

Research Article

Christina Schneegass*, Sophia Sigethy, Teodora Mitrevska, Malin Eiband, and Daniel Buschek

UnlockLearning – Investigating the Integration of Vocabulary Learning Tasks into the Smartphone Authentication Process

<https://doi.org/10.1515/icom-2021-0037>

Abstract: Frequent repetition of vocabulary is essential for effective language learning. To increase exposure to learning content, this work explores the integration of vocabulary tasks into the smartphone authentication process. We present the design and initial user experience evaluation of twelve prototypes, which explored three learning tasks and four common authentication types. In a three-week within-subject field study, we compared the most promising concept as mobile language learning (MLL) applications to two baselines: We designed a novel (1) *UnlockApp* that presents a vocabulary task with each authentication event, nudging users towards short frequent learning session. We compare it with a (2) *NotificationApp* that displays vocabulary tasks in a push notification in the status bar, which is always visible but learning needs to be user-initiated, and a (3) *StandardApp* that requires users to start in-app learning actively. Our study is the first to directly compare these embedding concepts for MLL, showing that integrating vocabulary learning into everyday smartphone interactions via *UnlockApp* and *NotificationApp* increases the number of answers. However, users show individual subjective preferences. Based on our results, we discuss the trade-off between higher content exposure and disturbance, and the related challenges and opportunities of embedding learning seamlessly into everyday mobile interactions.

***Corresponding author: Christina Schneegass,**
Ludwig-Maximilians-Universität München, Media Informatics,
Frauenlobstr. 7a, Munich, Bavaria, Germany, e-mail:
christina.schneegass@gmail.com

Sophia Sigethy, Teodora Mitrevska, Malin Eiband,
Ludwig-Maximilians-Universität München, Media Informatics,
Frauenlobstr. 7a, Munich, Bavaria, Germany, e-mails:
sophia@sigethy.de, tmitrevska97@gmail.com,
malin.eiband@ifi.lmu.de

Daniel Buschek, University of Bayreuth, Department of Computer
Science, Research Group HCI + AI, Universitätsstrasse 30, Bayreuth,
Bayern, Germany, e-mail: Daniel.Buschek@uni-bayreuth.de

Keywords: Mobile Language Learning, Embedded Interaction, Micro Interactions, Authentication

1 Introduction

Language learning is a long-term task that requires users to engage with the learning content on a regular basis. Repeating content frequently can consolidate the memory of it and thereby counteract forgetting [40]. Prior research has explored a variety of options for technology to support this process. The concept of micro-learning utilises the presentation of small chunks of information, i. e. micro-content units, through micro-interactions and has proven its potential for effective vocabulary acquisition and recall [7]. The sheer ubiquity of mobile devices makes them already a feasible tool for the presentation of learning content. We interact with our smartphones on a regular basis, frequently over the course of the day [17, 21].

Besides commonly available learning applications, research has evaluated concepts that give learning tasks more visibility during everyday smartphone interactions, such as presenting content as interactive push notifications [15] or as lockscreen wallpapers [13]. These concepts aim to increase users' exposure to the learning content by maximising the visibility of the tasks [13] or by exploiting idle moments such as waiting situations [10, 15]. We propose to embed the learning action even further into the smartphone interaction by combining it with the authentication process – a task we perform numerous times a day without any further use [20, 29].

In this paper, we evaluated the user experience of twelve prototypes exploring all combinations of three learning tasks with four interaction input types that are commonly applied for smartphone authentication. Based on the users' feedback and implementation feasibility, we selected the most promising prototype for further evaluation. By reporting and discussing this in-depth study of concepts, this article is an extended version of Schneegass, Sigethy, Eiband & Buschek (2021) [42].

We conducted a three-week within-subject field study to compare users' interactions and exposure to learning content using our new authentication learning app with two concept based on prior work. The three conditions differ in their level of embeddedness into the smartphone interaction and are described as follows:

1. First, our novel concept called *UnlockApp* extends the concept of Dearman and Truong's [13] learning wallpaper by connecting a simple multiple-choice vocabulary task that the user can answer by a single button press with the phone unlock action. Although users can still dismiss the learning task, this concept is strongly embedded into the unlock action as the app initiates the learning task. We aim to nudge users toward more frequent learning by lowering the threshold for users to interact with the content.
2. Second, our notification learning app (*NotificationApp*) presents a learning task in a continuously shown notification. The notification increases learning task's visibility during interactions with the lockscreen or status bar but still requires users to actively initiate learning.
3. Third, we implement a *StandardApp*, following a baseline self-contained and self-initiated learning design.

In our in-the-wild evaluation, we investigate how the different degrees of embedding and self-initiated vs app-initiated exposures impact interactions with the learning content over the day and how users perceive this form of learning in their everyday use.

The results show that the *UnlockApp* and *NotificationApp* designs, compared to the *StandardApp*, significantly increase the number of learning tasks users answer per day. The *UnlockApp*, in particular, further distributes the learning more evenly across the day. Interestingly, we observe a great variety in participants' subjective perceptions of the three concepts. While some people like that the *UnlockApp* nudges them to engage with the vocabulary regularly, others favored the *NotificationApp*, arguing that it is less distracting, especially when unlocking their phone with a specific goal in mind. In contrast, when participants had time to spare, they enjoyed long learning streaks with the *StandardApp*.

Based on these comparisons, we discuss challenges and opportunities of embedding learning tasks more ubiquitously into peoples' everyday lives. In particular, combinations of features from the different concepts and offering further personalisation options can target individual user preferences and balance the trade-off between nudging users to learn and disturbing routine smartphone tasks.

2 Related Work

2.1 Mobile & Micro-Learning

The ubiquity of mobile devices fosters the shift from learning being a classroom-based activity to an activity that is embedded into our daily life. By applying the micro-learning paradigm, learning content is broken down into small bits of information and presented to learners using simple interaction concepts [7]. Hereby, the learning task is spread out over time. Especially for vocabulary acquisition, a higher number of short interactions should be favored over one long learning streak [12]. By spacing out the learning content, prior work has demonstrated that learners could memorise vocabulary content better than immediate successive presentation [14]. The concept of "spaced-repetition" suggests that repeating learning content according to a repetitive schedule helps consolidate the memory of it [1, 43]. This schedule targets the transient forgetting of information that naturally occurs after a certain amount of time of the information is not rehearsed or applied [40]. As MLL apps require self-directed learning, it needs to be taken into account that the success of autonomous learning and the adherence to rehearsal schedules depends on the consideration of various dimensions (i. e. context of learners, learner interest and motivation, language proficiency, etc.) [26, 28]. Thus, the overall goal of micro-learning is to maximise the degree of exposure with the learning content due to increased presentation and, thus, frequent repetition of content. In particular for the teaching of languages, the application of the micro-learning approach has proven to improve vocabulary acquisition and recall [10, 16, 45]. Prior work presented different ideas for realizing micro-learning in mobile learning. For example, Dearman and Truong (2012) [13] implemented a mobile language learning application that can display learning tasks (i. e. vocabulary translations with three multiple-choice options) on the lockscreen. The authors show that users frequently interacted with the application and improved their knowledge of the language over the course of this study. Dingler et al. (2017) [15] achieved similar results by implementing and evaluating a vocabulary application that presents tasks in push-notification. Their app "QuickLearn" detects opportune moments for the presentation of the learning content, in particular, when users are bored and use their smartphone without a specific purpose. Dingler and colleagues concluded that their application increased the number of "quick" learning sessions on the go, which was appreciated by their participants.

2.2 Smartphone Usage

Users' interactions with their mobile devices can range in their level of intensity. Prior work has shown that half of our interactions with our smartphones take less than 30 seconds, with only one in ten interactions exceeding four minutes [48]. In particular, many interactions do not even reach the threshold of fifteen seconds, what Ferreira et al. (2014) [17] defined as so-called "micro-usage" situations. For example, users briefly reply to a message, check their phone for new push notifications, or want to know the time. For many of those interactions, users do not unlock their device [20, 21] or only use one application after unlocking [29]. However, unlocking the mobile device is still an action that happens frequently throughout the day. Mahfouz, Muslukhov and Beznosov (2016) [29] report that users unlock their phone on average 46 times, while other studies report an average of between 25 [20] and 47 unlocks [19] per day per participant. As this interaction has no other purpose than the authentication, prior work suggests combining the authentication event with a micro-interaction task, in particular, when the slide-to-unlock authentication is used. Truong, Shihpar and Wigdor (2014) [44] emphasised that authentication gestures serve no purpose other than to unlock the phone and outline their potential to serve other tasks such as data collection. Therefore, it has been proposed as a quick and easy method to enter journaling data [49], nutrition tracking [23], or sleep tracking [11].

2.3 Authentication Methods and Their Prevalence

In general, smartphone authentication mechanisms can be divided into four categories according to Wang et al. (2020) [47]:

1. **Knowledge-based:** Require the user to remember certain information, either text-based or graphical. *Example:* Passwords, PINs, patterns, or graphical passwords
2. **Physiological Biometric-based:** Makes use of unique identifying biological traits of users such as their fingerprint or facial features. This form of authentication can be performed upon explicit prompt or implicit assessment. *Example:* Fingerprint, voice, and face recognition
3. **Behavioral Biometric-based:** Captures users' unique behavioral patterns or characteristics. *Example:* Gait recognition, tapping behavior, hand gesture

4. **Mixed Models / Multi-factor:** Combination of two or more approaches to increase security. *Example:* Integrating knowledge factors with biometric authentication

In 2016, Malkin et al. (2017) [30] performed an online survey with more than 8000 smartphone users about their unlocking mechanism. While around one third (32.1 %) of the participants stated to use slide-to-unlock, 32.5 % used pattern authentication, 18.7 % PIN, and each around 7 % used a password or biometric authentication (e. g. fingerprint). With a decrease in popularity of passwords for mobile authentication due to an increased number of passwords one needs to remember and a comparably high effort and time to input them, visual passwords (i. e. patterns) gained prevalence in the first decade of the 20th century [46]. However, these patterns have a more limited diversity than passwords and are prone to guessing or observer attacks. With the implementation of fingerprint sensors in smartphones, implicit biometric authentication became popular. Although the prevalence is steadily increasing, common techniques (i. e. pattern, PIN, swipe) are still in use for many reasons. For example, users want to protect their privacy by not digitalizing their biometric information such as fingerprints. Further, implicit authentication can be prone to failure [2, 4, 36], with especially the fingerprint entry showing an increased number of errors [37]. Reasons for increased error rate can be reduced sensitivity of a fingerprint sensor due to dirt or water, making knowledge-based methods such as PIN, pattern, or passwords a valid complementary "fallback" method.

For increased security, knowledge-based mechanisms can be included in multi-factor authentication schemes. Thus, even with implicit authentication such as fingerprint input or FaceUnlock becoming more common, PINs, passwords, and swipe-to-unlock do not become obsolete. Findings from studies such as by Qiu et al. (2019) [37] show diverse user preferences and authentication behavior, thus, strengthening the argument that offering multiple different unlocking mechanisms for smartphones is necessary.

Therefore, the first part of this chapter will present an exploration and evaluation of embedding different learning tasks into a diverse set of smartphone authentication methods, i. e. PIN, pattern, swipe-to-unlock, or fingerprint input using fingerprint sensor gestures.

2.4 Summary & Research Gap

In summary, micro-learning applications for vocabulary-based language learning have proven to increase vocab-

ulary recall [10, 16, 45] and have shown great potential for pervasively embedding learning into everyday smartphone usage [13, 15]. Yet, it remains unclear how best to embed the learning task into everyday device use. This motivates our idea and investigation of our new concept of integrating learning directly into the authentication process, by displaying a vocabulary task every time the phone is unlocked using any form of authentication mechanism. Furthermore, to the best of our knowledge, no comparative analysis has been performed yet to investigate potential differences between the proposed pervasive presentation of vocabulary and a “standard” in-app learning application that requires the user to actively initiate learning. For example, while Dearman and Truong (2012) [13] showed a knowledge gain through their wallpaper app, they do not implement a “standard” app as a control condition.

3 Concept: Integrating Learning Tasks Into Authentication Methods

In the following, we outline the learning tasks and the authentication methods we explored in more detail. As this works’ idea aims for the embedding of learning tasks into the authentication action – creating one seamless interaction – we break down both the learning and authentication actions into their individual interactions (i. e. discrete vs. continuous input, single vs. multiple inputs).

3.1 Authentication Interaction

We characterised the authentication methods from an interaction perspective. Thus, it was less important what the user actually inputs (characters, numbers, etc.) but rather what type of interaction is required. We distinguish the following four types of interactions found in common smartphone authentication processes:

- **Single discrete input**, aka. “tap” – An individual tap that can be used for any button when applied on the touchscreen, to unlock the phone when applied on the unlock hard key or to use on the *fingerprint sensor*.
- **Multiple discrete inputs**, aka. “sequence” – A set of consecutive taps that can be used to input a *PIN* or *password* on the touchscreen.
- **Uni-directional continuous input**, aka. “swipe” – A continuous gesture that can be used as *swipe to unlock* on the touchscreen or as gesture input on the *fin-*

gerprint sensor (current sensors recognise four swipe directions)

- **Multi-directional continuous input**, aka. “pattern” – A continuous gesture that connects swipes in multiple directions without lifting the finger from the touchscreen, as it is known from common *pattern* authentication formats.
- **Implicit input** – This format includes all authentication methods that do not require the user to perform an action on the smartphone such as *FaceUnlock* or *voice recognition*.

3.2 Learning Task Interaction

We selected three different tasks for the interaction with the language learning content. All tasks are of simple design and focus on vocabulary recognition. They require only micro-interactions with the aim to be feasible for solving the task on the lockscreen and during the authentication process. We included the commonly used **Multiple-Choice** question format, in which the correct answer is presented accompanied by a number of incorrect answers (also called “lures”). The user has to pick the correct answer from the set. While this format can be used to present multiple correct answers, we decided to implement the simplest version of only one correct answer so only one click (single tap) will be needed to solve the task. With the simple interaction, this task is specifically feasible to be used for single touch authentication methods such as swipe-to-unlock or fingerprint authentication (using fingerprint sensor gestures, see details in subsection 3.1). The number of presented “lure” answers can further be adapted considering the limited space on the lockscreen but needs to be chosen carefully as too many answer options can increase the number of incorrect answers in later testing situations [6, 8] (i. e. “Negative Suggestion Effect” (NSE), for summary see [41]). For new learning content, Roediger and Marsh (2005) [39] recommend to use fewer alternatives in the beginning. Thus, we will present only one to two lure options as addition to the correct answer. Furthermore, at this prototypical stage, it is possible to present either the native word (L1) and ask the user for the second-language translation (L2) or present the L2 word and ask for the L1 translation.

As a second format, we implemented a “**Check Word**” task along the vocabulary test yes/no format, for example, applied in the Eurocentres Vocabulary Size Test [32, 33]. This task format aims to present a large number of words while requiring minimal time and effort by the user. In our examples, the L1 words are presented along the L2 transla-

Table 1: Overview of all prototypes mapped to the three learning tasks (columns) as well as the four interaction types (rows).

| | Check Word | Multiple-Choice | Sentence-Building |
|------------------------------|------------|-----------------|-------------------|
| Single Discrete | PT1 | PT5 | (not applicable) |
| Multi Discrete | PT2 | PT6 | PT11 |
| Uni-directional Continuous | PT3 | PT7, PT8 | (not applicable) |
| Multi-directional Continuous | PT4 | PT9 | PT10, PT12 |

tions while the user has to state if he/she knew the translation or not (single tap on a yes or no button). Similar to the the multiple-choice task, the yes/no answer format can be implemented also for single interaction authentications. To control for an overestimation of knowledge caused by users marking unknown words as known, the later application could include phonetically or orthographically similar pseudowords. In contrast to the yes/no test by Meara and Jones (1988) [33], this format (with included translation) tests not only the recognition of words but also checks if learners are aware of the word's meaning.

As a third task, we include a **Sentence-Building** exercise. Similar to the multiple-choice task design, the user is presented with a set of words from which they are asked to select those that make up a semantically and syntactically correct sentence. Due to the limited space of a lockscreen during authentication, we chose to present simple sentences with three words that need to be chosen out of a set of six words (three correct words, three lure answers). For example, a task could look as follows:

| | | |
|------|---------------|---------------|
| A | green | lemon. |
| An | yellow | tomato. |
| He | drinking | water. |
| They | drinks | wood. |

A sentence-building task can thus combine vocabulary recall with applying the correct grammatical construct. In the first example above, picking the correct article (a/an) and remembering the meaning of the words green/yellow and lemon/orange. In the second example, choosing the correct combination of pronoun (he/they) and verb form (drink/drinking) and remembering the meaning of the words water and wood.

4 Prototypes

For the generation of our prototypes we combined all authentication methods described in Section 3.1 with all

learning tasks presented in Section 3.2. This section gives an overview of the prototypes and presents an evaluation of users' experience in interacting with them.

4.1 Mapping of Learning Tasks and Authentication Methods

In Table 1 we show how our different prototypes address combinations of learning task and authentication method. All interaction concepts were implemented as high-fidelity interactive prototypes (PT1 to PT12) using the Justinmind¹ prototyping tool for mobile applications. For example, PT6 is depicted in Mockup 1 (see Figure 1a). The users enter their PIN and tap the screen once more to select one out of three translations for the L2 word presented below. Prototype PT3 implements Mockup 2 (see Figure 1b) by combining a *check word* task with a *uni-directional continuous* input, i. e. a swipe gesture starting at the on-screen fingerprint sensor to either of two buttons to indicate if the translation is known or unknown. This prototype is also applicable for other uni-directional continuous input types such as slide-to-unlock. Since sentence-building inherently requires connecting multiple words, it is not applicable for single input (discrete or continuous) authentication methods such as swipe-to-unlock. A sentence can be built with a multi-directional continuous input as visualised in Mockup 3 (PT12) see Figure 1c). Here, the user enters their predefined authentication pattern and, without lifting the finger, continues the gesture through connecting the words presented below to form a sentence.

4.2 Exploratory Evaluation of Prototypes

To investigate users' preferences for our different prototypes, we performed a user experience evaluation with ten in-depth semi-structured interviews. We showed the par-

¹ Justinmind: <https://www.justinmind.com/>, last accessed December 13th, 2021.



(a) Mockup 1
PIN with
Multiple-Choice

(b) Mockup 2
Fingerprint with
Check Word

(c) Mockup 3
Pattern with
Sentence-building

Figure 1: Three mockups² to visualise our concept for integrating learning into the authentication process. **Mockup 1** outlines a multiple choice task after PIN entry. **Mockup 2** Asks the user to indicate if the word is known (green button with arrow) or unknown (red button with cross). **Mockup 3** shows how sentence-building can be integrated into pattern authentication.

ticipants the interactive prototypes of the 12 concepts on a smartphone, asked them to interact with them, and collected their feedback. The focus of this user study was to gain insights into the usability and user experience of the concepts.

4.2.1 Procedure

At the beginning of each session, we welcomed the participants and informed them about the purpose of the study. After agreeing to the data protection policies along the GDPR regulations, participants signed a consent form. A short questionnaire was presented asking for demographic information such as age, gender, degree of education, current occupation, their experience with mobile learning, their current authentication technique of their smartphone, and its operating system.

Afterward, we introduced the participants to our prototypes and varied the order of presentation between the participants. The prototypes were shown on a Sony Xperia

Z smartphone with a 5-inch screen through the Justinmind app. People interacted with one prototype after another. We asked them to follow the think-aloud protocol, verbalizing their thoughts about the interface, interaction, and usability as well as problems of understanding. If the user was not able to perform the required interaction, we explained and/or demonstrated it and the user was asked to retry the interaction. We explicitly advised all participants to not input any authentication PIN or Pattern during the study that is similar to their own.

For each prototype, the interview was guided by the following user experience metrics: *Ease of use*, *Enjoyment*, *Interaction Speed*, and *Willingness to use* the prototype as a learning method in the future. Those facets were posed as yes/no questions but with the collections of further comments. The participants were furthermore encouraged to state any concerns regarding the prototype as well as potential improvements that come to their mind. Overall, the study took around 60 minutes depending on the extent of comments users had about the prototypes. We transcribed people's comments and applied a thematic analysis approach for comments and suggestions. This method is commonly used for identifying, analyzing, and reporting themes found within a qualitative set of data such as interview transcripts [5, 35].

² Mockups created using Figma (<https://www.figma.com/>, last accessed December 13th, 2021) and the Noun Project (<https://thenounproject.com/>, last accessed December 13th, 2021).

Table 2: Users' assessment of the prototypes along the dimensions ease of use, enjoyment, interaction speed, and their willingness to use the prototype in their daily lives. Questions were posed in a yes/no format and the answers in the table are visualised as “+” when 8 or more users agreed; similarly, a “–” indicates two or less users agreeing.

| Task Type | Action | Prototype | Easy to use | Enjoyable | Fast | Would use it |
|-----------------|--------|-----------|-------------|-----------|------|--------------|
| Check Word | SD | PT1 | + | + | + | + |
| | MD | PT2 | + | 0 | + | + |
| | UdC | PT3 | + | + | + | + |
| | MdC | PT4 | + | 0 | + | 0 |
| Multiple-Choice | SD | PT5 | + | 0 | 0 | + |
| | MD | PT6 | + | + | + | 0 |
| | UdC | PT7 | + | + | 0 | 0 |
| | MdC | PT8/9 | + | + | + | + |
| Sentence | MD | PT10 | 0 | 0 | 0 | 0 |
| | UdC | PT11 | 0 | 0 | – | – |
| | MdC | PT12 | + | 0 | – | 0 |

4.2.2 Participants

Ten participants took part in our study (2 female, 8 male) with a mean age of 24.2 ($SD = 1.55$). From those ten participants, nine were currently students and one stated to be a software engineer. In total, six participants reported to have obtained a Bachelor's degree, three a Master's degree, and one a PhD. Six participants owned an android phone, three used iOS, and one user used both operating systems due to owning two devices. The most commonly used method of authentication was the fingerprint, used by eight participants, whereof three stated having implemented a fallback authentication (one stated PIN, one password, and one Pin and FaceUnlock). One participant stated to use a pattern to authenticate and the last one used FaceUnlock. Furthermore, four participants stated to currently use their mobile device for language learning.

4.2.3 Results

We transcribed the responses of the participants and report on them here, clustered by the three task types *Check Word*, *Multiple-Choice*, and *Sentence-Building*. The answers of two participants for one prototype each (P6's PT19 and P8's PT10) had to be excluded due to technical problems with the audio recording device. We will report the summative evaluation of the users' answers to the four guiding questions on ease of use, interaction speed, enjoyment, and willingness to use, and report on the open comments in the style of a thematic analysis [5, 35].

4.2.3.1 Check Word

The majority of the participants rated all Check Word prototypes as easy to use in their interaction. For PT2 (multi-

discrete) and PT4 (multi-directional continuous), one participants each stated that the mechanism lacked intuitiveness. PT1, PT2, and PT3 were unanimously rated as fast, while P7 was the only participant rating PT4 as not fast. In particular, PT4 was not perceived very positively by the participants. Five out of ten indicated that they would not be open to using this prototype in their daily lives, arguing it is “no fun” (P3, repeated by P5). Others would not use it as they stated to never use Swipe-to-unlock in general due to low security (P7, P8). In general, participants doubted the efficiency of the check word task and expected a more engaging learning task.

4.2.3.2 Multiple Choice

All participants considered PT5, the simple tap on one out of two answer solutions, very easy to use and very fast. P1 and P9 considered it not very enjoyable, while three participants stated to not be willing to use it due to a lack of security. Compared to the single tap needed for PT5, participants perceived PT6 with multiple discrete taps needed as not fast (P3, P7, P8, P9). P8 attributes the lack of speed to the PIN authentication method in general. Similarly to PT5, three participants would not use it due to the authentication method and the “cluttered screen” (P9). Out of the four multiple-choice prototypes, PT7/8 was best received by our participants. Nine out of ten mentioned they would use it and all considered it fast and enjoyable. Two participants considered it not intuitive enough and expected a tap instead of a swipe gesture. Lastly, for PT9, the multi-directional continuous input, the majority of participants considered it easy to use (9/10) and would use it in their daily lives (8/10). However, P8 noted that it felt slow because one first has to think about the pattern input and then, while not releasing the finger, decide which answer

to pick. Similar to the multi-directional continuous input for the other tasks, the screen was perceived as cluttered. As suggestion for improvement participants stated to reduce the number of answer options from three to two.

4.2.3.3 Sentence-Building

Overall, the sentence-building tasks were perceived as least intuitive compared to the check word and multiple-choice tasks. PT11 with its multiple discrete inputs required was considered time consuming by eight out of ten participants and also as not enjoyable (4/10). P4 highlighted that due to the many actions one has to perform, *“it’s not fun at all, so many steps just to unlock make this task feel like a chore”*. While the continuous multi-directional input (PT12) was rated to be faster and more enjoyable, the majority of the participants still indicated that they are not willing to use it. In particular, P7 mentioned that a complicated task like this does not belong on the lockscreen and that the multiple-choice or check word tasks were a better fit.

4.3 Limitations

During our study, participants expressed difficulties in rating the interaction with the embedded authentication learning task without being biased by the authentication method used. In general, several participants considered swipe-to-unlock as insufficiently secure and would not use this mechanism. We stressed that this technique could be integrated into fingerprint authentication and should be rated independently from users’ opinion on the authentication itself. However, we cannot be certain that our participants were able to avoid such bias. We further noticed that the alignment of the learning tasks below the authentication interface was not optimal. In particular, for continuous input gestures such as in PT9 and PT12, users had to interrupt the continuous gesture to have a look at the options again. The visibility of the options is ultimately defined by the end-point of the pattern input – i. e. when the pattern ends in the lower left corner and the user is right-handed, the hand will almost completely cover the answer options.

4.4 Summary

We have presented twelve prototypes for combining three learning tasks (check word, multiple-choice, and sentence-building) with four common forms of smartphone authentication mechanisms (represented through their required interaction: single discrete input, multi-

ple discrete, single continuous, multi-directional continuous). A first usability evaluation with ten participants showed that users’ preferences vary greatly. Most prototypes were reported to be easy to use and fast. While some participants preferred discrete taps, others especially enjoyed the single and continuous swipe input schemes.

The main take-away from this evaluation is that users desire a quick and easy learning interaction. While they value the idea of combining learning with their authentication, many participants emphasised that the learning should not impede or obstruct the unlock process. In particular, the sentence-building tasks were considered fairly slow and not a good fit for the presentation on the lockscreen. We thus excluded this type of task for our further evaluation. However, participants doubted the efficiency of the check-word tasks. Thus, we will further explore the multiple-choice task as the most promising concept.

Lastly, we noticed that the speed perception varied greatly among participants. While some found the interaction slower than necessary for an authentication method, others perceived the learning task itself as rather quick. Along the glass half-full vs. glass half-empty situation, we argue that the perception strongly depends on the users’ perspective on the situation. As the proposed methods combine authentication and learning, it is clear that the interaction will take longer than authentication alone. However, when comparing the combined interaction with the action of having to open a stand-alone learning app to learn a vocabulary, it is undeniable that the newly proposed prototypes enable a faster interaction. Thus, our following evaluations include a standalone app as control condition for participants to compare the interaction fairly.

5 Implementation

For our main investigation, we implemented three concepts with varying levels of integration following existing research ideas presented in Section 2.1, as Android applications (for Android 6.0 and higher). The *UnlockApp* (see Figure 2a) presents vocabulary immediately following the unlock action. A word is presented in a foreign language (L2) and the user is asked to choose between two presented translations in their mother tongue (L1) by pressing a button on the screen. We chose to implement a multiple-choice approach with two answer options as it enables quick and simple interaction with the content.

We further implemented a push notification learning app, which presents vocabulary in a pervasively displayed

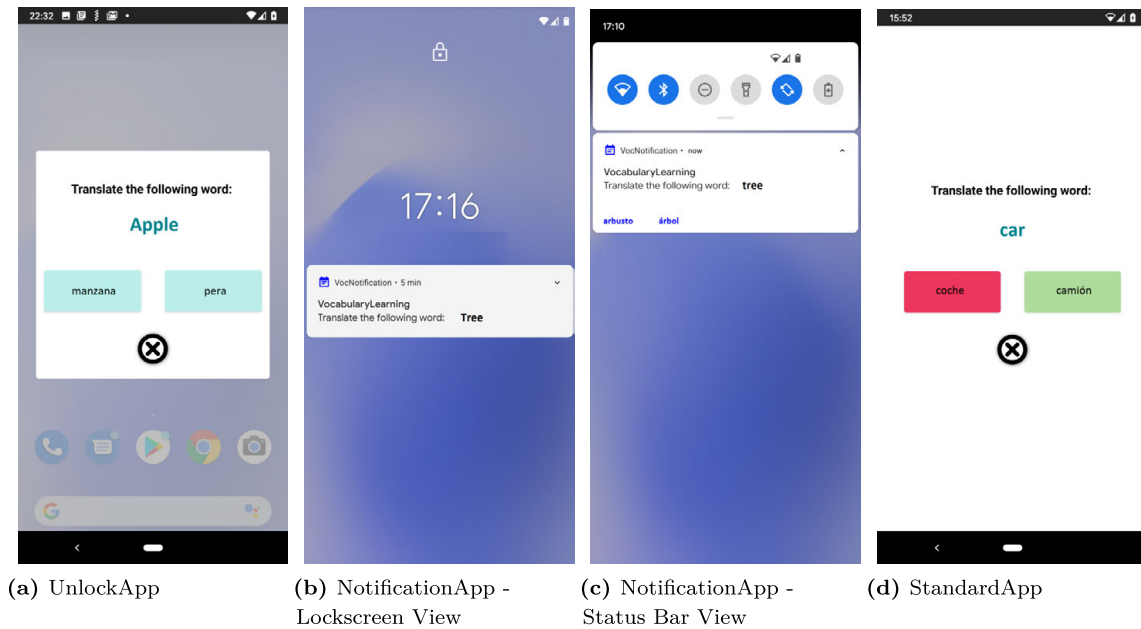


Figure 2: The *UnlockApp* (a) displays a vocabulary task immediately after the user performs the authentication. The *NotificationApp* (b+c) presents the learning task in a continuously displayed push notification. This notification is visible on the lockscreen (b) and in the (pulled down) status bar (c). Via notification action buttons (also displayed on the lockscreen if interactivity is enabled), the user can select one out of two translations for the respective word. Figure (d) shows the *StandardApp*, in the state of showing corrective feedback.

notification in the status bar (named *NotificationApp*, see Figure 2c, along the concept proposed by Dinger et al. (2017) [15]) and on the lockscreen (see Figure 2b).

Finally, we implemented a baseline vocabulary app, similar to the interactions with established apps, which the user has to actively start and quit (named *StandardApp*, see Figure 2d).

Beyond the different presentations, a key conceptual difference is that the *UnlockApp* nudges the user to interact with the content, while the *NotificationApp* and *StandardApp* require users to actively initiate the learning themselves. However, in contrast to the *StandardApp*, the *NotificationApp*'s constant visibility acts as a continuous reminder for users to engage with the learning content.

5.1 Learning Content

All three apps include vocabulary allowing for learning Spanish, French, and Swedish with translations into the native language of our University's country. The word list comprised of 446 common nouns (originally from the British National Corpus,³ last accessed December 13th, 2021) and have been translated to the other languages using Google Translate while removing inaccurate transla-

tions and highly ambiguous words, similar to the procedure of Cai et al. (2014) [10] and Dingler et al. (2017) [15].

In the background, all three applications are connected to a Firebase database, which stores both the learning content and interaction logging data. In the database, each user is assigned a unique id to ensure anonymity. All three applications are described in more detail in the following.

5.2 UnlockApp

To integrate the learning task into the authentication action, we display the vocabulary translation interface immediately when the user unlocks the screen (see Figure 2a). As the current Android versions do not allow a direct manipulation of the lockscreen, we chose to implement a context-registered broadcast receiver. This receiver listens for the unlock event (`ACTION_USER_PRESENT` indicates that the user is present after the device is woken up⁴) while running continuously in the background and places the app in the foreground when an unlock occurs. Since it is possible that applications are shut down for battery optimisation purposes, we further included a fore-

³ <http://www.natcorp.ox.ac.uk>

⁴ Android User Intent: <https://developer.android.com/reference/android/content/Intent>, last accessed December 13th, 2021.

ground service that allows for automatic restart of the application.

When an unlock event occurs, the *UnlockApp* asks the user to translate one word from the native language to one of the three languages of choice – Spanish, French, or Swedish. To keep the learning task as short as possible, the interface only presents the translation task, two potential translation options in the form of buttons. Further, we include a “skip” option to keep learners from guessing the answer, as guessing increases the risk of experiencing the negative testing effect [31] (i. e. remembering the false answer one gave to a multiple-choice question rather than the correct solution).

After the user selects one of the two translation options, the *UnlockApp* displays corrective feedback by highlighting the answer option in either red or green. Providing feedback in multiple-choice learning tasks is essential and counteracts the negative testing effect [9]. After a short delay for users to perceive the feedback, the app moves into the background.

We chose a translucent background to not distract the user too much from the primary task they were about to accomplish when unlocking the phone. Prior research by Hodgetts and Jones (2006) [22] has shown that when the primary task was still visible during an interruption, participants are able to resume the task faster. To ensure consistency, we used the Android material design and color palettes.

The *UnlockApp* further allows to use a fingerprint sensor gesture (swiping left or right on the fingerprint sensor) as input to answer the tasks. We utilise the Fingerprint Gesture Controller accessibility service.⁵ This controller registers gestures performed on the fingerprint sensor, independent from sensor location (depending on the model, the sensor can be located on the back of the phone, on the side, or on the front below or on the screen). While prior work has already shown that fingerprint sensor gestures can be a feasible tool for quick input such as unlock journaling [18], this feature is not yet supported by many Android devices. Thus, we will not evaluate the usage of it in this study.

5.3 NotificationApp

In this app, learning content is presented in a continuously displayed push notification, both in the (pulled down) sta-

tus bar of the smartphone (see Figure 2c) but also on the lockscreen (see Figure 2b). This enables the user to interact with the content at any point in time and answer as many tasks as they like to, the continuous notification working as an active and constant reminder. The notification displays the task and word to be translated as well as two potential translation options via action buttons.⁶ Those buttons are frequently applied by other applications to allow immediate interaction with the app sending the notification. For example, Gmail⁷ employs action buttons to enable users to delete or reply to emails directly from the notification. We implemented a broadcast receiver to listen for the input in the notification and check the correctness. In the same style as the *UnlockApp*, green and red coloring of the two potential translations indicates corrective feedback. Afterward, the notification is updated and a new word is shown. Likewise, when the user dismisses a notification or restarts the app, it presents a new word. The notification itself cannot be removed by the user but remains constantly visible in the pulled down status bar.

5.4 StandardApp

In contrast to the *UnlockApp* and *NotificationApp*, the *StandardApp* does not show any continuous reminder of the learning task. The users have to actively open the application and engage with the learning content. Once the users open the app, they can complete as many learning tasks (i. e. request as many translation tasks) as they like. The appearance of the application’s interface is consistent with the interface of the *UnlockApp*: The app shows the native word, the two translation options in buttons, a dismiss button, and corrective feedback for the task (see Figure 2d). In contrast to the *UnlockApp*, the background is fully opaque.

6 Evaluation

Our evaluation follows a within-subject design, where all participants interact with all three different applications to allow them to draw comparisons. Each app is used over the course of one week (seven days) before switching to

⁵ FingerprintGestureController <https://developer.android.com/reference/android/accessibilityservice/FingerprintGestureController>, last accessed December 13th, 2021.

⁶ Android Notification Action Buttons: <https://developer.android.com/training/notify-user/build-notification#Actions>, last accessed December 13th, 2021.

⁷ Gmail: <https://www.google.com/gmail/about/>, last accessed December 13th, 2021.

the next app. The vocabulary progress is preserved so that people can continue in the next app where they left off in the previous one. The order of the apps is counterbalanced across participants to avoid sequence effects.

In our subsequent analysis, we explore differences in users' interaction with the three apps (independent variable), in particular frequency and duration of interactions, as well as perceived usability and experience (dependent variables). Based on related work, we particularly evaluate differences among the three apps with regard to how much they expose users to the learning content, stating the following main hypothesis:

H_1 Embedding vocabulary tasks into everyday smartphone interactions as in the *UnlockApp* and *NotificationApp* leads to a higher number of learning tasks solved by users over the course of the study compared to the *StandardApp*.

Furthermore, we gather individual feedback on participants' preferences and experiences when learning with the three applications.

6.1 Procedure

We provided participants with a detailed installation guide enabling people to download the three applications from the Google Play Store (setup as a non-public test version), after they gave informed consent. The guide also encouraged them to restart the (current) app in case they restart their phones or the app is closed for any other reason and does not restart automatically.

At the beginning of the study, people filled in a questionnaire on demographics and current authentication method, smartphone usage habits, language proficiency, and motivation to learn a new language. In this process, people selected a language to learn with the apps (Spanish, French, or Swedish). We informed participants that the vocabulary taught in the apps is on a beginner's level and thus recommended to choose a language they are not very proficient in.

During the study, we logged data on the users' interaction with the apps, including the number of solved vocabulary tasks, their correctness, response times and task dismissals (in the *UnlockApp*).

6.2 Sample

We recruited 30 participants via University mailing lists, social media channels, and word-of-mouth, nineteen

identifying as female and eleven as male. Their average age was 29.8 years ($SD = 15.55$, range 19–78 years) and the majority (18) reported to be studying or working full time in jobs such as Engineer, Physicist, Accountant, Doctor, or Dentist, while two were pensioners. Fifteen stated having a high school degree and twelve a university degree (including bachelor, master, or phd). One third (19 people) reported using fingerprint authentication on their smartphones, and seven reported using a PIN or password, two patterns, and two face recognition. All participants stated to use their device at least multiple times a day (15) if not multiple times per hour (15).

At the time of the study, ten participants confirmed learning or actively improving on a language (four Spanish, three French, two English, and one Japanese). They further stated to be proficient in at least one and up to five foreign languages ($Md = 3$), and reported to have great interest in learning a new language ($M = 6.07$, $SD = 0.85$, Likert-scale from 1 = "I fully disagree" to 7 = "I fully agree"). As languages to learn in this study, thirteen people chose Spanish, twelve French, and five Swedish. For their successful participation all people received a 25 Euro Voucher for an online store, or an equivalent amount of study credit points.

6.3 Results

We report on the data of 30 people. However, one person did not complete any of the final questionnaires, and two additional people did not submit the questionnaire for the *StandardApp*. Therefore, we report the interaction data of 30 and the questionnaire results of 29 people for the *UnlockApp* and *NotificationApp*, 27 for the *StandardApp*, respectively. For the statistical analysis of the questionnaires, we performed pair-wise exclusion for incomplete data sets.

We used R [38] for significance testing; concretely, (generalised) linear mixed-effects models (LMMs, packages *lme4* [3] and *lmerTest* [27]). The LMMs accounted for individual differences (*participants*) and *app order* via random intercepts. Note that app order was counterbalanced yet we still included it here following best practices. As fixed effects, we included *app*, plus the *number of days* since the start of the study.

6.3.1 Overall Usage

In total, we recorded 7715 answered vocabulary tasks over all participants across the three weeks of our study.

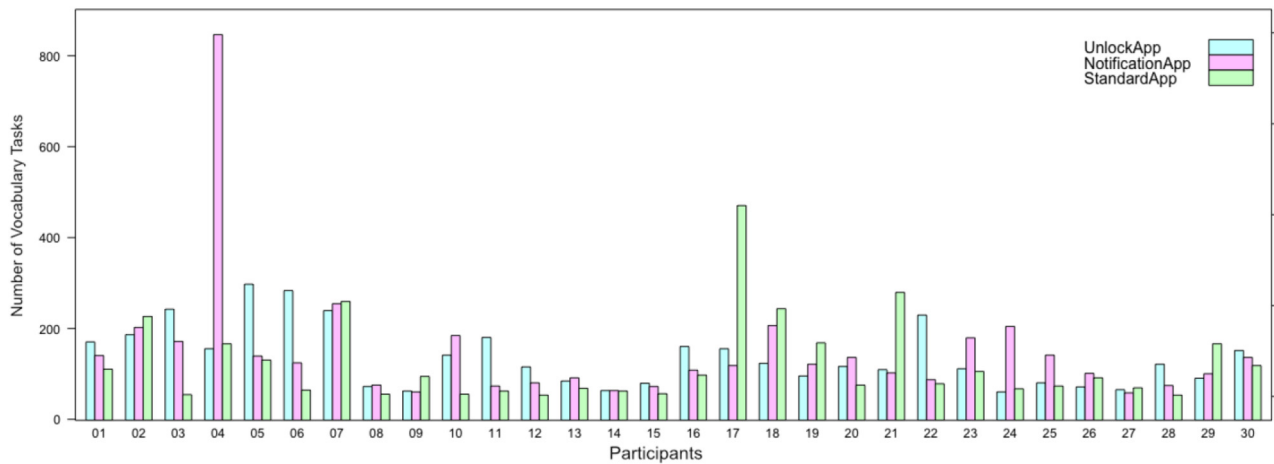


Figure 3: Overview of participants' usage of the three apps during the study: (per participant from left to right) the overall number of vocabulary tasks solved in the *UnlockApp*, *NotificationApp*, and *StandardApp*.

The most tasks were answered using the *NotificationApp* (2945), followed by the *UnlockApp* (2604) and the *StandardApp* (2166). On average per day, people used the apps to solve 10-13 vocabulary tasks. Here, the *StandardApp* showed the lowest mean usage ($Md = 2$, $M = 10.28$, $SD = 20.05$), while the *UnlockApp* ($Md = 8$, $M = 12.28$, $SD = 14.73$) and *NotificationApp* ($Md = 5.5$, $M = 13.7$, $SD = 28.43$) were used more frequently. While the *NotificationApp* shows the highest overall usage, the Median is lower than the Median of the *UnlockApp*. This is due to one exceptional case in the *NotificationApp* usage (see P4 in Figure 3): P4 reported having used the *NotificationApp* extensively out of enjoyment, with over 800 answered vocabulary tasks (increasing the overall usage count but with little effect on the Median). Since this was intended use, we do not remove this as an outlier from our analysis.

For significance testing, we fitted a generalised LMM (Poisson family) on the answer count data (i. e. number of answered vocabulary questions). The model had *app* as a significant positive predictor (*UnlockApp*: $\beta = .43$, $SE = .06$, $CI_{95\%} = [.31, .55]$, $p < .0001$; *NotificationApp*: $\beta = .36$, $SE = .06$, $CI_{95\%} = [.25, .48]$, $p < .0001$): Therefore, compared to the *StandardApp*, using the *UnlockApp* was estimated by the model to result in $\exp(\beta) = 1.54$, that is, 54 % more answered vocabulary questions. Similarly, using the *NotificationApp* was estimated to result in 43 % more. Moreover, the model had *day* (since start of the study) as a significant negative predictor ($\beta = -.15$, $SE = .01$, $CI_{95\%} = [-.17, -.12]$, $p < .0001$): The number of answered questions was estimated by the model to decline over the course of the study (estimated as $\exp(\beta) = 0.86$ i. e. -14 % per day). The interaction of *day* and *app* was significant for *Unlock-*

App ($\beta = -.07$, $SE = .02$, $CI_{95\%} = [-.10, -.04]$, $p < .0001$), but not for *NotificationApp* ($p = .176$).

We further plotted people's interactions with the three applications over the course of the whole day and the week of use: Figure 4 visualises the number of tasks solved by the users, indicating that they interacted with each app most frequently on the first day of use. It has to be noted here that participants started the study on different days of the week (Monday: 5, Tuesday: 2, Wednesday: 2, Thursday: 5, Friday: 5, Saturday: 6, Sunday: 5) and the starting weekday then stayed the same for each person across app conditions. Moreover, the plots reveal that interactions with the *UnlockApp* are scattered more across the time of the day compared to the interactions with the *StandardApp* and *NotificationApp*. Furthermore, the plot matches the results of the LMM that the number of answered questions declined from day one to day seven.

6.3.2 Learning Task Correctness

In general, the users answered more than 90 % of the learning tasks correctly over all three applications (see Table 3). The correctness rates for the *StandardApp* and *NotificationApp* are slightly higher than the rates for the *UnlockApp*. A Friedman test revealed no significant difference ($p > .05$) for the task correctness of the answers with regard to the three applications.

For the *UnlockApp*, we can further measure the time between when the learning task is presented and the moment the answer is recorded. On average, it took participants 3.4 seconds to answer a task using the *UnlockApp* ($SD = 1.02$). The time to complete the task for incorrect an-



Figure 4: The relative distribution of vocabulary tasks solved for the *UnlockApp* (left), *StandardApp* (middle), and *NotificationApp* (right) in percent ($0.06 = 6\%$). The amount of vocabulary tasks is normalised for each user with respect to their overall number of solved tasks and visualised according to the seven usage days (x-axis) and the hours of a day (midnight to midnight, y-axis).

Table 3: Overview of people's use of the three apps in terms of the number of answered vocabulary questions.

| App Type | # Answers Total | # Correct | # Incorrect | % Correct |
|-----------------|-----------------|-----------|-------------|-----------|
| StandardApp | 2166 | 2026 | 140 | 93.54 % |
| NotificationApp | 2945 | 2762 | 183 | 93.79 % |
| UnlockApp | 2604 | 2380 | 224 | 91.40 % |

swers is higher ($M = 3.98$, $SD = 2.71$) compared to correctly answered tasks ($M = 3.38$, $SD = 1.02$).

The *UnlockApp* enabled users to dismiss a learning task in case they did not want to answer it. During the seven study days, users chose to skip between zero and 80 tasks, with an average of 9.6 skips ($SD = 14.48$, $MD = 4$).

6.3.3 Favorites, Ratings & Subjective Results

We asked participants directly which of our three applications they would prefer for everyday use. As their most favorite application, 18 out of 30 people named the *NotificationApp*, followed by seven mentions of the *UnlockApp*. Vice versa, 15 participants stated the *UnlockApp* as their least favorite of the three applications, followed by 13 naming the *StandardApp*.

Moreover, we analysed participants' subjective impressions of the three applications (expressed in the final questionnaire as 7-point Likert scale items from 1 = "I totally disagree" to 7 = "I totally agree"). For each item we performed a Friedman test with post-hoc Wilcoxon pairwise-comparisons. There was a significant difference

in how much participants liked the applications (a: $\chi^2(2) = 6.812$, $p < .05$, see Figure 5a). Post-hoc analysis with Wilcoxon signed-rank tests showed a significantly higher score for the *NotificationApp* compared to both the *StandardApp* ($Z = -2.159$, $p < .05$) and the *UnlockApp* ($Z = -3.100$, $p < .01$).

We further asked participants (b) if they felt they used the app only when they had time to learn and (c) if they used the app even when they had no time to learn. Both items revealed significant differences in participants' subjective perception of their usage among all three applications (b: $\chi^2(2) = 16.568$, $p < .001$, see Figure 5b; c: $\chi^2(2) = 15.571$, $p < .001$, see Figure 5c). For item (b), participants report to be more likely to use the *StandardApp* when they had time to learn compared to the *NotificationApp* ($Z = -2.060$, $p < .05$) and the *UnlockApp* ($Z = -3.960$, $p < .001$), and more likely to use the *NotificationApp* when compared to the *UnlockApp* ($Z = -2.737$, $p < .01$). For item (c), participants report to be more likely to use the *UnlockApp* even when they had no time to learn compared to the *NotificationApp* ($Z = -2.786$, $p < .01$) and the *StandardApp* ($Z = -3.945$, $p < .001$), and more likely to use the *NotificationApp* when compared to the *StandardApp* ($Z = -3.945$, $p < .001$).

We found no significant differences in participants' ratings of the design of the three applications and their intuitiveness of use. Moreover, the Friedman test revealed no significant difference in terms of participants' willingness to continue using the applications in their everyday life. The descriptives, however, show a slight preference in favor of the *NotificationApp* ($M = 4.37$, $SD = 1.85$) compared to the *UnlockApp* ($M = 3.70$, $SD = 2.05$) and the *StandardApp* ($M = 3.30$, $SD = 1.73$), see Figure 5d.

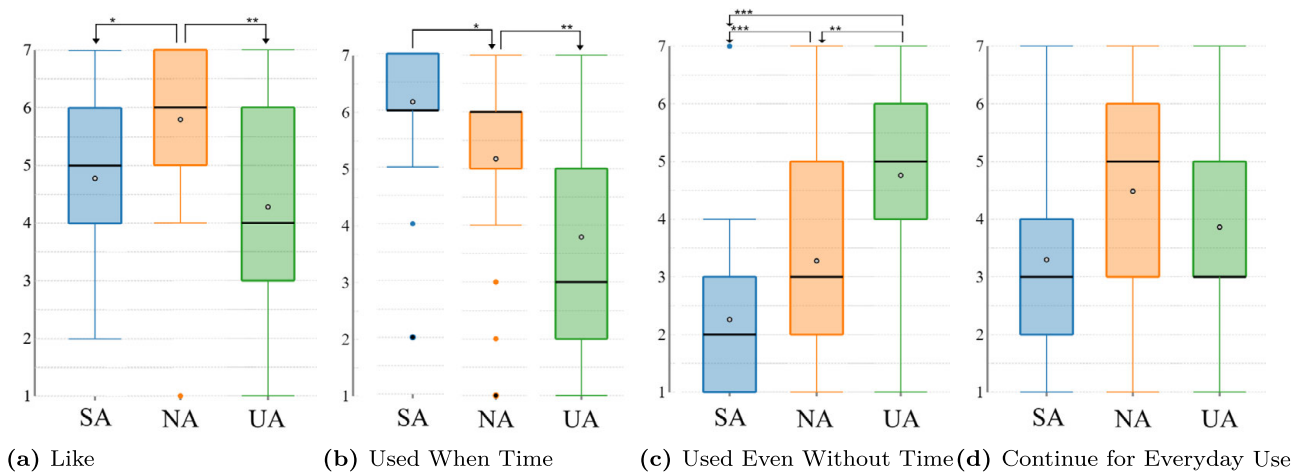


Figure 5: Participants' ratings of the *StandardApp* (SA), *NotificationApp* (NA), and *UnlockApp* (UA) for four questionnaire items (from 1 = "I totally disagree" to 7 = "I totally agree"): (a) I liked the [SA/NA/UA]; (b) I only answered tasks with the [SA/NA/UA] when I had the time to learn; (c) I answered tasks with the [SA/NA/UA] even when I did not have time to learn; (d) I would continue using the [SA/NA/UA] frequently in my everyday life for learning a language. Asterisks indicate all statistically significant differences (* < .05 | ** < .01 | *** < .001).

6.3.4 Open Comments and Suggestions for Improvement

Regarding the concept of the *NotificationApp*, people reported positive as well as negative impressions: One participant considered the continuous presentation not optimal and rather wants to take the time to actively engage with language learning (P15). P7 perceived the display of the notification on the lockscreen as distracting. Two other participants suggested increasing the delay between the presentation of two vocabulary tasks so that the app is not running all the time but just presents a new word from time to time (P21, P27). On the other hand, some participants positively emphasised the unobtrusiveness of the notification app, saying it is "[...] *not distracting, it provides the opportunity to quickly solve a couple of learning tasks*" (P11) and that it is "*very suitable for daily use*" (P25). To further improve the application, participants wished for the inclusion of grammar knowledge (P2, P5, P14), free text entry of words to foster recall rather than recognition of words (P2), audio output to help with the pronunciation (P5, P8, P14), and gamification features (P6).

Similar suggestions were stated for the *UnlockApp*: Participants wished for free text entry opportunities (P22), pronunciation support (P9, P26), and grammar knowledge (P3, P5, P7, P19). Since vocabulary in itself is not sufficient to learn a new language, P6 considers the app a nice addition for people currently attending language learning courses and P19 recommends using it to freshen up a language. Regarding the overall concept of combining the learning task with authentication, many participants stated positive impressions: While P9 highlighted the sim-

ple design, P19 emphasises that "*it's good that you always have to solve at least one task and therefore learn continuously*". Further, P6 stated to like the app and in particular the idea, similar to P8. However, P8 adds that "[...] *it bothers me that I have to answer the task or dismiss the app when I just quickly want to do something on my phone*". P20 shares the experience and describes that when they just quickly want to use the phone, they are "*not focused enough to answer the questions conscientiously*". To address this issue, other participants suggested including more personalisation options. For example, P7 wishes for a feature to adjust the number of vocabulary tasks presented at each authentication (which was fixed to one in the study), describing that if a user currently has more time to learn, then they could increase the number of words presented per unlock event. Further, P15 proposes to define time intervals during the day in which the app is active (P15).

7 Discussion

7.1 Limitations

Our user study was conducted during the COVID-19 pandemic, probably leading to anomalies in users' daily routines, mobility, and smartphone usage due to lockdown and work-from-home phases. All these factors potentially influence the use of mobile language learning applications; thus, we can not ensure the generalisability of our

results. However, we included a control condition via the *StandardApp*.

The focus of our evaluation lies on the users' experiences and users' interaction with the three apps of varying integration into smartphone usage. Although our applications do not constitute an exhaustive representation of all MLL applications, we chose these three as a concise comparison of how different levels of embedding could be implemented. Further, with our focus on user's interaction we can not report on actual vocabulary retention. However, as retention of vocabulary is increased by frequent interaction with the content [12, 14], we expect our applications to positively impact people's vocabulary recall and recognition. Nonetheless, actual vocabulary retention, particularly concerning the consolidation properties of the apps over time, needs to be further evaluated in future work.

By providing people the three applications to use for each seven days during this user study, we gained interesting insights into users' learning behavior. We decided in favor of a shorter within-subject study over a longer between-subject study to enable users to draw comparisons among their interaction with the three apps and to make the best use of the limited sample size. However, we are aware that our study only presents a narrow view onto users' actual behavior. As the users show a great diversity in individual preferences, only a long-term evaluation with a larger sample will be able to show which behaviours will prevail in users' daily lives. Our study data shows a more extensive usage of each application on the first day, declining over the seven days of usage. This pattern could indicate a form of curiosity or novelty effect. Educational technology research has shown that customisation with new learning technology can negatively impact learners' preferences regarding technology-based learning [25] and reduce users' motivation [24]. In contrast to the *StandardApp*, the *UnlockApp*'s concept partly counteracts this decline by continuously engaging the learner in solving vocabulary tasks with each authentication event. However, a future long-term evaluation in particular of the new *UnlockApp* concept is required to reveal the strength of the novelty effect in everyday usage.

7.2 UnlockApp and NotificationApp Increase Vocabulary Exposure

Our analysis confirms our initial hypothesis, showing that the *NotificationApp* and *UnlockApp* result in a higher number of solved vocabulary tasks per participant. Thus, we

conclude that the learning content exposure was higher for these two apps when compared to the baseline, the *StandardApp*. Furthermore, the presentation of the vocabulary tasks after the authentication leads to a more spread out exposure to the tasks across the day, confirming Dingler and colleagues' [15] conclusion that proactively triggering vocabulary learning sessions can effectively space out learning. Further, the continuous presentation of content through the *NotificationApp* allows for short but also extensive learning sessions (cf. the outlier P4). The similar distribution of correctly and incorrectly answered tasks across the three applications gives no indication that users might have been less focused when learning with the *UnlockApp* or *NotificationApp*. Additionally, we observed that users occasionally skip tasks with the *UnlockApp*. This suggests that even though the skipping requires the same amount of effort as selecting an answer (one button press), the users can judge if they have the mental capacity and/or time to engage in the learning or not.

7.3 Potential for Adaptation & Personalisation

The subjective feedback revealed individual differences regarding people's attitudes toward the three applications: While many state to like the concepts of the *NotificationApp*, opinions on the *UnlockApp* are mixed. Specifically, some people felt distracted by the *UnlockApp*'s vocabulary presentation when they unlocked their phone with a specific task in mind. People do not oppose the concept in general but express their need for further personalisation. Suggested adaptation features include the definition of learning time frames for the *UnlockApp* or adjusting the number of tasks presented with each authentication event. Moreover, we see potential for automated mechanisms that learn from people's interactions with the *UnlockApp* (in particular dismissals) and adjust the presentation accordingly. I. e. the app should stop promoting users with learning tasks after unlock events at times during the day when they are frequently dismissed.

7.4 Extending the UnlockApp Concept to Different Authentication Methods

Based on the results here, we deem it useful to discuss possible extensions of the *UnlockApp* concept to further authentication methods: Concretely, the majority of participants in our study used fingerprint authentication to

unlock (yet with phones that do not support fingerprint gestures). Our app and concept already support fingerprint gestures to respond to vocabulary tasks (see implementation section), which can help to embed the learning task even more implicitly into the authentication (i. e. no switch from fingerprint to touch required). Prior work has found that using fingerprint sensor gestures for journaling is perceived faster and less intrusive compared to notifications [18]. Both touch input and fingerprint gestures might also be combined with unlocking methods that do not require further input themselves, such as face unlock, or when Android's smart screen lock feature is enabled (i. e. no unlock at home). Beyond this, we already gathered experiences with knowledge-based authentication methods: One third of our sample stated to use PIN, password, or pattern authentication. In these cases, the *UnlockApp* also presents the task immediately after the unlock event.

7.5 Extending the Learning Features of Embedded Learning Applications

Our application focused on the presentation of vocabulary translations to reduce the complexity of the content. With this, we aim to simultaneously increase the control of potential effects among the three apps. The participants of our study strongly emphasised the demand for additional features such as grammar knowledge or pronunciation exercises in open comments, and compared our application to applications available on the market. To accommodate for people's need for more complex learning tasks, we see three potential solutions: (1) By combining the embedding concepts with a fully-featured learning app, we could address users' preferences for long learning streaks when they have time to spare (i. e. with the *StandardApp*), during which the app could also teach grammar or pronunciation knowledge. The differentiation between two complexity levels could also be realized by (2) differentiating between short and long learning sessions as suggested by Cai and colleagues [10] (e. g. by exploiting different types of waiting situations). Additionally, the *UnlockApp* could refresh users' vocabulary knowledge by continuously presenting translation tasks after each authentication to engage users with a language on a daily basis. Option (3) is to investigate extensions of learning tasks that are short enough to be embedded into the *UnlockApp* or *NotificationApp*. These tasks need to be solvable with simple interactions and with low effort and time. Possible examples include fill-in-the-blank tasks offering a word in two tenses or with two suffix options.

8 Conclusion & Future Work

We evaluated twelve concepts for embedding learning tasks into the authentication process, as it is a highly frequent yet otherwise useless daily action. While participants appreciated the idea of being “nudged” to learn more frequently through the embedded authentication learning approach, they emphasized the need for a quick and simple interaction. If the task is too complex or slow, they could not imagine using it over a longer period. Based on these results, we implemented the *UnlockApp*, which presents a multiple-choice vocabulary task after each unlock event, and performed a follow-up evaluation.

We compared our *UnlockApp* concept to a second application presenting learning tasks in a continuously shown notification (*NotificationApp*) and a standalone learning app. When similar ideas were proposed in the literature, these individual concepts had been evaluated independently of each other. Thus, in this paper, our main contribution to the literature is their direct empirical comparison in a single study.

In this light, our key result and conclusion is that embedding vocabulary learning tasks into everyday smartphone interactions – concretely: the status-bar notifications or the unlock procedure – significantly increases exposure over a dedicated self-initiated learning app.

To critically reflect on this finding, we might ask if this embedding and increased exposure is “worth it” for users, considering potential downsides such as distraction or longer unlock interactions. Based on our results and user feedback here, we conclude that the overall answer is yes, yet our insights point towards individual differences: Both embedding approaches (notification, unlock) were favoured by a considerable proportion of people, who provided various concrete reasons for their choices.

In the broader context of the literature, the results from our direct comparison here thus support such embedding and micro-learning concepts, and therefore strengthen their motivation and this line of research overall. Practically speaking, our findings motivate that language learning apps should consider offering these embedding concepts to users, for example, as complementary widgets to their main app content.

Looking ahead, our findings encourage further investigations in this direction: For example, future work could explore the design space of unlock integration in detail (e. g. supporting further learning tasks and unlock mechanisms), also to improve the individually perceived trade-offs of learning and disturbance (e. g. via personalised and context-aware adaptations). Also, learning tasks could be connected to other smartphone actions, such as giving ad-

ditional purpose to less productive actions defined by the user (e. g. starting a social media app).

References

- [1] David P Ausubel and Mohamed Youssef. The effect of spaced repetition on meaningful retention. *The Journal of General Psychology*, 73(1):147–150, 1965.
- [2] Matthias Baldauf, Mohamed Khamis, Sebastian Steiner, and Sarah-Kristin Thiel. Investigating the user experience of smartphone authentication schemes-the role of the mobile context. In *Proceedings of Hawaii International Conference on System Sciences (HICSS 2019)*, 03 2019.
- [3] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.
- [4] Rasekhar Bhagavatula, Blase Ur, Kevin Iacovino, Su Mon Kywe, Lorrie Faith Cranor, and Marios Savvides. Biometric authentication on iphone and android: Usability, perceptions, and influences on adoption. In *USEC'15: Workshop on Usable Security*, 8 February 2015, San Diego, CA: Proceedings, pages 1–10, 2015.
- [5] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.
- [6] Alan S Brown, Hildur EH Schilling, and Melinda L Hockensmith. The negative suggestion effect: Pondering incorrect alternatives may be hazardous to your knowledge. *Journal of Educational Psychology*, 91(4):756, 1999.
- [7] Peter A Bruck, Luvai Motiwalla, and Florian Foerster. Mobile learning with micro-content: A framework and evaluation. In *Bled eConference*, page 2, Atlanta, GA, USA, 2012. Association for Information Systems (AIS).
- [8] Andrew C Butler, Elizabeth J Marsh, Michael K Goode, and Henry L Roediger III. When additional multiple-choice lures aid versus hinder later memory. *Applied Cognitive Psychology*, 20(7):941–956, 2006.
- [9] Andrew C Butler and Henry L Roediger. Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & cognition*, 36(3):604–616, 2008.
- [10] Carrie J. Cai, Philip J. Guo, James Glass, and Robert C. Miller. Wait-learning: Leveraging conversational dead time for second language education. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems, CHI EA '14*, page 2239–2244, New York, NY, USA, 2014. Association for Computing Machinery.
- [11] Eun Kyoung Choe, Bongshin Lee, Matthew Kay, Wanda Pratt, and Julie A. Kientz. Sleeptight: Low-burden, self-monitoring technology for capturing and reflecting on sleep behaviors. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '15*, page 121–132, New York, NY, USA, 2015. Association for Computing Machinery.
- [12] William L Cull. Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, 14(3):215–235, 2000.
- [13] David Dearman and Khai Truong. Evaluating the implicit acquisition of second language vocabulary using a live wallpaper. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12*, page 1391–1400, New York, NY, USA, 2012. Association for Computing Machinery.
- [14] Frank N Dempster. Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology*, 79(2):162, 1987.
- [15] Tilman Dingler, Dominik Weber, Martin Pielot, Jennifer Cooper, Chung-Cheng Chang, and Niels Henze. Language learning on-the-go: Opportune moments and design of mobile microlearning sessions. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '17*, New York, NY, USA, 2017. Association for Computing Machinery.
- [16] Darren Edge, Elly Searle, Kevin Chiu, Jing Zhao, and James A. Landay. Micromandarin: Mobile language learning in context. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11*, page 3169–3178, New York, NY, USA, 2011. Association for Computing Machinery.
- [17] Denzil Ferreira, Jorge Goncalves, Vassilis Kostakos, Louise Barkhuus, and Anind K. Dey. Contextual experience sampling of mobile application micro-usage. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '14*, page 91–100, New York, NY, USA, 2014. Association for Computing Machinery.
- [18] Pascal E. Fortin, Yuxiang Huang, and Jeremy R. Cooperstock. Exploring the use of fingerprint sensor gestures for unlock journaling: A comparison with slide-to-x. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '19*, New York, NY, USA, 2019. Association for Computing Machinery.
- [19] Marian Harbach, Emanuel Von Zezschwitz, Andreas Fichtner, Alexander De Luca, and Matthew Smith. It's a hard lock life: A field study of smartphone (un) locking behavior and risk perception. In *10th Symposium On Usable Privacy and Security (SOUPS2014)*, pages 213–230, Berkeley, CA, USA, 2014. USENIX Association.
- [20] Daniel Hintze, Rainhard D. Findling, Sebastian Scholz, and René Mayrhofer. Mobile device usage characteristics: The effect of context and form factor on locked and unlocked usage. In *Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia, MoMM '14*, page 105–114, New York, NY, USA, 2014. Association for Computing Machinery.
- [21] Daniel Hintze, Philipp Hintze, Rainhard D Findling, and René Mayrhofer. A large-scale, long-term analysis of mobile device usage characteristics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(2):1–21, 2017.
- [22] Helen M. Hodgetts and Dylan M. Jones. Contextual cues aid recovery from interruption: The role of associative activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(5):1120–1132, 2006.
- [23] Jisu Jung, Monica Nour, Margaret Allman-Farinelli, and Judy Kay. Harnessing the “ambience” of the mobile-phone lockscreen for ultra-lite logging. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction, OZCHI*

- '17, page 21–30, New York, NY, USA, 2017. Association for Computing Machinery.
- [24] John Keller and Katsuaki Suzuki. Learner motivation and e-learning design: A multinationally validated process. *Journal of educational Media*, 29(3):229–239, 2004.
- [25] Kathy A Krendl and Mary Broihier. Student responses to computers: A longitudinal study. *Journal of Educational Computing Research*, 8(2):215–227, 1992.
- [26] Agnes Kukulska-Hulme. Chapter one: Language learning defined by time and place: A framework for next Generation designs. In *Left to my own devices: Learner autonomy and mobile-assisted language learning*, pages 1–20. Brill, Bingley, UK: Emerald Group, 2012.
- [27] Alexandra Kuznetsova, Per B. Brockhoff, and Rune H. B. Christensen. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13):1–26, 2017.
- [28] Chun Lai and Dongping Zheng. Self-directed use of mobile devices for language learning beyond the classroom. *ReCALL*, 30(3):299–318, 2018.
- [29] Ahmed Mahfouz, Ildar Muslukhov, and Konstantin Beznosov. Android users in the wild: Their authentication and usage behavior. *Pervasive and Mobile Computing*, 32:50–61, 2016.
- [30] Nathan Malkin, Marian Harbach, Alexander De Luca, and Serge Egelman. The anatomy of smartphone unlocking: Why and how android users around the world lock their phones. *GetMobile: Mobile Computing and Communications*, 20(3):42–46, 2017.
- [31] Elizabeth J Marsh, Pooja K Agarwal, and Henry L Roediger III. Memorial consequences of answering sat ii questions. *Journal of Experimental Psychology: Applied*, 15(1):1, 2009.
- [32] Paul Meara. Some notes on the eurocentres vocabulary tests. *Foreign language comprehension and production*, pages 103–113, 1990.
- [33] Paul Meara and Glyn Jones. Vocabulary size as a placement indicator. In Grunwell, Pamela, Ed. *Applied Linguistics in Society*. 20th Annual Meeting of the British Association for Applied Linguistics, 1988.
- [34] Lotte Meteyard and Robert A.I. Davies. Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, 112:104092, 2020.
- [35] Lorelli S Nowell, Jill M Norris, Deborah E White, and Nancy J Moules. Thematic analysis: Striving to meet the trustworthiness criteria. *International journal of qualitative methods*, 16(1):1609406917733847, 2017.
- [36] Sarah Prange, Lukas Mecke, Alice Nguyen, Mohamed Khamis, and Florian Alt. Don't use fingerprint, it's raining! how people use and perceive context-aware selection of mobile authentication. In *Proceedings of the International Conference on Advanced Visual Interfaces*, pages 1–5, 2020.
- [37] Lina Qiu, Alexander De Luca, Ildar Muslukhov, and Konstantin Beznosov. Towards Understanding the Link Between Age and Smartphone Authentication, page 1–10, New York, NY, USA, 2019. Association for Computing Machinery.
- [38] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2020.
- [39] Henry L Roediger III and Elizabeth J Marsh. The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5):1155, 2005.
- [40] Daniel L Schacter. The seven sins of memory: insights from psychology and cognitive neuroscience. *American psychologist*, 54(3):182, 1999.
- [41] Christina Schneegass and Fiona Draxler. Cognitive biases and their effect on mobile learning: The example of the continued influence bias and negative suggestion effect. In *Workshop on Detection and Design for Cognitive Biases in People and Computing Systems at the 2020 CHI Conference on Human Factors in Computing Systems (CHI'20)*, 2020.
- [42] Christina Schneegass, Sophia Sigethy, Malin Eiband, and Daniel Buschek. Comparing concepts for embedding second-language vocabulary acquisition into everyday smartphone interactions. In *Mensch und Computer 2021*, pages 11–20, 2021.
- [43] Behzad Tabibian, Utkarsh Upadhyay, Abir De, Ali Zareade, Bernhard Schölkopf, and Manuel Gomez-Rodriguez. Enhancing human learning via spaced repetition optimization. *Proceedings of the National Academy of Sciences*, 116(10):3988–3993, 2019.
- [44] Khai N. Truong, Thariq Shhipar, and Daniel J. Wigdor. Slide to x: Unlocking the potential of smartphone unlocking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14*, page 3635–3644, New York, NY, USA, 2014. Association for Computing Machinery.
- [45] Andrew Trusty and Khai N. Truong. Augmenting the web for second language vocabulary learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11*, page 3179–3188, New York, NY, USA, 2011. Association for Computing Machinery.
- [46] Emanuel von Zezschwitz, Paul Dunphy, and Alexander De Luca. Patterns in the wild: A field study of the usability of pattern and pin-based authentication on mobile devices. In *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '13*, page 261–270, New York, NY, USA, 2013. Association for Computing Machinery.
- [47] Chen Wang, Yan Wang, Yingying Chen, Hongbo Liu, and Jian Liu. User authentication on mobile devices: Approaches, threats and trends. *Computer Networks*, 170:107118, 2020.
- [48] Tingxin Yan, David Chu, Deepak Ganesan, Aman Kansal, and Jie Liu. Fast app launching for mobile devices using predictive user context. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services, MobiSys '12*, page 113–126, New York, NY, USA, 2012. Association for Computing Machinery.
- [49] Xiaoyi Zhang, Laura R. Pina, and James Fogarty. Examining unlock journaling with diaries and reminders for in situ self-report in health and wellness. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16*, page 5658–5664, New York, NY, USA, 2016. Association for Computing Machinery.