Language Learning On-The-Go: Opportune Moments and Design of Mobile Microlearning Sessions

Tilman Dingler
VIS, University of Stuttgart
tilman.dingler@vis.uni-stuttgart.de

Dominik Weber
VIS, University of Stuttgart
dominik.weber@vis.uni-stuttgart.de

Martin Pielot
Telefonica Research
martin.pielot@telefonica.com

Jennifer Cooper
University of Minnesota
cooperj@umn.edu

Chung-Cheng Chang
University of Washington
changcc@uw.edu

Niels Henze
VIS, University of Stuttgart
niels.henze@vis.uni-stuttgart.de

ABSTRACT
Learning a foreign language is a daunting and time-consuming task. People often lack the time or motivation to sit down and engage with learning content on a regular basis. We present an investigation of microlearning sessions on mobile phones, in which we focus on session triggers, presentation methods, and user context. Therefore, we built an Android app that prompts users to review foreign language vocabulary directly through notifications or through app usage across the day. We present results from a controlled and an in-the-wild study, in which we explore engagement and user context. In-app sessions lasted longer, but notifications added a significant number of "quick" learning sessions. 37.6% of sessions were completed in transit, hence learning-on-the-go was well received. Neither the use of boredom as trigger nor the presentation (flashcard and multiple-choice) had a significant effect. We conclude with implications for the design of mobile microlearning applications with context-awareness.

Author Keywords
Microlearning; Microinteractions; Language Learning; Prediction; Attention; Mobile Device.

ACM Classification Keywords
H.5.m Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
Learning a new language is an activity that many people would like to pursue, but are challenged by its time-intensive nature. According to the The Foreign Service Institute (FSI) of the US Department of State, it can take up to 750 class hours (30 weeks) for a native English speaker to become proficient in German[1].Many language learners simply cannot fulfill such a time commitment or lack the motivation to spend a great deal of time and resources on intensive or immersive instructions.

MobileHCI '17, September 04-07, 2017, Vienna, Austria

ISBN 978-1-4503-5075-4/17/09…$15.00

DOI: http://dx.doi.org/10.1145/3098279.3098565

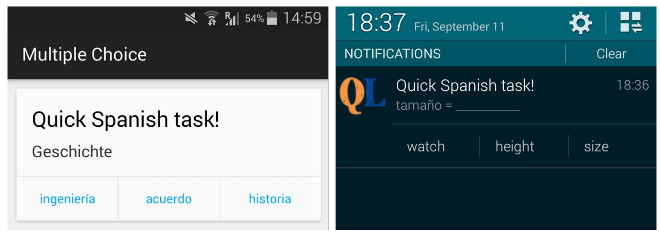

Figure 1. *QuickLearn* allows users to engage in microlearning sessions (left) and nudges users to re-visit vocabulary through system-initiated, interactive notifications (right).

While most products require learners to dedicate large chunks of time on regular basis, mobile apps are often designed to support *microlearning* sessions. Microlearning sessions break down a learning task into a series of quick learning interactions, thereby reducing learning units to more manageable chunks. These chunks can then be completed, for example, during idle moments, such as while waiting for the bus or standing in line at the supermarket. The fact that people tend to keep their smartphone in close reach most of the time [15] and exhibit a high attentiveness towards their devices [16], opens up the possibility to engage them in short learning tasks spaced out throughout the day. In addition to enabling people who cannot dedicate large chunks of time to learn a new language, research in psychology found that repetitions are more effective than dedicated long streaks of learning [12].

One major challenge is how to best design for such microlearning sessions. Due to the anytime-anywhere capabilities of mobile devices, researchers commonly suggest that microlearning is a great tool to convert spare time into productive phases [5, 6, 33]. Since some people may need nudges to keep on learning a language, notifications can be used to turn users' attention to the device. More recently, notifications started to be populated

[1] `http://web.archive.org/web/20071014005901/http://www.nvtc.gov/lotw/months/november/learningExpectations.html`

by interactive content, which allows users to quickly deal with short tasks through the notification drawer and therefore open up a design space for microlearning sessions. In these brief interactions users can, for example, review a few words of foreign language vocabulary in different situations throughout the day. Thereby it is still unclear whether mobile users prefer engaging with language learning tasks during short transient moments, e.g., during their commute on a train. Also, there is the question of how such notification-triggered sessions should be designed and when they should be triggered.

Therefore, we present *QuickLearn*, an app to explore and compare different designs of microlearning sessions. Through interactive notifications the app invites learners to frequently review language content. Word exercises can be attended to directly in the notification drawer or through self-initiated user sessions (see Fig. 1). We assess the feasibility of notification triggers, presentation modes, and interaction modalities with regard to microlearning sessions. We present findings from a mixed method user study: a controlled lab study and an in-the-wild experiment. Participants were able to acquire 18 new words per week through micro-sessions. While notification sessions lasted shorter, more words were reviewed per app session than per notification. Learning sessions in-between tasks and on-the-go were preferred rather than at home or work. Additionally, we collected phone usage data and analyzed features and phone usage patterns that correlate with microlearning sessions and user engagement. Here we noted that users welcomed notifications when they had interacted with the phone more recently.

The main contributions of this paper are:

1. An investigation of context factors of microlearning in mobile settings.
2. An assessment of the feasibility of using interactive notifications for learning sessions, which last shorter than app launches.
3. A comparison of two vocabulary review methods: flashcards vs. multiple choice, of which flashcards tend to be better suited for new word acquisitions.
4. An exploration of using idle moments as an opportunity for learning and memory strengthening.

BACKGROUND AND RELATED WORK

Our work is mainly inspired by and based on previous research in learning theories, technology-mediated learning, and work regarding user attention and interruptions.

The Effects of Repetition on Vocabulary Learning

Research has shown that the extent to which vocabulary gains are made is positively related to the number of times language learners encounter each vocabulary word [34]. Vocabulary learning is a gradual process in which gains are made in small increments with repeated encounters needed to achieve full knowledge of a word [26]. In addition to repetition, word review can be enhanced with spacing. The spacing effect refers to the retention advantage for information that is repeated in a distributed fashion relative to information that is repeated in a massed fashion [8]. Dempster [14] demonstrated that language learners retained greater number of vocabulary words when they were distributed and spaced apart instead of being presented in succession. Spacing effects not only apply within a learning session, but also between sessions. Bahrick and Phelps [2] demonstrated that words were better retained when studied and re-learned at 30-day intervals as compared to 1-day or 0-day intervals.

The use of computers can greatly enhance spaced repetition vocabulary learning. It is a lot easier for a computer to keep a record of the learner's performance and control the sequencing of vocabulary words appropriately. Language learners who study with physical flashcards need to monitor their own learning progress, in turn running the risk of inefficient learning [27]. A study conducted by Nakata [27] has shown that students were able to recall more vocabulary words using the computer when compared to word lists.

Microlearning

Microlearning is the technique of breaking down a learning task into a series of short learning interactions [4]. Researchers have created a large number of applications that integrate microlearning with day-to-day activities. Cai *et al.* [7], for instance, created an Instant Message web application where users learned words during the wait time of an ongoing chat. Trusty and Troung [32] built a browser extension that replaces a selected set of English words with their foreign translation.

Mobile devices provide a powerful platform for some forms of learning where individualization (or personalization) of learning content with anytime and anywhere access is critical [25]. The portability of the mobile device along with the low cost of retrieval make them a promising platform for microlearning. Ashbrook [1] demonstrated the importance of microinteractions, explaining that quick and easy access to a device is important for promoting frequent use. Prior to the ubiquity of smartphones, Cavus and Ibrahim [9] sent SMS messages of English vocabulary words. Edge *et al.* [20] created a mobile app that uses the phone's location to deliver contextual relevant vocabulary words. To ensure repeated exposure to a vocabulary word, Dearman and Truong [13] created a mobile app that displays the vocabulary word on the wallpaper of the phone.

Concluding, vocabulary-based microlearning experiments have shown to improve vocabulary acquisition and recall [7, 20, 32], while microlearning on a mobile device has also garnered high user acceptance [6, 9, 10, 31].

Flashcard and Multiple Choice Learning

Unlike lesson-oriented learning, flashcards operate at the granularity of facts (e.g. word translation) and measure the learner's ability to provide the correct response in the presence of a stimulus [20]. The use of flashcards is a common way to learn vocabulary as it allows to group difficult words and, which are then reviewed more frequently than comparatively easy words [27]. Furthermore, flashcards enhance vocabulary learning by reducing the list effect where learners tend to recall words more easily due to the order they appear on the word list.

Numerous researchers have explored microlearning applications based on flashcard presentation. Basoglu and Akdemir [3] invited students to use a mobile phone with a flashcard vocabulary application. Results indicated that using mobile phones as a vocabulary learning tool is more effective than traditional vocabulary learning tools. Edge *et al.* [19] created a mobile flashcard application with an adaptive spaced repetition algorithm.

Multiple choice learning has not been extensively applied to mobile devices; however, vocabulary skills are commonly assessed through multiple choice tests [30]. Primarily, multiple choice quizzes highlight deficiencies in language acquisition, thereby providing a similar baseline as flashcard learning.

Notifications & Interruptions

Microlearning mobile applications have shown a lot of promise and potential to improve vocabulary learning. However, studies on microlearning are typically not optimized for sustained or continuous use. One way to promote sustained use of the app is through push notifications, however the language learning app must also compete with numerous notifications by other apps. Research suggests that frequent push notifications cause interruptions and can induce stress [35]. Such negative effects induced by frequent and interruptive push notifications, however, can be lessened. Studies have shown that pushing a notification during a breakpoint reduces frustration [24] and is dealt with significantly more quickly [21]. Poppinga *et al.* [29] collected data from smartphone sensors and applied machine learning models to automatically predict and identify opportune moments to trigger notifications. In our previous work we conducted a similar experiment by leveraging smartphone sensor data and machine learning models to predict users' boredom. Results showed that mobile users are more likely to engage in suggested push notification content when they are bored [28].

Knowledge gains can be made with repeated exposure and learning sessions can be broken down into small chunks of quick interactions presented on mobile devices. However, mobile applications geared toward learning need to compete for the attention of the user who may have hundreds of other apps installed. In our investigation we focus on how to best design mobile learning applications as well as systematic push notifications to promote prolonged and sustained mobile learning sessions. Therefore, we address the following research questions:

- **RQ1**: which context factors are most opportune for language learning?
- **RQ2**: how effective are interactive notifications for learning tasks compared to a dedicated mobile app?
- **RQ3**: which review method leads to better learning results in mobile applications: flashcards or multiple choice questions?
- **RQ4**: do moments of boredom correlate with situations, in which users are more likely to engage in learning tasks?

QUICKLEARN: DESIGN AND IMPLEMENTATION

We developed *QuickLearn*, a vocabulary trainer for mobile devices, which runs on Android phones with OS 4.3 or newer and provides two distinct modes: one allows users to actively open the app and review vocabulary and the other consists of a background service that triggers notifications to remind users to review vocabulary. On its first launch the app shows a consent form explaining what data is collected and for which purpose. After explicitly stating their consent, users are guided through the permission granting process, which consists of giving access to the Android Accessibility Services and allowing the app to access notifications. This is needed to inform the notification schedule algorithm and to take into account the user's activities in order to trigger notifications. In the final step, the app asks the user to specify age, gender, optionally leave an e-mail address, and allows to set a mother tongue (L1), the language to be learned(L2), as well as the current level of proficiency in that language. Since the application supports five languages, we asked users to choose the language they were most comfortable with in case their mother tongue was unavailable. Once the setup is complete, the background service is started and the app shows the first set of words.

User-Initiated Learning Sessions

The application can be launched anytime. Upon start, it starts an activity that shows words to be reviewed in sets of three. After each set, users are asked whether they would like to continue with the next set of words or to quit the application. The design intends to support the notion of short learning sessions to give users the feeling of completion and accomplishment in minimum time.

System-Initiated Learning Sessions

While the application can be explicitly launched by the user, an additional service runs in the background and initiates learning sessions by triggering notifications. These notifications remind users to review a set of vocabulary. The review can be done directly in the notification drawer, thereby avoiding the need for opening the application and enforcing a context switch. By adding buttons to the notification, users are able to interact with the notification's content, which is why we consider them *interactive* notifications. The app can also be directly launched by clicking on the notification itself.

Results from our previous work suggests that people are more receptive to external stimulation when they are bored [28]. Since the success of microlearning strongly depends on peoples' willingness to engage with tasks in respective moments, boredom might be an appropriate state to study language vocabulary. To examine this hypothesis (RQ4), we applied our model of Pielot *et al.* [28] and integrated the boredom classifier into the notification trigger algorithm. The model uses features, such as the recency of communications, demographics, or intensity of recent phone usage to estimate whether the phone user is bored with an accuracy of 74.5% AUCROC. The notification trigger mechanism works as follows: whenever the user turns the phone's screen on, the app service tries to schedule a notification. Therefore, it checks whether a notification is already scheduled or whether the minimum time of 20 minutes since the last posted notification has elapsed. To not disturb users at night, notifications are only posted between 7am and 11pm. If these conditions are all met, a timeout is scheduled with a random delay between 10 seconds and 5 minutes. When the timeout runs out, the classifier estimates the user's boredom state. In case the user is detected to be bored, a notification is posted. If the boredom state is predicted to be negative, the notification is only posted in 1 out of 9 cases; in pilot tests we found this number to make sure that an equal amount of notifications is posted for each of the two boredom states (bored and non-bored). In case the app is already in the foreground, no notification is posted.

User Interface Design

Words are presented in two display modalities: *flashcards* and *multiple choice*. The modality is randomly assigned when the app is first launched. The flashcard method consists of a series of screens, beginning with the presentation of a foreign noun with no translation (see Fig. 2). Users can then click "translate" to view the translation and are instructed to acknowledge whether they already knew the word ("knew it") or have had no previous knowledge of the word ("did not know"). This way, the words that are tagged as previously known by the user can be separated from the words that need additional practice.

In the multiple choice modality, a foreign word is shown to the user together with three possible translations (see Fig. 2): a random order of words including the correct translation as well as two suggestions randomly retrieved from the vocabulary set. The three options are arranged in a random order. Users then have to choose which translation they believe to be correct. The application provides immediate feedback on whether the selection is correct or wrong. Correctly guessed words are tagged accordingly and have a lower frequency of repetition than those that were guessed wrong.

Word List

We used a vocabulary list consisting of common English nouns originally taken from the British National Corpus[2]. The list comprises high frequency English nouns since in second language learning nouns are typically acquired before verbs and are less context-dependent. We used Google Translate to translate these words to Spanish, French, German, and Arabic. Finally, two native speakers manually went through the list, corrected inaccurate translations, and flagged highly ambiguous words. The final word list consists of 476 nouns in each language. The translations are uni-directional, meaning that a word was reviewed by being presented in its foreign translation (L2) before revealing it in the user's comfort language (L1).

Word Scheduling

Vocabulary words are delivered in sets of three, providing obvious stopping points should the user feel inclined to stop. After each set of words, users are asked if they would like to continue or quit. A mechanism of spacing words for repetition is employed using a variation of the Leitner style schedule [22]: this schedule is based on the principle of spaced repetition. Given that humans exhibit a negatively exponential forgetting curve [18], repetitions need to occur at increasingly spaced intervals. Hence, new words are encountered just as they are about to be forgotten. Hence, whenever the user reviews a word, it is tagged as either correctly or incorrectly guessed. In the flashcard modality, this is done by admitting to have known the word or not. In the multiple choice modality, this is automatically done by evaluating the given answer. When the app is first launched, the word list is randomized. With every word review the app sequentially goes through that list. If a word is seen for the first time and guessed correctly, it is appended to the very end of the word list. If this word is guessed incorrectly, it is replicated and spaced throughout the remaining wordlist in positions 4, 8, 16, 32, ... and so on until the end of the word list is reached. Whenever a word is encountered that is represented multiple times in the list, it is simply removed regardless of being guessed correctly or incorrectly, hence making sure that the list does not grow indefinitely.

CONTROLLED USER STUDY

To gain insights into the use of microlearning sessions for language acquisitions, we conducted a user study where we explicitly recruited participants to use the app over a course of two weeks.

Method

We collected usage data from 17 participants who installed the app under our supervision and returned after each week to fill in a survey and talk about their experiences. Regarding the vocab review methods, we used a repeated measure design with the review method being the independent variable. We counterbalanced the starting condition so that half of the participants started using the app with flashcards, and another half started out with using multiple choice answers. After seven days, the method changed automatically. As dependent variables, we collected app logs, general phone usage data as listed in Table 2, and information about the notifications triggered by the app. We assessed language learning effects by administering vocabulary tests after each week. Finally, participants were asked to provide subjective user feedback and comprehensive comments during a final semi-structured interview.

Participants

We initially recruited 19 participants (9 female) through university mailing lists and personal connections with ages ranging between 22 and 44 ($M = 28, SD = 5.2$) years. One participant needed to be excluded from the quantitative analysis due to technical reasons and another because he had his phone stolen during the study. Most of the remaining 17 participants were students, of which nine indicated German to be their first language, one French, two Arabic, two English, and three Spanish. Table 1 shows an overview of their target languages and their current proficiency levels. Three participants indicated to be currently enrolled in a second language course.

Data Collection

Once the setup was complete, the app started the notification schedule service as described above. It also included a log

[2] http://www.natcorp.ox.ac.uk

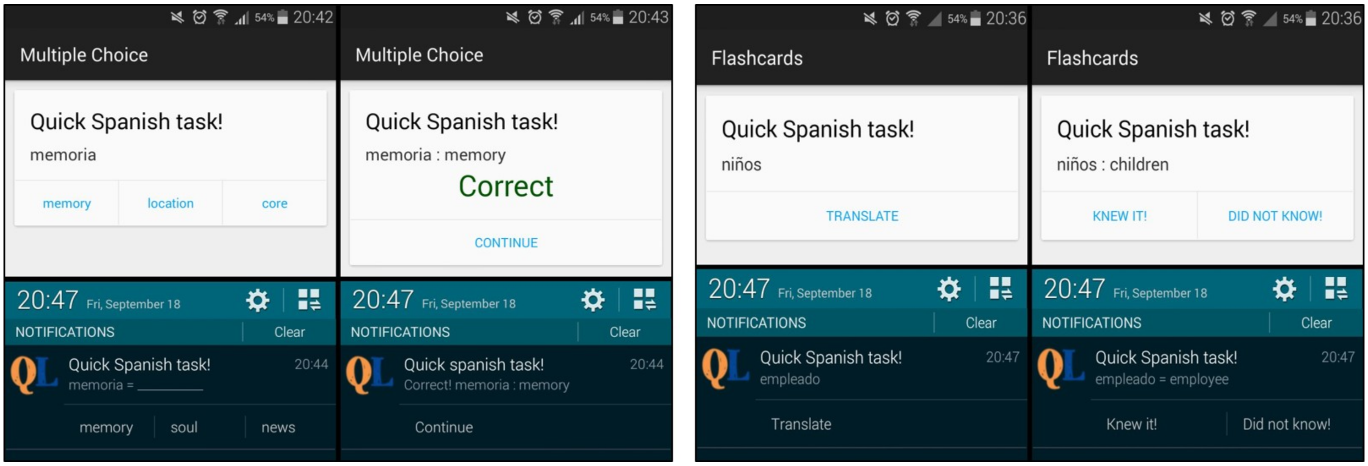

Figure 2. The multiple choice modality via app and notification showing three possible answers and correct response (left), and the flashcard modality via app (top row) and notification (bottom row) displaying initial screen and translation screen (right).

Language	None	Elementary	Limited	Professional	Full
German	1	2	2	2	0
English	0	0	1	0	0
Spanish	1	3	0	0	0
French	2	2	1	0	0

Table 1. Number of participants and their selection of their target language proficiency in the controlled user study.

system where every app interaction was tracked: whenever the app was opened by the user, how it was launched (clicking the application icon, the notification, or direct interaction in the notification), which words were reviewed and with which review method. Further, we collected context parameters, such as time and location. Sensor and usage data were sent to our servers through a secure connection. To prevent straining users' data plans, the app stored the collected data locally and transmitted only when connected through WiFi. Whenever a notification was triggered, interacted with, dismissed or ignored, and when the app was explicitly launched, we logged the current boredom state as predicted by the classifier.

Procedure

We invited participants to our lab, where we walked them through the purpose of the study and had them fill in an initial survey collecting demographic data as well as their language preferences and previous knowledge. We explained the types of data that would be collected and how this data would be transmitted and stored. Then, we installed the app on their phone, went through the setup procedure and had them complete a first set of words. We explained details about the app's intention to post notifications and explicitly asked users to not feel obliged to answer every incoming app notification, but rather deal with them as found convenient. We counterbalanced the review mode to start with either the flashcard or multiple choice mode. This first setup meeting took about 15 minutes, after which participants used the app for seven consecutive days.

A week later, we invited them back to take a vocabulary test and to fill in a survey. The test was dynamically created for each participant and contained only words that fulfilled the following condition: each word listed had been reviewed by the user and had been guessed incorrectly when first encountered. This way we intended to make sure that we would test only new words that were unknown to the user before using the app. The test would list the word in the second language (L2) and ask the participant to type in its translation (L1), thus employing a *recognition* task. We specifically did not list multiple choices or had participants self-assess their performance like in the flashcard review mode, because we wanted to test whether the words had actually been memorized irrespective of the review mode. The survey consisted of the 11 5-point Likert-style questions listed in Table 3 asking about the subjective experience with the notifications, the learning in general and its perceived effectiveness. The last part of the survey was a text-based feedback form assessing what participants liked, disliked, and would improve about the app. Depending on the amount of words that had to be reviewed, filling in the survey took up to 15 minutes. The filling in of the survey was done under the supervision of an experimenter, in some cases where people could not come in after seven days, we sent them an online link to the survey with the explicit request not to use any auxiliary sources for filling in the vocabulary test. For the next seven days, the app changed its vocabulary review mode to flashcard or multiple choice respectively.

After another seven days of app usage had passed, we asked participants to come in for a final wrap-up session where we applied another vocabulary test. Words presented in this test fulfilled the same condition described above and were words that had only been encountered in the second week of the study. This way we made sure that each vocabulary test contained words reviewed in one review mode respectively. Finally, participants attended a final semi-structured interview containing 16 questions.

Hence, in total participants came in three times. For the initial setup and the intermediate session after one week, we compensated them with 5 EUR respectively, and another 10 EUR for completing the study after two weeks.

Phone Context	
ringer mode	mute, vibration, ringer
charging mode	unplugged, charging
battery status	related to usage intensity
display orientation	portrait or landscape mode, orientation changes
light sensor	changes in lightness allow us to derive whether phone is covered (carried in a pocket or taken out)
proximity sensor	phone in pocket
location	GPS data allows the inference of locations visited, and in-place vs. on-the-go states
motion	significant motion sensor, change of position
Phone Usage	
Calls	incoming, outgoing
SMS	incoming, outgoing
Notifications	received, dismissed, ignored, interacted with
Screen	on/off events
Unlocks	phone unlocks
Data usage	upload/download
Applications	applications in foreground, switches, usage duration

Table 2. Context data collected from phone sensors.

Results

In the following, we report our results with regard to the learning success, notification interaction, user ratings and qualitative feedback as given by participants.

Learning

After two weeks of casual usage, each participant had encountered on average 523.33 words ($Mdn = 331, SD = 477.14$). Of those, on average 222.75 ($Mdn = 196, SD = 127.2$) words were unique (not repeat-words) and 54.75 previously unknown ($Mdn = 39, SD = 53.91$). We categorized words as new or unknown when they had been marked as *"did not know"* or incorrectly answered when encountered for the first time. Participants completed on average 56 learning sessions ($Mdn = 45.5, SD = 31.26$). In total we recorded 557 learning sessions with an average of 10 words reviewed ($Mdn = 3, SD = 19.31$) in 3 word sets per session ($Mdn = 1, SD = 6.43$). A learning sessions took between 1sec and 20min ($M = 48.8sec, Mdn = 19.2, SD = 1.62min$). In the vocabulary tests, users translated 35 previously unknown words (64%) correctly into the foreign language despite never having had to type an actual word during the study. These 35 words ($Mdn = 20, SD = 26.3$) account for about 18 new words learned per week. The user who was exposed to the most new words (76) translated 32 correctly to L2.

Learning through notifications vs. through the app

Over the course of 2 weeks, participants reviewed on average 451 words ($Mdn = 257, SD = 473,7$) through the app, and 72 words ($Mdn = 70, SD = 40.3$) directly through notifications. On average, 39 (SD=25.08) learning sessions were completed by actively launching the app and 21 (SD=19.5) sessions by interacting through notifications. A Wilcoxon signed-rank test showed that statistically significant more reviews took place through an explicit app launch than through notification interaction ($Z = -2.510, p = 0.0.12$). Similarly, significantly more sessions were completed within the app than through notifications ($Z = -2.476, p = 0.013$). Participants spent on average 23.7 seconds ($Mdn = 11.4, SD = 92.9$) per notification session vs. 59.8 seconds ($Mdn = 27.5, SD = 97.3$) per app session. A Wilcoxon signed-rank test showed that notification sessions lasted significantly shorter ($Z = -2.510, p = 0.12$). In notification sessions, participants reviewed between 2 and 15 words ($M = 3.5, Mdn = 3, SD = 1.6$) and in app sessions between 1 and 216 words ($M = 13.3, Mdn = 5, SD = 22.5$) per learning session. Significantly more words were reviewed per app session than per notification session ($Z = -2.746, p = 0.006$).

Statement after 1st week	Median	SD
I feel the app helped me effectively improve my spanish vocabulary.	3	1
I liked that the app reminded me to review my vocabulary.	3	0.97
I liked learning words directly from the notification window.	3	1.22
I found the notifications well timed.	1	0.82
I found myself dismissing the notifications a lot.	2	1.01
The notifications interrupted my ongoing activities.	1	1.12
I liked being able to spend a few seconds on learning during the day.	4	0.42
I was able to review one set of words in a rather short amount of time.	3	0.73
The app allowed me to squeeze in reviewing vocabulary in between tasks.	3.5	0.84
I opened the app myself whenever I felt bored.	3	1.35
Statement after 2nd week	**Median**	**SD**
I think learning with MultipleChoice was very useful.	2	1.1
I think learning with FlashCards was very useful.	3	1.14
I preferred learning with MultipleChoice.	2	1.72
I preferred learning with FlashCards.	2.5	1.77
During this last week I noticed to get tired of the notifications.	2	1.2
During this last week I did not use the app as much.	2	1.29
I found myself learning with app at home a lot.	1.5	1.08
I found myself learning with app at work a lot.	1	1.4
I found myself learning with the app a lot when when I was on the go.	3	1.01
I found myself reviewing my vocabulary when other people were around.	1.5	0.99
I was very likely to review vocabulary when I found a notification was already there when I checked the phone.	3	1.17
I was very likely to review vocabulary when a notification was triggered while I was using the phone.	3	1.12
The app motivated me to use additional resources (e.g. books, courses, other apps,..) to improve my Spanish.	0.5	0.91

Table 3. Subjective user ratings in 5 point Likert-scales: 0 = strongly disagree, 4 = strongly agree.

Learning through flashcards vs. multiple choice

Participants reviewed on average 271 words ($Mdn = 117, SD = 375.3$) with flashcards, and 252 words ($Mdn = 137.5, SD = 255.3$) with multiple choice. On average, 27.8 learning sessions ($Mdn = 22.5, SD = 20.7$) were completed with flashcards and 28.4 ($Mdn = 28.5, SD = 16.2$) sessions with multiple choice. A Wilcoxon signed-rank test showed no statistical significance between the number of words reviewed ($Z = -.784, p = 0.433$), nor between the number of learning sessions completed per condition ($Z = -.550, p = 0.582$). Participants spent on average 51 seconds ($Mdn = 17.8, SD = 115.3$) per flashcard session compared to 46 seconds ($Mdn = 20.5, SD = 75.6$) per multiple choice session. In flashcard sessions, participants reviewed between 1 and 216 words ($M = 10.3, Mdn = 3, SD = 21.4$) and in multiple choice sessions between 1 and 121 words ($M = 10.3, Mdn = 3, SD = 17$) per learning session. A Wilcoxon signed-rank test showed no statistical significance for the average session duration, nor for the average number of words reviewed per condition ($Z = -.863, p = 0.388$). For the flashcard condition participants recalled a median of 10 (3 to 13) previously unknown words, and a median of 5 (1 to 12) in the multiple choice condition. A Wilcoxon signed-rank test yielded no statistically significant differences between conditions, ($Z = -1.023, p = 0.306$), despite a tendency of participants to perform better recalling words learned through flashcards.

Learning Locations

Of in total 556 recorded learning sessions across 2 weeks, we registered 105 (18.9%) sessions having been completed at home, while 59 (10.6%) having been completed at work. 209 (37.6%) took place at other (*i.e.*, neither work nor home) and 183 (32.9%) at unknown locations, where no location information could be acquired. Semantic locations were estimated

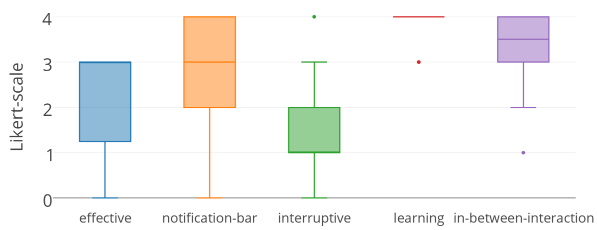

Figure 3. User ratings from 0 = strongly disagree to 4 = strongly agree taken from the first week's questionnaire, where we assessed learning *effectiveness*, utility of learning through the *notification bar*, the notifications' *interruptiveness*, *learning* in small chunks throughout the day, and being able to schedule sessions *in-between* tasks.

based on where users spent most of their time during waking/sleep hours. Participants reviewed on average 257 ($SD = 605.5$) words at home, 57 ($SD = 101.2$) at work, and 219 ($SD = 242.5$) in transit (*i.e.* "other"). We applied a Friedman test with a post-hoc analysis using a Wilcoxon signed-rank test with a Bonferroni-corrected significance level set at $p < 0.017$. There was a statistically significant difference in where participants reviewed the most words ($\chi^2 = 11.375, p = 0.003$). The post-hoc analysis revealed a statistically significant difference between the number of words reviewed in transit compared to at work ($Z = -2.937, p = 0.003$), but no significant difference between work and home ($Z = -2.043, p = 0.041$) or home and transit ($Z = -1.448, p = 0.148$). However, there is a tendency towards more words reviewed at home ($MD = 48$) than at work ($MD = 23$) and more reviewed in transit ($MD = 134$) than at home or work. Users learned more words in transit, regardless of the review type (flashcards. vs multiple choice).

We further looked at the locations, in which more words were reviewed through notifications or in-app sessions: With regard to interactive notifications, there was a statistically significant difference in locations ($\chi^2 = 9.475, p = 0.009$). The post-hoc analysis revealed a statistically significant difference between the number of words reviewed in transit compared to at work ($Z = -2.585, p = 0.01$), but no significant difference between work and home ($Z = -1.884, p = 0.06$) or home and transit ($Z = 1.139, p = 0.255$). There was a tendency towards more words reviewed through notifications at home ($MD = 5, SD = 35.8$) than at work ($MD = 1, SD = 13.6$) and more app reviews in transit ($MD = 18, SD = 32.1$) than at home or work. Similarly, there was a statistically significant difference in where participants used the app directly to review words ($\chi^2 = 11.460, p = 0.003$). The post-hoc analysis revealed a statistically significant difference between the number of words reviewed through the app in transit compared to at work ($Z = -2.844, p = 0.004$), but no significant difference between work and home ($Z = -1.578, p = 0.115$) or home and transit ($Z = -1.306, p = 0.191$). Here also there was a tendency towards more words reviewed through in-app usage at home ($MD = 13, SD = 507.6$) than at work ($MD = 12, SD = 94$) and more app reviews in transit ($MD = 89, SD = 193$) than at home or work.

Boredom as Trigger

To explore whether participants were more open to learn vocabulary when they were bored, we analyzed the data we collected in situations of presumed boredom and non-boredom. We consider the boredom estimation as independent variable of a quasi-experimental design with two conditions: bored and normal. The dependent variables were (1) *click-through rate*: the fraction of times that people clicked on the notifications created by *QuickLearn*, and (2) *words per session*: the number of words learned in a session. First, we cleaned up the data by removing participants who received less than 20 notifications and/or who clicked on less than 5 of them. For the lab study, 15 of the 19 participants passed this filter.

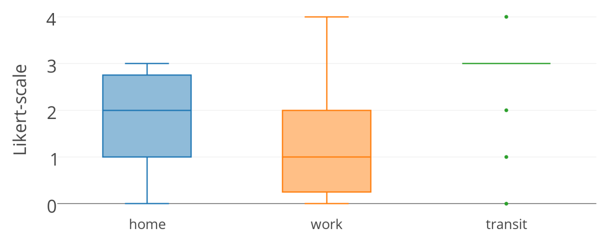

Figure 4. Results from subjective user assessment of *where* they found themselves engaging with microlearning sessions.

Click-Through Rates

In the data set of the lab study, we did not find differences between the fraction of notifications that the participants clicked in each condition. In the *bored* condition, scores ranged from 6% to 87% ($M = 31.2$, $SD = 22.4$). In the *normal* condition, scores ranged from 2% to 88% ($M = 35.2$, $SD = 25.4$). A Levene's test showed that variance of the scores of the two conditions were sufficiently equal to use parametric tests ($F(1,28) = .17, p = .68$). A dependent t-test revealed no significant effect ($t(14) = -.8, p = .43$).

Words Per Session

We did not find differences between the average word count that the participants learned per session. In the *bored* condition, scores ranged from 0 to 8 words ($M = 6.07$, $SD = 4.33$). In the *normal* condition, scores ranged from 3 to 74 words ($M = 12.87$, $SD = 17.89$). A Levene's test showed that variance of the scores of the two conditions were sufficiently equal to use parametric tests ($F(1,28) = 1.47, p = .20$). A dependent t-test revealed no significant effect ($t(14) = -1.69, p = .11$).

User Ratings

Table 3 contains the results from the intermediate and final survey taken after each week of app usage. Questions differed for each week except for the assessment of how participants rated the current learning mode. The ratings were given on a 5-point Likert scale with 0 indication "I strongly disagree" and 4 "I strongly agree". Results show that the app was very well received, in particular the notifications acting as reminders to review vocabulary ($Mdn = 3, SD = 0.97$) and the interaction possibility through the notification drawer ($Mdn = 3, SD = 1.22$). Fig. 3 shows box plots for a selection of questions participants were asked after the first week. *Effective* refers to whether participants were able to effectively improve their second language skills, *notification-bar* visualizes users' overall affirmation of liking to learn words directly through interactive notifications, thereby not being too *interruptive*. Participants strongly agreed on the statement that they enjoyed being able to spend a few seconds on *learning* during the day ($Mdn = 4, SD = 0.42$). Interaction time was perceived as very low as words could be reviewed in a rather short amount of time ($Mdn = 3, SD = 0.73$) and *in-between* tasks ($Mdn = 3.5, SD = 0.84$). The novelty effect of introducing the app seems to be limited as participants overly disagreed with the statement that they got tired of notifications ($Mdn = 2, SD = 1.2$), neither did they stop using the app as much ($Mdn = 2, SD = 1.29$). Using the app on-the-go ($Mdn = 3, SD = 1.01$) seemed to be the preferred mode as compared to at home ($Mdn = 1.5, SD = 1.08$) or at work ($Mdn = 1, SD = 1.4$), as depicted in Fig. 4.

There was a statistically significant difference in where participants reported to have used *QuickLearn* ($\chi^2 = 9.333, p = 0.009$). No statistically significant differences were detected between work and home ($Z = -0.677, p = 0.498$), transit and home ($Z = -1.840, p = 0.066$), or transit and work ($Z = -2.271, p = 0.023$). However, there was a tendency of more people indicating transit to be the preferred place where they found themselves reviewing vocabulary.

Qualitative Measures

After each week, participants filled in a survey, which - besides testing vocabulary recall and collecting the subjective user ratings - contained free-form questions, where participants could leave qualitative feedback.

Most participants (9) positively commented on the ease of word access and interaction through the notification bar: P1 stated *"I didn't consciously realize that I was learning. Since it was such short periods of time."* P6 commented on the notifications' reminder function: *"I liked that it reminded me to invest some time. A reminder to do something useful."* Notification interactions were welcomed during idle moments (8), e.g. while *"fooling around on my phone."* (P1), *"after finishing a task on the phone"* (P7) or *"whenever I was alone or doing something boring."* (P8). Most participant (9) explicitly rejected the notion that notifications for microlearning were perceived as a disruption. *"It's not disruptive and allows me to continue easily with what I was doing"* (P14) while another observed *"you don't have to leave the current app, therefore you don't lose the focus"* (P7). Well received was the combination of proactive reminders and the possibility to launch the app explicitly: *"I can open it whenever I want, but at the same time, it reminds me to keep learning"* (P16). Participants appreciated the brevity of interaction: *"Takes pretty much no time to learn something new"* (P18). Short interactions conveyed a feeling of accomplishment: *"I was motivated to use the app because it was easy and not time-consuming"* (P13) and *"sets of 3 words give me a sense of achievement"* (P15). Participants found opportune moments for learning when they were, for instance, *"pulling out the phone to just check the time and there was a notification there, I would do a couple of words"* (P1). Users welcomed notifications while engaged in information consumption rather than actively doing something on the phone: *"when the browser was opened and I was reading unspecifically, then it would fit. It did not fit when I was concretely doing something, such as calendar entries"* (P10). Where participants were most likely to review vocabulary, a majority indicated public transportation and while in-transit to be most fitting, also during waiting situations, *e.g.*, between breaks at the gym or at the bus stop.

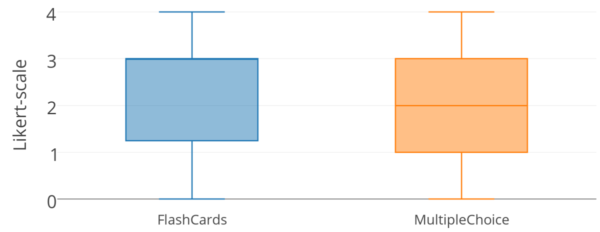

Figure 5. Participants' subjective rating of presentation modes: Flashcards vs. MultpleChoice on a 0 - 4 Likert scale.

Six participants found themselves explicitly opening the app in waiting situations: *"Mostly when I was on the train or when I was waiting. Wasting time."* (P5), and *"when I am on the lift alone. When I am waiting, queuing for some reason: supermarket, doctor's appointment [...]"* (P8). 5 participants welcomed the reminders while they were using the phone (*"While reading emails I was also doing some words on the side"* (P6)) while others (4) preferred reminders to be triggered when they were not paying attention to their phone (*"When I was not using the phone, or right after I was using the phone"* (P10)).

Multiple participants (5) complained about the battery drain of the app. Since we collected a great variety of sensor data, this affected battery drainage in a noticeable way. Multiple choice seemed to polarize participants' opinions. Some liked its *"game-like character"* (P10), but a majority (9) saw problems with cognates, *i.e.* words that look similar in different languages, and the problem of being able to guess a lot of the words: *"My level of spanish is not really high, but multiple choice often allows me to guess the answer of words I have never seen before, given my knowledge of French"* (P14) and *"Multiple choice options are certainly not optimal to learn vocabulary, though okay for the quick interactions"* (P16). However, as Figure 5 shows, there was barely a difference between learners preferring Flashcards over MultipleChoice. Some participants (4) commented negatively on the timing of the notifications, for example when they had just checked their phone only to be reminded to review their vocabulary after putting it away: *"sometimes notifications were delayed and came after I just checked the time"* (P17). As for app improvements, 6 participants mentioned the limited choice of vocabulary and how they would like to add custom words or choose their own word topics. As for review modes, 4 participants requested text entry quizzes, speech output for pronunciation, sentence examples, or adding articles to nouns. One feature consistently requested was a way to track their learning progress.

IN-THE-WILD STUDY

To increase our body of context sensor data we released *QuickLearn* for free on Google Play.

Methodology

This part of the study comprises data from users in the wild. We collected sensor data as well as boredom prediction results, and collected subjective app feedback through an in-app surveys.

Participants

Besides releasing the app on Google Play, we launched a corresponding website and Facebook page to be able to promote the app through social channels. One year after the app release in September 2015, we registered 83 active users, 19 of which being lab-study participants who decided to keep using the app for a while beyond the study. 28 indicated to be female, 55 to be male. The reported mean age was 31 ($SD = 11.7$) and according to the most frequent device locales (35 en-US, 9 de-DE, 18 en-GB, 27 other) and timezones, most users were from the U.S., Germany, and UK. 20 selected German as mother tongue, 39 English, 11 arabic, 7 French, and 6 Spanish. 25 chose to study English, 23 Spanish, 13 German, 18 French, and 4 Arabic.

Procedure

After downloading the app, users went through the same setup procedure consisting of agreeing to the terms on the consent form and granting the application access to the Android Accessibility Service as well as to notifications. Users were asked to provide their age, gender, language preference, and current proficiency. The display modality was randomly assigned when the app was first launched. In contrast to the controlled study, the modality changed every three days.

Results

In total, 14,632 notifications were triggered, of which 2765 (18.9%) were interacted with. Further, we registered 1435 app launches resulting in more than 19,000 words reviewed.

Boredom as Trigger

After cleaning up the data as was done for the lab study we ended up with 16 participants who provided valid data sets.

Click-Through Rates

In the data set of the in-the-wild study, did not find statistically significant differences between the click-through rates either. In the *bored* condition, scores ranged from 2% to 81% ($M = 20.69$, $SD = 21.91$). In the *normal* condition, scores ranged from 0% to 50% ($M = 18.75$, $SD = 15.92$). A Levene's test showed that variance of the scores of the two conditions were sufficiently equal to use parametric tests ($F(1,30) = .33, p = .57$). A dependent t-test revealed no significant differences between the groups ($t(15) = .38, p = .71$).

Words Per Session

In the data set of the in-the-wild study, we neither found differences between the click-through rates. In the *bored* condition, scores ranged from 1 to 49 words ($M = 7.71$, $SD = 11.49$). In the *normal* condition, scores ranged from 2 to 21 words ($M = 6.71$, $SD = 5.02$). A Levene's test showed that variance of the scores of the two conditions were sufficiently equal to use parametric tests ($F(1,32) = .55, p = .46$). A dependent t-test revealed no significant differences between the groups ($t(16) = .49, p = .63$). In summary, we found no evidence to indicate whether phases of boredom were better or worse for microlearning.

Opportune Moments for Language Learning

To answer RQ1: "*Which context factors are most opportune for language learning?*", the *QuickLearn* application collected usage logs as depicted in Table 2. From this log data, we analyzed 35 features from 14,632 instances of notifications sent to 37 different participants (for the full list of features, see [28]). We had previously filtered out participants who received less than 20 notifications and interacted with less than 5 of those. With the goal to predict whether a participant will engage in a learning session triggered by a notification, we ranked the features by building a model with XGBoost [11]. XGBoost is a tree algorithm with gradient boosting, which creates a ranking by calculating each feature's importance in the form of an F-score. The importance is defined by the fraction of times a feature is chosen to be used in a tree. We applied XGBoost's standard configuration, with one exception: since the standard classifier strongly tended towards a positive prediction (*i.e.*, an opportune moment for learning) while only 18.9% of the notifications had led to an actual learning session, we reduced *scale_pos_weight* by factor 0.25, thus increasing the penalty of false positives. The resulting tree model achieved a precision of 0.430, a recall of 0.712, and an F1-score of 0.526. Compared to the baseline, notifications posted by using such a classifier would have boosted the conversion rate 2.21 times. Table 4 shows the importance of the 10 best predictors as reported by the XGBoost model. We further analyzed the correlations between those 10 features and the ground truth via Spearman's Rank correlation. Positive predictions of the classifier had significant, non-negligible correlations with:

Feature	Importance
time since last unlock	0.2163
age	0.1450
time since last incoming phone call was denied	0.0814
time since last notification	0.0687
number of apps used in time window (5mins)	0.0662
time since notification center was last accessed	0.0662
number of phone unlocks in time window (5mins)	0.0433
last notification category: NONE	0.0407
time since last incoming phone call was accepted	0.0407
time since last phone call was placed	0.0382

Table 4. Feature importance as reported by the XGBoost model. Importance values depict F-scores summing up how many times each feature is split on (*i.e.*, in what percentage of splits the particular feature is being used).

- Less time passed since the phone had been last unlocked ($r = -0.303, p < 0.001$).
- Higher age ($r = 0.144, p < 0.001$).
- More apps used during the last 5 minutes ($r = 0.213, p < 0.001$).
- Less time passed since the last access of the notification center ($r = -0.186, p < 0.001$).
- More screen unlocks during the last 5 minutes ($r = 0.264, p < 0.001$).
- Not having received a notification in the last 5 minutes ($r = 0.109, p < 0.001$).

Primarily, these results indicate that participants were more likely to engage in quick learning sessions when they had interacted with the phone more recently. This finding is in line with previous work that showed that more intense phone use correlates with boredom [28] and stimulation-seeking, *ritualistic* phone use [23].

DISCUSSION AND LIMITATIONS

Combining the two studies presented we provide empirical evidence that using a proactive mobile learning app facilitates microlearning and is an effective way of spacing out learning sessions throughout the day. This is mainly due to users' tendency to review vocabulary on-the-go.

We were able to shed some light on context factors, in which microlearning was found feasible (**RQ1**), namely when in transit as well as in idle moments, such as waiting situations. The design of the app has supported this notion of learning on-the-go by providing interactive notifications. Learning sessions through notifications were found to be shorter than explicit app launches. Therefore, notifications seem to offer great possibilities for engaging users in quick learning tasks that only take up a few seconds, however as to its effectiveness compared to dedicated app launches we were not fully able to answer **RQ2**.

Also, despite a tendency towards flashcards being more effective (as in new words learned and remembered) than multiple choice questions, a statistically significant difference could not be determined. In a mobile setting and with regard to users often finding themselves killing time with the phone positive aspects of MultipleChoice seem to be the game-like mechanism and the pleasure of guessing as stated by study participants (**RQ3**).

When notifications were triggered out of predicted boredom situations as suggested by our previous work [28], we were in contrast not able to detect a statistical significant difference when compared to non-bored situations. Therefore, the study did not find conclusive evidence whether people are more likely to engage in micro-learning tasks depending on whether they are bored or not (**RQ4**). This finding is in contrast to the mentioned previous work, where participants clicked significantly more often on suggested articles. The difference in the findings might be explained by the difference in the suggested content: easy-to-consume articles versus a mentally demanding learning task. While boredom is characterized as a state in which people seek stimulation - hence might hypothesized to be open for the suggestion to learn vocabulary - it is also characterized as the inability to find stimulation in current activities [17]. While people who are currently experiencing boredom might in theory have the time to engage in microlearning, they may not be able to get up the energy or may not expect the stimulation they desire from this mentally demanding task.

By analyzing phone usage context, we provide an assessment of opportune moments for scheduling learning sessions. Proactive learning reminders were especially welcomed when the phone had been recently used as the correlations with phone unlock and app usage patterns show. Also, the higher participants' age, the more likely they would engage with learning content. Unfortunately, the nature of the in-the-wild-study and the limited knowledge of our users prevents us from drawing clear causal relationships here. However, by updating the boredom classifier with our innate model we would be able to double the compliance rate of the app's learning reminders.

For a 1-year in-the-wild study we registered relatively low usage numbers. Besides the fact that we did not actively promote the app, its data settings and privacy implications probably had a deterrent effect. While *QuickLearn* does not suffice to teach a foreign language to a learner, we focused on vocabulary learning as a proof of concept. However, due to the well-received notion of microlearning sessions throughout the day, such technology could pose as a complement to existing learning techniques. For example, combined with an e-Reader or audio book application, new words from articles the user has read could automatically be placed into the learning cue and spaced out in systematic repetitions in order to help users eventually commit them to long-term memory.

CONCLUSION

Microlearning in combination with ubiquitous technologies bears great potential for learners who naturally lack time and motivation to tackle a daunting task, such as learning a foreign language. In our work we explore different aspects of microlearning as a tool on-the-go, namely presentation mode, automatic reminders, interaction modalities, and learning context. Therefore, we created the Android app *QuickLearn*, through which users learned on average 18 new words per week by being exposed to about 37 words per day: whether in relatively quick word reviews through the notification bar or through longer sessions through explicit app launches.

Our results show that people are more open to engage in quick learning sessions when they are mobile. Notification interactions and in-app sessions registered more word reviews when people were in transit compared to at home or at work. Microlearning systems should therefore take advantage of interactive notifications and design for learning sessions to be completed while on-the-go. While boredom does not seem to correlate with opportune moments for learning, we showed that training a classifier with context information retrieved from the phone's sensors can increase compliance rates.

We see microlearning as a complement for existing learning techniques: throughout the day we read various articles, make notes, and try to remember new information encountered. As part of a greater knowledge management system, small learning units can be provided to users in idle moments. In this work we showed the feasibility of such an approach, but the possibilities are far greater when, for example, resources beyond foreign language vocabulary are added. Through smart scheduling of content repetitions, mobile devices can help users make use of idle moments, commit knowledge to long-term memory and therefore increase their personal effectiveness.

ACKNOWLEDGEMENTS

We would like to thank the participants of our studies and acknowledge the financial support of the Future and Emerging Technologies (FET) programme within the 7th Framework Programme for Research of the European Commission, under FET grant number: 612933 (RECALL) and the financial support of the NSF under the grant number IIA-1358096.

REFERENCES

1. Daniel Lee Ashbrook. 2010. Enabling mobile microinteractions. (2010).
2. Harry P Bahrick and Elizabeth Phelphs. 1987. Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13, 2 (1987), 344.
3. Emrah Baki Basoglu and Omur Akdemir. 2010. A Comparison of Undergraduate Students' English Vocabulary Learning: Using Mobile Phones and Flash Cards. *Turkish Online Journal of Educational Technology - TOJET* 9, 3 (July 2010), 1–7. http://eric.ed.gov/?id=EJ898010http://files.eric.ed.gov/fulltext/EJ898010.pdf
4. Jennifer S. Beaudin, Stephen S. Intille, Emmanuel Munguia Tapia, Randy Rockinson, and Margaret E. Morris. 2007. Context-Sensitive Microlearning of Foreign Language Vocabulary on a Mobile Device. Springer Berlin Heidelberg, 55–72. http://link.springer.com/chapter/10.1007/978-3-540-76652-0_4
5. Chihab BenMoussa. 2003. Workers on the move: New opportunities through mobile commerce. *Stockholm mobility roundtable* (2003), 22–23.
6. Peter A. Bruck, Luvai Motiwalla, and Florian Foerster. 2012. Mobile Learning with Micro-content: A Framework and Evaluation. Bled, Slovenia. http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1041&context=bled2012
7. Carrie J. Cai, Philip J. Guo, James Glass, and Robert C. Miller. 2014. Wait-learning: Leveraging Conversational Dead Time for Second Language Education *(CHI EA '14)*. ACM, New York, NY, USA, 2239–2244. DOI: http://dx.doi.org/10.1145/2559206.2581183
8. Shana K. Carpenter and Edward L. DeLosh. 2005. Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology* 19, 5 (2005), 619–636. DOI: http://dx.doi.org/10.1002/acp.1101
9. Nadire Cavus and Dogan Ibrahim. 2009. m-Learning: An experiment in using SMS to support learning new English language words. *British Journal of Educational Technology* 40, 1 (Jan. 2009), 78–91. DOI: http://dx.doi.org/10.1111/j.1467-8535.2007.00801.x
10. Chih-Ming Chen and Ching-Ju Chung. 2008. Personalized mobile English vocabulary learning system based on item response theory and learning memory cycle. *Computers & Education* 51, 2 (Sept. 2008), 624–645. DOI: http://dx.doi.org/10.1016/j.compedu.2007.06.011
11. T. Chen and C. Guestrin. 2016. XGBoost: a scalable tree boosting system. In *Proc. of the ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD)*. 785–794.
12. William L Cull and others. 2000. Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology* 14, 3 (2000), 215–235.
13. David Dearman and Khai Truong. 2012. Evaluating the Implicit Acquisition of Second Language Vocabulary Using a Live Wallpaper *(CHI '12)*. ACM, New York, NY, USA, 1391–1400. DOI: http://dx.doi.org/10.1145/2207676.2208598
14. Frank N Dempster. 1987. Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology* 79, 2 (1987), 162.
15. Anind K. Dey, Katarzyna Wac, Denzil Ferreira, Kevin Tassini, Jin-Hyuk Hong, and Julian Ramos. 2011. Getting Closer: An Empirical Investigation of the Proximity of User to Their Smart Phones. In *Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp '11)*. ACM, New York, NY, USA, 163–172. DOI: http://dx.doi.org/10.1145/2030112.2030135
16. Tilman Dingler and Martin Pielot. 2015. I'Ll Be There for You: Quantifying Attentiveness Towards Mobile Messaging. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '15)*. ACM, New York, NY, USA, 1–5. DOI: http://dx.doi.org/10.1145/2785830.2785840
17. John D. Eastwood, Alexandra Frischen, Mark J. Fenske, and Daniel Smilek. 2012. The Unengaged Mind: Defining Boredom in Terms of Attention. *Perspectives on Psychological Science* 7, 5 (2012), 482–495.
18. Hermann Ebbinghaus. 1913. *Memory: A contribution to experimental psychology*. Number 3. University Microfilms.
19. Darren Edge, Stephen Fitchett, Michael Whitney, and James Landay. 2012. MemReflex: Adaptive Flashcards for Mobile Microlearning *(MobileHCI '12)*. ACM, New York, NY, USA, 431–440. DOI: http://dx.doi.org/10.1145/2371574.2371641
20. Darren Edge, Elly Searle, Kevin Chiu, Jing Zhao, and James A. Landay. 2011. MicroMandarin: Mobile Language Learning in Context *(CHI '11)*. ACM, New York, NY, USA, 3169–3178. DOI: http://dx.doi.org/10.1145/1978942.1979413
21. Joel E. Fischer, Chris Greenhalgh, and Steve Benford. Investigating Episodes of Mobile Phone Activity As Indicators of Opportune Moments to Deliver Notifications. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (2011) *(MobileHCI '11)*. ACM, 181–190. DOI: http://dx.doi.org/10.1145/2037373.2037402
22. Robert Godwin-Jones. 2010. Emerging technologies from memory palaces to spacing algorithms: approaches to secondlanguage vocabulary learning. *Language, Learning & Technology* 14, 2 (2010), 4.

23. Alexis Hiniker, Shwetak N. Patel, Tadayoshi Kohno, and Julie A. Kientz. 2016. Why Would You Do That? Predicting the Uses and Gratifications Behind Smartphone-usage Behaviors. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 634–645. DOI: http://dx.doi.org/10.1145/2971648.2971762

24. Shamsi T. Iqbal and Brian P. Bailey. Effects of Intelligent Notification Management on Users and Their Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2008) *(CHI '08)*. ACM, 93–102. DOI: http://dx.doi.org/10.1145/1357054.1357070

25. Agah Tugrul Korucu and Ayse Alkan. 2011. Differences between m-learning (mobile learning) and e-learning, basic terminology and usage of m-learning in education. *Procedia-Social and Behavioral Sciences* 15 (2011), 1925–1930.

26. William E Nagy and Patricia A Herman. 1987. Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. *The nature of vocabulary acquisition* 19 (1987), 35.

27. Tatsuya Nakata. 2008. English vocabulary learning with word lists, word cards and computers: Implications from cognitive psychology research for optimal spaced learning. *ReCALL* 20, 01 (2008), 3–20. DOI: http://dx.doi.org/10.1017/S0958344008000219

28. Martin Pielot, Tilman Dingler, Jose San Pedro, and Nuria Oliver. 2015. When Attention is Not Scarce - Detecting Boredom from Mobile Phone Usage. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 825–836. DOI: http://dx.doi.org/10.1145/2750858.2804252

29. B. Poppinga, W. Heuten, and S. Boll. 2014. Sensor-Based Identification of Opportune Moments for Triggering Notifications. *IEEE Pervasive Computing* 13, 1 (Jan. 2014), 22–29. DOI: http://dx.doi.org/10.1109/MPRV.2014.15

30. John Read. 2000. *Assessing vocabulary*. Cambridge university press Cambridge.

31. Patricia Thornton and Chris Houser. 2005. Using mobile phones in English education in Japan. *Journal of Computer Assisted Learning* 21, 3 (June 2005), 217–228. DOI: http://dx.doi.org/10.1111/j.1365-2729.2005.00129.x

32. Andrew Trusty and Khai N. Truong. 2011. Augmenting the Web for Second Language Vocabulary Learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 3179–3188. DOI: http://dx.doi.org/10.1145/1978942.1979414

33. Maria Virvou and Eythimios Alepis. 2005. Mobile educational features in authoring tools for personalised tutoring. *Computers & Education* 44, 1 (2005), 53–68.

34. Stuart Webb. The Effects of Repetition on Vocabulary Knowledge. *Applied Linguistics* 28, 1 (????), 46–65. DOI: http://dx.doi.org/10.1093/applin/aml048

35. SungHyuk Yoon, Sang-su Lee, Jae-myung Lee, and KunPyo Lee. 2014. Understanding notification stress of smartphone messenger app. In *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems - CHI EA '14 (CHI EA '14)*. ACM, New York, NY, USA, 1735–1740. DOI: http://dx.doi.org/10.1145/2559206.2581167