | Extraversion | 0.779 | 0.344 | 2.262 | 0.0311 |
| | Neuroticism | −1.296 | 0.430 | −3.014 | 0.0052 * |
| | Active | 6.141 | 2.394 | 2.565 | 0.0159 * |
| | Reflective | 8.212 | 1.879 | 4.370 | 0.0001 ** |
| | Sensing | −4.925 | 2.051 | −2.401 | 0.0228 * |
| | Intuitive | −5.156 | 3.202 | −1.610 | 0.1179 |
| | TOEIC | −0.053 | 0.024 | −2.257 | 0.0314 |
| | factor(Q3)2 | −10.672 | 5.531 | −1.929 | 0.0632 |
| | factor(Q3)3 | −13.878 | 6.966 | −1.992 | 0.0555 |
| | factor(Q4)1 | 1.716 | 4.614 | 0.372 | 0.7125 |
| | Median | −1.166 | 0.782 | −1.491 | 0.1464 |
| Day 14 vs. Day 7 | Extraversion | 0.783 | 0.278 | 2.813 | 0.0082 * |
| | Reflective | 4.474 | 1.550 | 2.886 | 0.0068 * |
| | Intuitive | −3.427 | 2.328 | −1.472 | 0.1505 |
| | TOEIC | −0.043 | 0.023 | −1.861 | 0.0717 |
| | Q1 | −5.040 | 2.313 | −2.179 | 0.0366 |
| | Q2 | −1.665 | 1.055 | −1.579 | 0.1239 |
| | factor(Q3)2 | −1.010 | 5.169 | −0.195 | 0.8464 |
| | factor(Q3)3 | −16.115 | 6.163 | −2.615 | 0.0134 * |

In the results, as shown in Table 5, a negative estimated value means the independent variable had positive effect on the proposed system. A positive estimated value means the independent variable had an adverse effect on the proposed system.

The models themselves vary in how much of the performance they can explain. In Day 7 vs. Day 1, $R^2 = 0.2285$, in Day 14 vs. Day 1, $R^2 = 0.4417$, and in Day 14 vs. Day 7, the $R^2 = 0.3051$. While none of the models can explain the majority of the user outcome, it is promising that factors that were recorded can contribute to understanding for whom this system works. In the future, it might be possible to quickly assess and recommend tools to different types of learners.

In summary, we found the following:

- For some users, the proposed system has an overall positive effect on English vocabulary retention two weeks after the memorization term.

- The system is not beneficial for all users, for whom a simple flashcard seems to be more beneficial.
- Some indexes relating to personality traits and learning styles are key factors to explain whether the proposed system is effective for a learner.

## 6. Discussion

In a previous study, Yamaguchi et al. [12] showed that a context-based text flashcard application could potentially be effective, at least over a short duration, for some participants. In this experiment, the multimedia that was available to the participants and the time in which they used it was expanded. As a result, the outcomes clearly illustrate both the promise and limitation of new technologies which allow learners to study for language acquisition with flashcards that are augmented with media content. Although context-based learning definitely has benefits for some users, other users will not benefit from this approach. Based on the results of this study, certain personality traits indicate the presence or lack of affinity to the learning style, which means that it should be possible to recommend whether this application is likely to work for an individual or not.

When looking at the comparison results, we can see that certain factors are important. In the case of Day 7 vs. Day 1, as shown in row 1 in Table 5, Sensing had a positive effect on the proposed system. People who belong to the learning style of Sensing prefer to learn from concrete and practical information. In this experiment, most participants used educational videos such as TED. As a result, the participants who belong to the learning style of Sensing can effectively acquire vocabulary through video content with the proposed system.

In the case of Day 14 vs. Day 1, as shown in row 2 in Table 5, Neuroticism and Sensing score had positive effects on the proposed system. This result suggests that comprehension of the contexts leads to retaining vocabulary. Videos and subtitles can facilitate learning for those with high English skills. However, they can be a hindrance to learning for those without English skills because the contents are difficult to understand. Additionally, both Active and Reflective, which are defined as the opposite indexes in ILS, had adverse effects on the proposed system. This indicates that the proposed system outperforms people who can process information in both an active and reflective way.

In the case of Day 14 vs. Day 7, as shown in row 3 in Table 5, the proposed system affects people who are good at vocabulary learning and people who mainly use video contexts in video attached flashcards. This indicates that context comprehension and its effective use contribute to vocabulary retention. On the other hand, Extraversion and Reflective had an adverse effect on the proposed system. This trend also appears in the case of Day 14 vs. Day 1.

The results show that it is important to develop applications and systems with the individual user in mind. Overall, there is not a universal advantage between the proposed system and the typical digital flashcard, as opposed to the clear advantage that can be found between digital flashcards and paper flashcards [30–34]. Unlike the move from paper flashcards to digital flashcards, which can be recommended in general, the context-based approach is appropriate for some learners. As context-based personalized learning becomes more and more prevalent, there needs to be an emphasis by developers on recommending the right type of learning style for each user. This proposes several challenges. The methods for recommendations need to be accurate enough to be useful for users while preserving user privacy and engagement at the same time. At the same time, the power of the results is not strong enough to be prescriptive in nature. That is, it would not be possible to build a recommendation system on this analysis alone.

### *Limitations and Future Work*

There are several limitations to this paper. First, there is a lack of an objective measurement of the "stickiness" of the proposed system. We do not know how users will use this context-learning system free from any guidance. This lack of analysis may either limit

the potential impact of the proposed system or overstate it. That is, the proposed system might actually result in more usage, and therefore, a benefit, even for users who did better with the traditional flashcard, or vice versa. Another limitation is the lack of an analysis of long-term memorization. One method may universally lead to better long-term results.

An important limitation to acknowledge is the demographics of this study. The vast majority of participants in the study are young Japanese male university students. There may be cultural impacts within this study that limit the generalizability of the study. Further work with more diverse participant cohorts needs to be completed.

Another area of exploration beyond the scope of this paper is learning beyond English vocabulary for learners. The current focus of the research is on second-language vocabulary acquisition in adults. The approach may have the potential to encourage childhood vocabulary acquisition in mobile settings as well as being portable to other subjects where memorization and repetition are required.

A vital area of future work is furthering the understanding of and ability to describe the relationship between different learning styles and the affinity for learning with the context-based system. Future work will include analysis on both in-the-wild usage as well as longer-term studies, with a focus on collecting more data with user traits and the ability to learn with the different systems. The term goal is to build an effective recommendation system for learners.

## 7. Conclusions

In the paper, we present a system that allows mobile users to engage in context-based learning for English vocabulary acquisition. In particular, we expand the media that can be used from text-only to video and audio. A study with 46 participants using videos found that there is no universal benefit for all users. However, we did find that certain learning and personality traits indicate a proclivity towards or against the proposed system as compared to a more traditional flashcard application.

The study shows that there is no one best learning style for all learners, and this extends to the mobile space. The proposed system cannot be recommended to all learners, such as those who exhibit the personality trait of Extraversion and the learning traits of Active and Reflective, even if the technology seems attractive. But the approach does promise important benefits for vocabulary acquisition for some individuals, such as those who have the personality trait of Neuroticism and the learning trait of Sensing. Future work should focus on the creation of non-invasive recommendation systems for learning styles and platforms.

**Author Contributions:** Conceptualization, A.V., K.Y., M.I. and K.K.; methodology, A.V., K.Y., M.I. and K.K.; formal analysis, A.V. and K.Y.; writing—original draft preparation, A.V. and K.Y.; writing—review and editing, A.V., M.I. and K.K.; supervision, M.I. and K.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by JST CREST (JPMJCR16E1), JST Trilateral AI (JPMJCR20G3), JSPS Kakenhi (20H04213, 20KK0235, 23KK0188), and the MEXT "Innovation Platform for Society 5.0" Program (JPMXP0518071489).

**Institutional Review Board Statement:** This research was approved by the ethics committee of the Graduate School of Engineering of Osaka Prefecture University (Now: Osaka Metropolitan University) on 27 December 2021.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data are available upon request from the first author. The data are not publicly available due to privacy considerations.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| VIF | Variance Inflation Factor |
| GLM | Generalized Linear Model |

## References


1. Chen, Z.; Chen, W.; Jia, J.; An, H. The effects of using mobile devices on language learning: A meta-analysis. *Educ. Technol. Res. Dev.* **2020**, *68*, 1769–1789. [CrossRef]
2. Crompton, H.; Burke, D. The use of mobile learning in higher education: A systematic review. *Comput. Educ.* **2018**, *123*, 53–64. [CrossRef]
3. Ashcroft, R.J.; Cvitkovic, R.; Praver, M. Digital Flashcard L2 Vocabulary Learning Out-Performs Traditional Flashcards at Lower Proficiency Levels: A Mixed-Methods Study of 139 Japanese University Students. *EuroCALL Rev.* **2016**, *26*, 14–28. [CrossRef]
4. Klímová, B.; Berger, A. Evaluation of the use of mobile application in learning English vocabulary and phrases—A case study. In *International Symposium on Emerging Technologies for Education*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–11. [CrossRef]
5. Webb, S. The effects of context on incidental vocabulary learning. *Read. Foreign Lang.* **2008**, *20*, 232–245.
6. Frances, C.; Martin, C.; Duñabeitia, J.A. The effects of contextual diversity on incidental vocabulary learning in the native and a foreign language. *Sci. Rep.* **2020**, *10*, 13967. [CrossRef] [PubMed]
7. Pigada, M.; Schmitt, N. Vocabulary acquisition from extensive reading: A case study. *Read. Foreign Lang.* **2006**, *18*, 1–28.
8. Ahmad, J. Intentional vs. incidental vocabulary learning. *ELT Res. J.* **2012**, *1*, 71–79.
9. Hidi, S. Interest, reading, and learning: Theoretical and practical considerations. *Educ. Psychol. Rev.* **2001**, *13*, 191–209. [CrossRef]
10. Harackiewicz, J.; Smith, J.; Priniski, S. Interest Matters: The Importance of Promoting Interest in Education. *Policy Insights Behav. Brain Sci.* **2016**, *3*, 220–227. [CrossRef]
11. Augereau, O.; Jacquet, C.; Kise, K.; Journet, N. Vocabulometer: A Web Platform for Document and Reader Mutual Analysis. In Proceedings of the 2018 13th IAPR International Workshop on Document Analysis Systems (DAS), Vienna, Austria, 24–27 April 2018; pp. 109–114. [CrossRef]
12. Yamaguchi, K.; Iwata, M.; Vargo, A.; Kise, K. Mobile vocabulometer: A context-based learning mobile application to enhance English vocabulary acquisition. In Proceedings of the UbiComp/ISWC '20 Adjunct: Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers, Virtual, 12–17 September 2020; pp. 156–159. [CrossRef]
13. Yurdagül, C.; Öz, S. Attitude towards mobile learning in English language education. *Educ. Sci.* **2018**, *8*, 142. [CrossRef]
14. Biswas, B.; Roy, S.K.; Roy, F. Students perception of mobile learning during COVID-19 in Bangladesh: University student perspective. *Aquademia* **2020**, *4*, ep20023. [CrossRef]
15. Alhumaid, K.; Habes, M.; Salloum, S.A. Examining the factors influencing the mobile learning usage during COVID-19 Pandemic: An Integrated SEM-ANN Method. *IEEE Access* **2021**, *9*, 102567–102578. [CrossRef]
16. Korucu, A.T.; Alkan, A. Differences between m-learning (mobile learning) and e-learning, basic terminology and usage of m-learning in education. *Procedia-Soc. Behav. Sci.* **2011**, *15*, 1925–1930. [CrossRef]
17. Webb, S. Learning word pairs and glossed sentences: The effects of a single context on vocabulary knowledge. *Lang. Teach. Res.* **2007**, *11*, 63–81. [CrossRef]
18. Restrepo-Ramos, F. Incidental Vocabulary Learning in Second Language Acquisition: A Literature Review. *Profile Issues Teach. Prof. Dev.* **2015**, *17*, 157–166. [CrossRef]
19. Miles, K.P.; Ehri, L.C. Learning to read words on flashcards: Effects of sentence contexts and word class in native and nonnative English-speaking kindergartners. *Early Child. Res. Q.* **2017**, *41*, 103–113. [CrossRef]
20. Lee, S.; Pulido, D. The impact of topic interest, L2 proficiency, and gender on EFL incidental vocabulary acquisition through reading. *Lang. Teach. Res.* **2017**, *21*, 118–135. [CrossRef]
21. Santi, E.; Kholipa, R.; Putri, M.; Mujiono, M. Reading Interest Strength and Vocabulary Acquisition of EFL Learners: A Meta-Analysis. *J. Lang. Linguist. Stud.* **2021**, *17*, 1225–1242. [CrossRef]
22. Perez, M.M.; Van Den Noortgate, W.; Desmet, P. Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System* **2013**, *41*, 720–739. [CrossRef]
23. Li, X.; Brand, M. Effectiveness of music on vocabulary acquisition, language usage, and meaning for mainland Chinese ESL learners. *Contrib. Music. Educ.* **2009**, *36*, 73–84.
24. Rodgers, M.P. The images in television programs and the potential for learning unknown words: The relationship between on-screen imagery and vocabulary. *ITL-Int. J. Appl. Linguist.* **2018**, *169*, 191–211. [CrossRef]
25. Meganathan, P.; Yap, N.; Paramasivam, S.; Jalaluddin, I. Incidental and Intentional Learning of Vocabulary among Young ESL Learners. *3L Southeast Asian J. Engl. Lang. Stud.* **2019**, *25*, 51–67. [CrossRef]
26. Karami, A.; Bowles, F. Which Strategy Promotes Retention? Intentional Vocabulary Learning, Incidental Vocabulary Learning, Or A Mixture Of Both? *Aust. J. Teach. Educ.* **2019**, *44*, 25–43. [CrossRef]
27. Laufer, B. Reading, Word-Focused Activities and Incidental Vocabulary Acquisition in a Second Language. *Prospects* **2001**, *16*, 44–54.

28. Schmitt, N. Instructed second language vocabulary learning. *Lang. Teach. Res.* **2008**, *12*, 329–363. [CrossRef]
29. Lin, J.J.; Lin, H. Mobile-assisted ESL/EFL vocabulary learning: A systematic review and meta-analysis. *Comput. Assist. Lang. Learn.* **2019**, *32*, 878–919. [CrossRef]
30. Nikoopour, J.; Kazemi, A. Vocabulary learning through digitized & non-digitized flashcards delivery. *Procedia-Soc. Behav. Sci.* **2014**, *98*, 1366–1373. [CrossRef]
31. Xodabande, I.; Pourhassan, A.A.; Valizadeh, M. Self-directed learning of core vocabulary in English by EFL learners: comparing the outcomes from paper and mobile application flashcards. *J. Comput. Educ.* **2021**, *9*, 93–111. [CrossRef]
32. Xodabande, I.; Iravi, Y.; Mansouri, B.; Matinparsa, H. Teaching Academic Words With Digital Flashcards: Investigating the Effectiveness of Mobile-Assisted Vocabulary Learning for University Students. *Front. Psychol.* **2022**, *13*, 893821. [CrossRef]
33. Xodabande, I.; Atai, M.R. Using mobile applications for self-directed learning of academic vocabulary among university students. *Open Learn. J. Open Distance-Learn.* **2022**, *37*, 330–347. [CrossRef]
34. Zakian, M.; Xodabande, I.; Valizadeh, M.; Yousefvand, M. Out-of-the-classroom learning of English vocabulary by EFL learners: Investigating the effectiveness of mobile assisted learning with digital flashcards. *Asian-Pac. J. Second. Foreign Lang. Educ.* **2022**, *7*, 16. [CrossRef]
35. Higashimura, R.; Vargo, A.; Iwata, M.; Kise, K. Helping Mobile Learners Know Unknown Words through their Reading Behavior. In Proceedings of the Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22). New Orleans, LA, USA, 30 April–6 May 2022; pp. 1–5. [CrossRef]
36. La Pray, M.; Ross, R. The graded word list: Quick gauge of reading ability. *J. Read.* **1969**, *12*, 305–307.
37. Leitner, S. *So Lernt Man Lernen*; Angewandte Lernpsychologie ein Weg zum Erfolg; Herder: Freiburg im Breisgau, Germany, 1972.
38. Xu, W.; Callison-Burch, C.; Napoles, C. Problems in current text simplification research: New data can help. *Trans. Assoc. Comput. Linguist.* **2015**, *3*, 283–297. [CrossRef]
39. Pratt, J.W. Remarks on zeros and ties in the Wilcoxon signed rank procedures. *J. Am. Stat. Assoc.* **1959**, *54*, 655–667. [CrossRef]
40. Goldberg, L.R. The development of markers for the Big-Five factor structure. *Psychol. Assess.* **1992**, *4*, 26. [CrossRef]
41. Felder, R.M.; Silverman, L.K. Learning and teaching styles in engineering education. *Eng. Educ.* **1988**, *78*, 674–681.