




Article

# Investigating the Use of Task Resumption Cues to Support Learning in Interruption-Prone Environments

Christina Schneegass \*, Vincent Füseschi, Viktoriia Konevych and Fiona Draxler \* 

LMU Munich, Media Informatics Frauenlobstr. 7a, 80337 Munich, Germany; vincent.fueseschi@icloud.com (V.F.); victoriashkurat24@gmail.com (V.K.)

\* Correspondence: christina.schneegass@ifi.lmu.de (C.S.); fiona.draxler@ifi.lmu.de (F.D.)

**Abstract:** The ubiquity of mobile devices in peoples' everyday life makes them a feasible tool for language learning. Learning anytime and anywhere creates great flexibility but comes with the inherent risk of infrequent learning and learning in interruption-prone environments. No matter the length of the learning break, it can negatively affect knowledge consolidation and recall. This work presents the design and implementation of memory cues to support task resumption in mobile language learning applications and two evaluations to assess their impact on user experience. An initial laboratory experiment ( $N = 15$ ) revealed that while the presentation of the cues had no significant effect on objective performance measures (task completion time and error rate), the users still perceived the cues as helpful and would appreciate them in a mobile learning app. A follow-up study ( $N = 16$ ) investigated revised cue designs in a real-world field setting and found that users particularly appreciated our interactive test cue design. We discuss strengths and limitations of our concept and implications for the application of task resumption cues beyond the scope of mobile language learning.



**Citation:** Schneegass, C.; Füseschi, V.; Konevych, V.; Draxler, F. Investigating the Use of Task Resumption Cues to Support Learning in Interruption-Prone Environments. *Multimodal Technol. Interact.* **2022**, *6*, 2. <https://doi.org/10.3390/mti6010002>

Academic Editor: Mu-Chun Su

Received: 25 August 2021

Accepted: 19 December 2021

Published: 30 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** mobile learning; interruptions; task resumption; memory cues; laboratory experiment; field study

## 1. Introduction

Mobile devices enable us to learn outside the classroom, whenever and wherever we want. Especially applications based on the micro-learning paradigm are used in a variety of situations, such as at home, at work, in libraries, or on public transportation during commutes [1,2]. Applications following the micro-learning approach favor high repetition counts over long streaks of continuous and attentive learning interaction [3]. This method enables users to integrate learning sessions into their everyday lives even when they can only grant learning a short amount of time before moving on to a different task. While this flexibility is a great advantage of mobile learning applications, it comes with the inherent cost caused by having to switch between the learning task and the environment. For example, using mobile learning apps on the commute to work can get interrupted by the learner having to switch from train to bus; learning in bed late at night can be cut short because the learner is getting tired; using a mobile learning app in the waiting room of the doctor can get disturbed by being called into the doctor's office. These interruptions and task switches make it challenging to engage in a topic for longer sessions which are needed for proper information processing and memory consolidation.

In general, even short interruptions can have a severe effect on the performance of the primary task, in particular, regarding error rate and task completion time [4–6]. Longer interruptions or learning breaks can lead to memory decay if the content is not rehearsed frequently [7]. To support mobile learning despite interruptions, this work will explore a new mobile learning application feature to mitigate the negative effects of interruptions. While prior work on interruption delay and management showed great

potential, the variability of everyday contexts in which mobile learning apps are used leads to interruptions that can rarely be predicted or postponed. Thus, in this work we will focus on supporting users in the learning task resumption by designing memory cues to reactivate the memory of prior learning sessions to compensate for disruptive effects of interruptions.

We explore different designs for memory cues to help provide a seamless connection between individual mobile learning units. In a first study ( $N = 15$ ), we present four different cues derived from related literature and evaluate their mitigation potential in a lab-based mobile learning study. Participants reported to perceive the cues as helpful and supportive, however, the quantifiable effect on error rate and completion time were negligible. As some participants reported not feeling very interrupted due to the controlled lab setting, we followed up with a second in-the-wild evaluation ( $N = 16$ ). For this purpose, we revised the cues according to the feedback gathered in the first study and embedded it again in a mobile language learning application. In this second study, we observed great differences in participants' individual preferences regarding the cue design, while the interactive test cue was appreciated by the majority of our participants, their opinion on the other cue designs was mixed. We discuss the results of our study and outline implications for the application of task resumption cues in learning applications and beyond.

## 2. Background

### 2.1. Mobile Learning and Micro-Learning

With the use of the micro-learning approach, mobile devices can be used for teaching content anywhere and anytime [8]. Through high repetition rates displayed as *micro-content* units in *micro-interactions*, learning on mobile devices has proven to be well-suited for the use on the go, in particular to make use of idle situations [9]. Prior surveys have shown that users learn with their mobile device when and where they have the opportunity to do so [2], which can represent a variety of situations, from home, over public transportation to public spaces such as libraries [1,2,8]. To allow learning even in short breaks, mobile learning apps give preference to simple interactions (e.g., buttons over typing text [10]) and content that can be broken down into small chunks (e.g., vocabulary translations over complex grammar knowledge [11]). These constraints originate in the usage context but also in the limitations of the device itself. Mobile phones come with limited screen space that benefits from employing simple interactions with immediate feedback [12]. In contrast, these devices also come with capabilities that can be used to support the learner, such as computational power, context-awareness through sensors, adaptation and personalization features, and their sheer ubiquity in our everyday life.

### 2.2. Interruptions and Memory Cues

In our everyday life, multitasking, especially with technology, has become a common phenomenon. We eat while watching television, call a friend while driving the car, or play games while riding the bus. In such situations, both the technology and the environment can become a source of distraction and interruption. In this work, along the interruption process outlined by Trafton et al. [13], we define an *interruption* as an event that draws the user's attention away from the primary task, in our case the learning application on the mobile phone. The events or action leading to an interruption (so-called *secondary tasks*) can be of different origins. In related literature, the most common differentiation is between *external* (environment, e.g., people approaching or having to switch trains), *internal* (user, e.g., feeling tired or experiencing mind-wandering), or *device* (smartphone, e.g., receiving a notification or call) interruptions [5,14]. Prior work by Draxler et al. [15] confirmed the prevalence of interruptions during mobile learning and observed that interruptions often lead to suspension and termination of the learning activities.

Even though not all of these interruptions are equally demanding to a user, prior work has shown that being interrupted can have severe negative consequences on the

execution of the primary task. For example, an interruption can affect the primary task's error rate and completion time [4–6]. In particular, unpredictable interruptions can cause stress [16] and long interruptions can entail additional time to resume the primary task [17]. Especially in the phase of early memory consolidation, an essential process in learning, interruptions can have permanent disruptive effects [18,19].

The Memory-for-Goals theory [20] describes interruptions as a suspension of the primary task's goal. The goal can be retrieved with the help of priming through a memory cue, triggering the recall of previously stored information from one's long-term memory [20]. To restore the task context of the primary task, Trafton et al. [13] outline two directions for goal encoding, namely: (1) retrospective ("What was I doing before?") and (2) prospective ("What was I about to do?"). Memory cues can support both ways of encoding. However, prospective memory cues are best encoded before the interruption (e.g., by noting down the next steps one wanted to take), whereas retrospective cues can be presented after the interruption without prior priming. Examples of such task resumption cues are cues that support resuming a task, are push-notifications, bookmarks, or summaries.

Since many interruptions in users' everyday lives cannot be anticipated or avoided, such as people approaching or having to switch trains, we will focus on the investigation of task resumption cues. These cues are presented after the interruption to mitigate their adverse effects and restore task context.

### 2.3. Task Resumption Support

There are various ways of supporting a learner in resuming a learning task. From a pedagogical perspective, putting the learner in the prior lesson's context is called memory re-activation, while *open* memory re-activation aims to guide learners back to a certain context or broad topic, and *specific* memory re-activation targets one specific chunk of information, e.g., by posing a question about it [21].

This technique of re-activating a piece of information from the users' long-term memory can also be referred to as *cuing*. Memory cues are stimuli that trigger users' memories, either on an implicit level using subtle highlights (e.g., [22]) or complex information to restore the full context of the primary task (e.g., [23]). In the HCI domain, memory or task resumption cues have been evaluated in a variety of interruption scenarios. They can be applied to support users resuming reading after an interruption, support the transition between multiple devices, or facilitate multi-tasking in critical jobs such as emergency operators.

Very little research has been done in applying this task resumption cue concept to the domain of mobile learning, although learning is a common task on mobile devices that requires a certain level of attention and is, therefore, affected by interruptions. Schneegass and Draxler [14] performed a structured literature analysis on task resumption cues in HCI applications. They show that visual cues have been frequently deployed and positively evaluated, such as implicit bookmarks generated by users' gaze [24,25], visual or auditory labels representing the task context before an interruption [26,27], or visuliazations of the last activities before an interruption [23,28].

None of the 30 publications describing task resumption cues showed their application in mobile learning scenarios, but mainly concerned stationary settings with large and/or multiple screens. Schneegass and Draxler [14] proposed a set of guidelines on how to design and implement memory cues for the specific use case of learning and stressed that task resumption cues need to be adapted to the task at hand and to the specific requirements of mobile devices.

### 2.4. Task Resumption Features in Mobile Learning Applications

When looking at learning applications commonly available on the market, we see that they implement subtle forms of memory cues. For example, the apps Duolingo (Duolingo: <https://www.duolingo.com/learn>, accessed on 26 September 2021) or Busuu (Busuu: <https://www.busuu.com/en>, accessed on 26 September 2021) send reminder notifications

after periods of inactivity and suggest learners to continue learning where they left off. The mobile app for Khan Academy (Khan Academy: <https://www.khanacademy.org/>, accessed on 26 September 2021) allows users to set bookmarks and shows the latest activity under the tab “recent lessons” with the name of the lesson. These features can be classified as implicit visual memory cues functioning as specific reminders. When the learning app further proposes to repeat content that has been previously taught immediately after entering the app, this revision can be classified as an interactive retrospective cue.

**Research Gap:** Prior work has investigated the effects of interruptions and proposed several ideas for supporting task resumption in desktop or multi-device settings. However, research investigating the use of task resumption or memory cues in mobile environments or for the specific application of mobile learning is sparse. Due to the high prevalence of interruptions in mobile environments and their severe effect on mobile learning (cf. [15]), we emphasize the need for the implementation of task resumption features that go beyond existing approaches in this use case. Furthermore, the effectiveness of task resumption cues has yet to be evaluated in this context. We aim to close this research gap by proposing designs for memory cues adapted to the specific use case of mobile learning and evaluating their effect in a laboratory and in-the-wild setting.

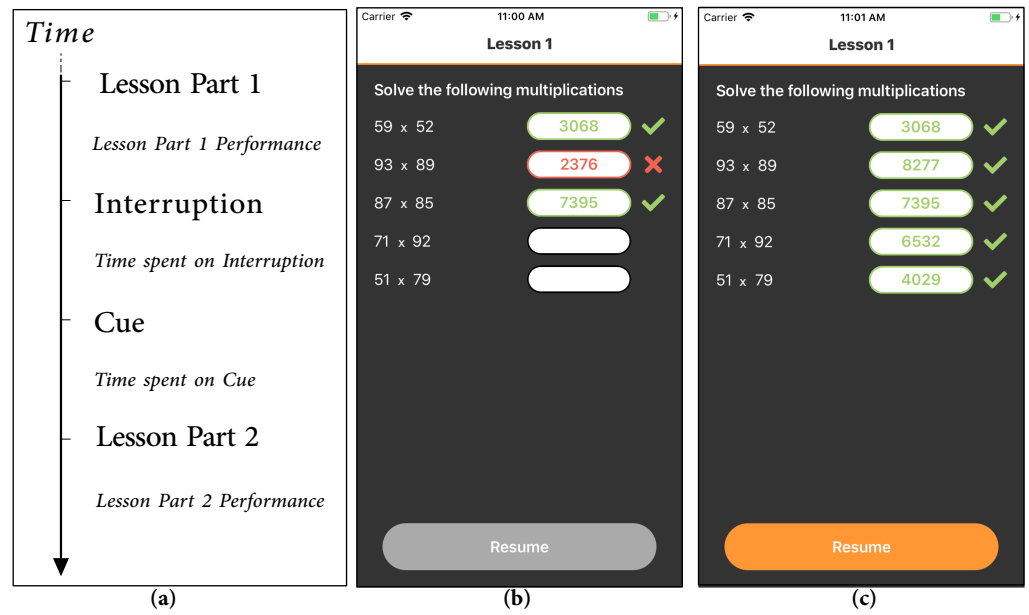
### 3. Implementation

We developed an iOS-based mobile learning application (Swift 4.2.1) to evaluate different task resumption cues based on prior work and adapted to a mobile learning context. The application was aligned (in terms of design, tasks, and content) with language learning applications such as Duolingo and Memrise, as they apply a micro-learning approach and are commonly used. We followed the basic structure of implementing multiple-choice question-based learning. We decided to implement a beginners’ level language learning course using a language rarely spoken in our university’s country to avoid effects caused by prior knowledge. Still, the language should not require teaching a new alphabet. Thus, we decided to use Polish for the sake of this user study.

When first starting the application, a login screen asked participants to enter a randomly assigned ID to match their interaction and performance data to the questionnaires and feedback. Afterward, the app presented an overview screen of the five lessons we implemented for this study. Each lesson consisted of two parts with a set of questions stored as JSON objects. In the first part with around twelve questions, the user is shown images or icons and has to relate them to the Polish words. In the second part (around eight questions), the knowledge acquired in the first half needs to be applied through multiple-choice or drag-and-drop recognition tasks. The app displayed corrective feedback immediately after answering using color coding, a brief message (green highlight, “You are correct!”), and if incorrect the correct solution for the task (red highlight, “Correct solution:”). We further used SnapKit (SnapKit: <https://snapkit.io/>, accessed on 17 October 2021) (version 4.0.1) for the layout of the app.

#### 3.1. Lesson Design

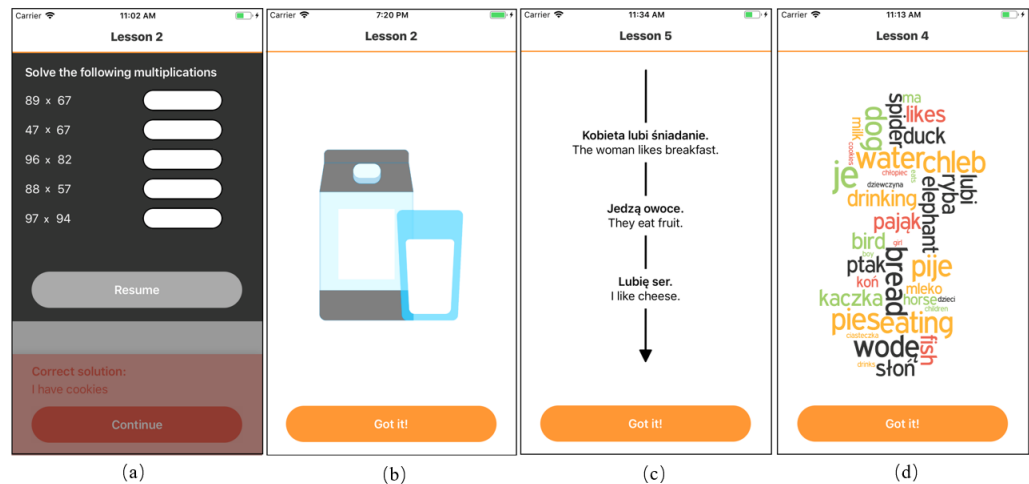
The app included several self-contained vocabulary lessons of 20 questions, grouped into lessons by topics such food or clothing. Every lesson consisted of two parts (cf. Figure 1a) and focused on explicitly teaching vocabulary and implicitly teaching simple grammar constructs through sentence building tasks. The first part of each lesson consisted of active and passive recognition tasks [29], in which users had to select the correct Polish/native translation for a displayed Polish/native word from a set of alternatives (multiple-choice). Pictures were used to help the initial acquisition of new words. The second task format was to translate a Polish/native sentence by assembling (Polish/native) words through sequential selection from a pool of words. The sentence-building task is considered more difficult as it presents more options and thus, does not allow for mere elimination of incorrect answer options to solve the task.



**Figure 1.** (a) Timeline of the lesson structure: The interruption occurred after the first lesson part and is followed by the cue presentation before the learning resumes with a second lesson part. (b) Interruption screen showing five multiplication tasks in progress and (c) correctly solved ready to resume the learning task. In the *Half-Screen Cue* condition the multiplication task covered only half the screen.

3.2. Interruptions

To distract participants from the learning task and test the effect of the task resumption cue designs, we interrupted them during lessons. We chose to use mathematical tasks as interruption as it has been previously used in other studies as interruption source (cf. [30–32]). Inside the application, the users were shown a series of five double-digit multiplication tasks they had to solve before they could continue learning. The app presented the mathematical tasks as a screen overlay (see Figure 2d).



**Figure 2.** The four task resumption cue designs created for the application in mobile learning, (a) *Half-Screen Cue*, (b) *Image Cue*, (c) *History Cue*, and (d) *WordCloud Cue*. The cues are presented when the user resumes the learning task after an interruption to help task resumption.

3.3. Task Resumption Cues

We designed two implicit and two explicit memory cues for our language learning application. The four different designs were presented as a full-screen overlay once the user re-enters the learning application after an interruption. We randomized the order of

the cue presentation. The users can view the displayed cue as long as they want and press a button labeled “Got it!” at the bottom of each cue view to continue with the next question.

### 3.3.1. *Half-Screen Cue*

The *Half-Screen Cue* left a part of the primary task interface (learning app) visible while the secondary task (multiplications) was performed (see Figure 2d). Thus, in contrast to the other resumption cues, it was shown during, and not after, the interruption. This cue is based on several studies conducted in desktop settings. For example, when the primary task interface remained partially visible, study participants were better at maintaining a spatial representation of the primary task [33], more quickly returned to prior tasks [34], and had a shorter resumption lag [35]. We consider the *Half-Screen Cue* an implicit cue because it does not include any additional information about the current topic.

### 3.3.2. *Image Cue*

In this cue, an image or graphical symbol is shown representing the lesson the user interacted with before an interruption (see Figure 2c). This form of visual memory cue was already suggested in prior work (cf. [36]). We consider it an implicit cue because it is a very simple reminder hinting the content of the prior lesson. Further, Chen et al. [37] noted that instructions in learning should be targeted to the individual learner, while some benefit more from verbal information (e.g., words in the *WordCloud Cue*), others are better supported with visual information such as images.

### 3.3.3. *History Cue*

The most explicit memory cue to support the resumption of the learning task is designed to show the user their progress over time. As shown in Figure 2a, this cue visualizes the last questions the user answered in L2 in chronological order as well as their solutions in L1 below, thus, providing a sense of context to them. Visualizing progress history has been explored in prior work, for example in programming settings [28], search tasks [38], and mission command or aircraft tasks, where it has shown its potential to increase accuracy and performance [23,39].

### 3.3.4. *WordCloud Cue*

This memory cue displays words learned in the lesson prior to the interruption in the form of a tag cloud (see Figure 2b). It is closely related to the *History Cue* as it is a summary of content learned before the interruption. However, it is less structured since it does not include a temporal component of when the word was presented or learned but gives a more general overview. The concept of word or tag clouds gained popularity in other application areas for the summarization of text analysis tasks [40] or search results [41].

## 4. Lab-Based User Study

We performed a within-subject laboratory-based user study investigating the effect of the four different cue types. The implementation of task resumption cues in prior work affected the users’ interaction with the respective system. These effects can be evaluated using subjective and objective metrics. As objective metrics, prior work frequently focused on assessing the resumption time (e.g., [34,35,42–44]), or changes in the overall task completion time. The metric measures the time the user needs to resume the task, which is usually increased due to interruptions. The prolonged resumption time is called “resumption lag” (cf. [45,46]). Further, the error rate can be an indicator that the interruption affected the users’ performance. Therefore, error rates are used to measure task resumption cue effectiveness (cf. [22,23,42,47]).

The analytics data was logged in a CSV file via the open-source secure logging framework SwiftyBeaver (SwiftyBeaver: <https://swiftybeaver.com/>, accessed on 17 October 2021) (version 1.6.1) and included user id, lesson id, task id, correctness, cue type, time

stamp, and if the task occurred before or after the interruption. All logfiles were stored locally on the device used for the laboratory study.

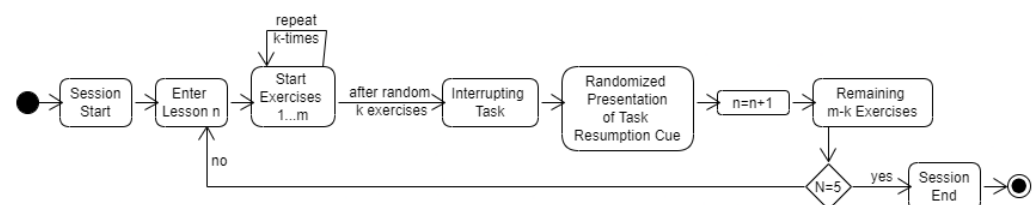
#### 4.1. Study Design

We performed our within-subject study investigating the effect of the four different cue types and an additional no-cue condition (independent variable) on the participants' task performance. In particular, we measured the error rate as well as the answer duration (dependent variables) after the interruption occurred and the cue was shown. Based on prior work we hypothesized that both error rate and answer duration decrease when the task resumption is guided by a memory cue as compared to the control condition without a cue. We expected the explicit cues (*WordCloud Cue*, *History Cue*) to have a stronger effect, thus decreasing error rate and answer duration more than the implicit cues (*Half-Screen Cue*, *Image Cue*) and no cue condition. The order of presentation of the cues was counterbalanced over the course of the five content lessons. Each lesson was interrupted once and the interruption was either followed by one of the cues or resumed immediately (no cue condition). Furthermore, we assessed the perceived helpfulness (dependent variable) of the different cues through Likert-scale ratings and a qualitative interview after the study was completed.

#### 4.2. Procedure

At the beginning of the study, participants were informed about the procedure and asked for consent. Then, they filled in a questionnaire to assess demographics and prior knowledge as well as experience with language learning applications. Next, the participants were given the study task—to complete five lessons in the mobile learning application designed for this study. Figure 3 visualizes the functionality within the study's application: Each lesson was once interrupted after a randomized number of exercises. After the interruption, a task resumption cue (or no cue) was presented and the lesson was continued until all five lessons were solved.

We provided pen and paper to help solve the multiplication interruption tasks and logged every interaction with the application to assess answer duration and error rate. As the task time was measured, we asked our participants not to take breaks during the lessons but instead between two lessons. However, they were not encouraged to rush but take as much time as they needed to answer the presented tasks correctly conscientiously. Finally, we conducted post-hoc semi-structured interviews to inquire about the perception of the different cues. The guiding questions of the interview concerned their general opinions of the cues, the cues' helpfulness, and the match of cue types for materials of different complexity levels.



**Figure 3.** The process of our application as used for the lab evaluation. Interrupting tasks were triggered randomly and were always followed by a task resumption cue. The study ended after all five sessions were completed.

#### 4.3. Sample

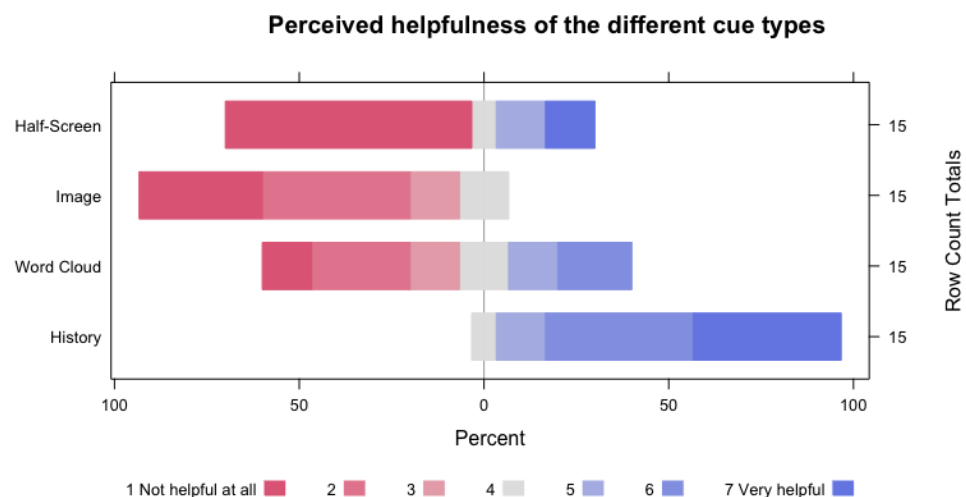
We invited participants through university mailing lists and social media channels, resulting in a set of 15 participants (8 male, 7 female) ranging between 20 and 33 years of age ( $M = 23.7$ ,  $SD = 3.6$ ). None of the participants had prior knowledge of Polish or any closely related languages. The user study took around 1 h, and as compensation for the participation, everyone received a 10€ voucher or an equal amount of study credit points.

#### 4.4. Results

Our data set included 75 learning sessions (five per participant), 75 interruptions, and the presentation of 60 cues (one no-cue condition per participant). Post-hoc ratings showed that the lesson content was challenging for the participants, as three participants rated it “very difficult”, seven as “somewhat difficult”, five as “adequate”, and none as (somewhat/very) easy. Out of 15 participants, 14 would appreciate it if language learning applications displayed task resumption cues as presented in this study for everyday usage.

##### 4.4.1. Perceived Helpfulness of Cues

We asked our participants to rate the helpfulness of the cue types on a 7-point Likert scale from 1 (Not helpful at all) to 7 (very helpful), while the *Half-Screen Cue* ( $M = 2.53$ ,  $SD = 2.28$ ), *Image Cue* ( $M = 2.07$ ,  $SD = 1$ ), and *WordCloud Cue* ( $M = 3.47$ ,  $SD = 1.75$ ) were perceived as semi to little helpful, participants valued the *History Cue* ( $M = 6.13$ ,  $SD = 0.88$ ) and stated that it was the most helpful by far (cf. Figure 4). A (non-parametric) Friedman test showed that there were significant differences between the perceived helpfulness of different cue types ( $\chi^2 = 26.1$ ,  $p < 0.001$ ; Kendall's  $W = 0.37$ ). Post-hoc Conover comparisons with Bonferroni correction revealed significantly higher helpfulness ratings of the *History Cue* over the *Half-Screen Cue* ( $t = 4.45$ ,  $p < 0.001$ ), the *History Cue* over the *Image Cue* ( $t = 4.30$ ,  $p < 0.001$ ), and the *History Cue* over the *WordCloud Cue* ( $t = 2.62$ ,  $p < 0.05$ ).



**Figure 4.** Perceived helpfulness of the different cue types.

##### 4.4.2. Post-Hoc Interviews

During the interview, participants confirmed that they liked the *History Cue* most. However, most participants did not notice that the questions displayed there were the last three before the interruption but assumed the cue displayed a random set of questions. One participant stated that they preferred the *WordCloud Cue* over the *History Cue* as it presents more details. As the *History Cue* can technically present more complex information than the *WordCloud Cue*, participants suggested to use it for presenting grammar knowledge and the *WordCloud Cue* for vocabulary.

For the *Image Cue*, four participants reported not noticing the cue or not looking at it at all. In particular, one of the *Image Cues*, an image of the Polish flag (horizontal white and red stripe), was not recognized by two of our participants. One of these participants thought it might be an indicator of correct and incorrect questions. Out of the 15 participants, only five noticed the *Half-Screen Cue*, whereof three thought it might be a bug in the interface.

In summary, all participants considered the implementation of task resumption cues to be a very helpful feature for mobile learning apps. The actual helpfulness, however, depends on the design of the cues.



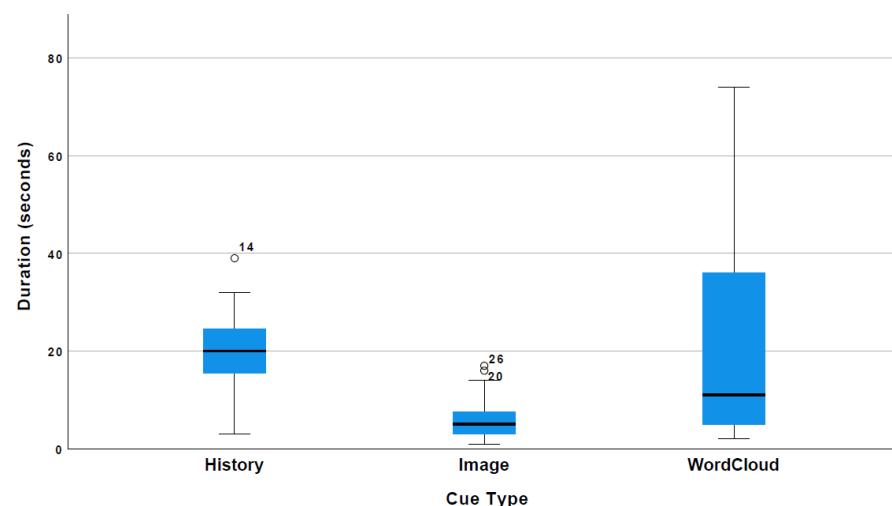
#### 4.4.3. Cue View Duration

The view duration of the different cues varied among the three cue types and among participants (as the *Half-Screen Cue* was implicitly embedded in the interrupting task, this cue has no view duration). Figure 5 shows that the *WordCloud Cue* was examined by the users with the greatest diversity in duration, between 2 and 74 s, with an average viewing time of 22.33 s ( $SD = 20.67$ ). In comparison, the *History Cue* viewing duration shows less variety among participants, ranging from 3 to 39 s but is higher on average ( $M = 19.8$ ,  $SD = 9.34$ ). Lastly, the *Image Cue* is viewed for the shortest duration between 1 and 17 s ( $M = 6.27$ ,  $SD = 5.12$ ). We found a significant difference in means between the three types (one-way ANOVA (A Kolmogorov–Smirnov test indicated a violation of the normality assumption ( $p > 0.05$ ). However, due to the robustness of ANOVAs in regard to this violation, we continued with this analysis),  $F = 5.82$ ,  $p < 0.01$ ) and post-hoc comparisons with Bonferroni correction revealed significant differences between *History Cue* and *Image Cue* ( $p > 0.05$ ) and *WordCloud Cue* and *Image Cue* ( $p > 0.01$ ).

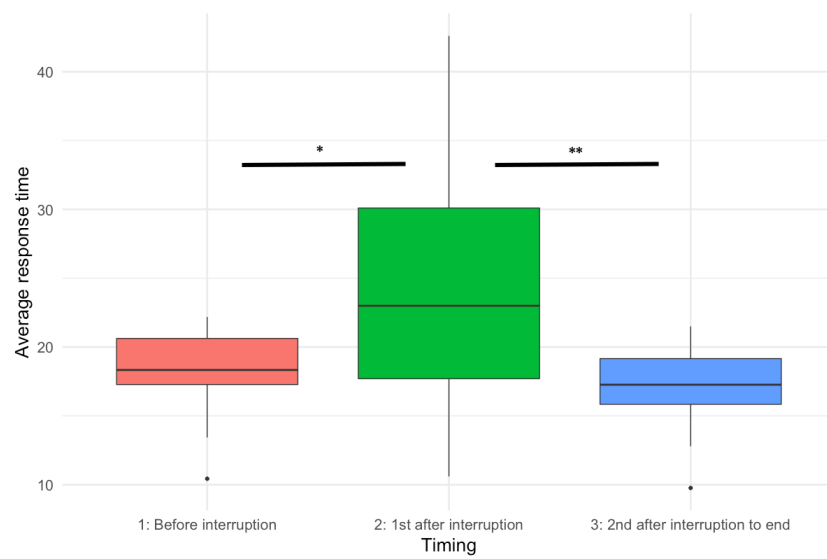
#### 4.4.4. Task Completion Time

In general, we find that the response time for the first question after an interruption differed significantly from the questions before the interruption and all subsequent questions (repeated-measures ANOVA,  $F = 9.0$ ,  $p < 0.01$ , see Figure 6a), while the average response time across all exercises of all five conditions was 18.6 s, the average response time for the first question after an interruption was 24.24 s (cf. Table 1). Post-hoc comparisons with Bonferroni correction showed that the task completion time for tasks before an interruption was lower than for the first task after an interruption ( $t = -3.6$ ,  $p < 0.01$ ). Similarly, the task completion time for the first task after an interruption was higher than for all other following tasks ( $t = 3.7$ ,  $p < 0.01$ ). In other words, the first task after an interruption took users significantly longer to answer than any other task.

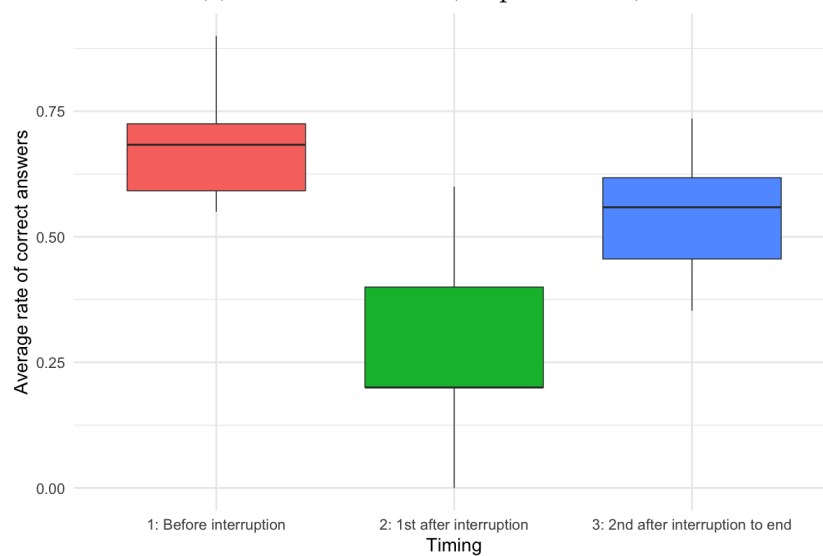
We further hypothesized that the presentation of task resumption cues can reduce the task completion time after an interruption. However, a repeated-measures ANOVA revealed no significant differences ( $p > 0.05$ ) when comparing the cue conditions with the no cue condition. In fact, the overall response time for all exercises following the interruption was lowest in the no cue condition ( $M = 16.1$  s,  $SD = 7.8$ ), followed by the *Half-Screen Cue*, *WordCloud Cue*, *Image Cue*, and *History Cue* (cf. Table 1).



**Figure 5.** The viewing duration in seconds of the three cue types *History Cue*, *Image Cue*, and *WordCloud Cue*.



(a) Answer Duration (Response Time)



(b) Correctness Rate

**Figure 6.** Effects of interruptions on (a) answer duration and (b) correctness rate. Asterisks indicate all statistically significant differences (\* < 0.05 | \*\* < 0.01).

**Table 1.** Response time and correctness rate with different cues for the first or all exercises after an interruption.

	No Cue	Half-Screen	Image	WordCloud	History
Response time 1st exercise (s)	19.5 (6.2)	22.5 (13.3)	30.6 (29.7)	23.2 (13.7)	25.4 (21.6)
Response time all exercise (s)	16.1 (7.8)	18.3 (11.0)	19.1 (15.9)	18.7 (12.4)	20.8 (16.8)
Correctness rate 1st exercise (s)	0.40 (0.51)	0.47 (0.51)	0.20 (0.41)	0.27 (0.46)	0.20 (0.41)
Correctness rate all exercise (s)	0.49 (0.34)	0.51 (0.32)	0.48 (0.29)	0.49 (0.29)	0.53 (0.34)

#### 4.4.5. Error Rate

In line with the results from the task completion time analysis, we found that the interruptions significantly affected the correctness of the first task after the interruption occurred (repeated-measures ANOVA,  $F = 69.3$ ,  $p < 0.001$ ; see Figure 6b). Post-hoc comparisons with Bonferroni correction revealed significantly higher error rates for the first task solved after an interruption compared to all tasks before ( $t = 11.74$ ,  $p < 0.001$ ) and all following tasks ( $t = -6.53$ ,  $p < 0.001$ ). Furthermore, we find higher error rates in tasks solved after an interruption (excluding the first task solved immediately after the interruption) compared to tasks solved before ( $t = 5.18$ ,  $p < 0.001$ ). i.e., participants made more errors in the exercises of the second lesson part, after they were interrupted.

Regarding the influence of our different cue types (either as four individual cases or as a cue/no-cue comparison), a Chi-square ( $\chi^2$ ) analysis did not show a significant effect of cue type on correctness (yes | no) after the interruption ( $p < 0.05$ ). Similarly, we also did not find an effect of the cue types on the correctness of the first task after an interruption (Figure 7).

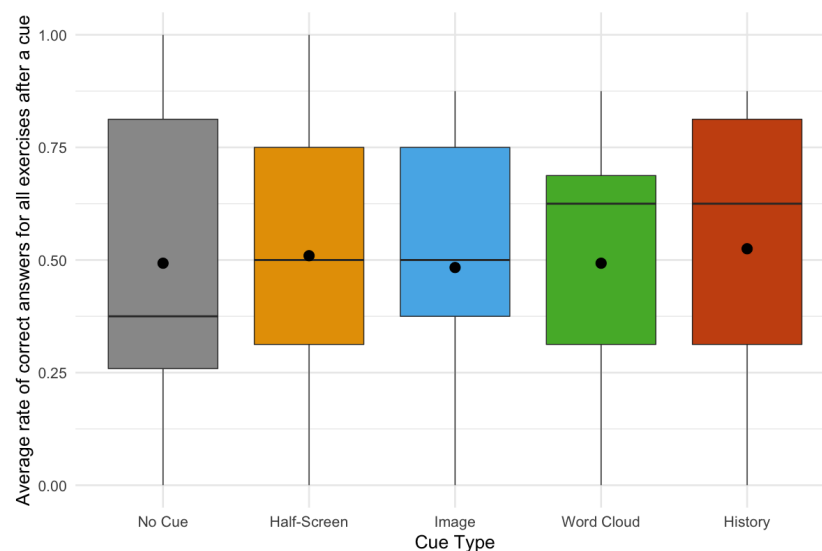


Figure 7. Performance after an interruption.

## 5. Discussion

### 5.1. Limitations

While this laboratory experiment aimed to assess the basic effects of interruptions on mobile learning performance, participants stated in the interviews that the situation felt very artificial. According to their perception, experiencing only one interruption type in an otherwise controlled setting did not reflect everyday situations. Therefore, they did not expect to be strongly influenced by the interruption and felt as if they could still recall the latest learning session. Future work is needed to assess the effect of task resumption cues with diverse interruptions in the wild to reflect real-world usage.

The data furthermore shows a difference in the difficulty of the presented learning lessons. Lesson 1 showed higher error rates and longer duration, which the interruptions or cues cannot explain as their presentation was counterbalanced to enforce randomization. We suspect that this perceived difficulty is due to Polish being a new language for our participants. We recommend increasing the number of learning tasks in future evaluations to overcome differences in task difficulty.

Moreover, we observed a limitation in the implementation of the *Half-Screen Cue*. In this cue design, the interrupting multiplication task UI overlays around 60% of the screen, while the lower third of the screen left the learning app visible. However, the remaining space was overlaid by the keyboard once participants started entering the results of the

multiplication tasks. Thus, the *Half-Screen Cue* was only visible for a certain amount of time and not during the whole interruption. Together with the fact that several participants did not notice the cue, the results in regards to the *Half-Screen Cue* should be viewed with a grain of salt.

### 5.2. Effects of Interruptions

In evaluating participants' performance across the different tasks, we observed that the interruptions affected their performance. Especially for the first task after an interruption, we recorded longer task completion times and higher error rates independent from any task resumption cue. For the remaining tasks after the interruptions, the performance improved again. Yet, participants expressed their doubt about the effect of the interruptions. In the interviews, they stated that the interruptions felt artificial and that they had no problem remembering the content of the learning session before the interruption. From the results of our study, we assume that while the interruptions were short and contained, they still (implicitly) affected our users' focus. We assume that real-world interruptions that take the users out of the context of the learning activity (mentally and physically) and potentially last significantly longer than our experimental interruptions will have a greater negative impact on the learners' performance. For future evaluations, we suggest including a greater variety of interruptions or choose a field study setting with a natural environment altogether.

### 5.3. Objective vs. Subjective Helpfulness of Cues

The quantitative analysis of our study results showed that the effect of the task resumption cues on task completion time and error rate was limited. Compared to the no-cue condition, the presentation of any of the four cue designs led to longer response times and no improvement in correctness in the exercises after the interruption. Nonetheless, all participants stated that the implementation of task resumption cues in mobile learning apps would be a helpful feature. While the *WordCloud Cue* was considered helpful for lessons introducing many new words to the vocabulary base, the participants could imagine the *History Cue* to be particularly helpful for grammar lessons (i.e., summarizing the prior lessons' rules). We hypothesize, aligned with the expectations expressed in [14], that the perceived helpfulness of the task resumption cues depends on the alignment of task complexity and resumption cue complexity. It is also possible that the effect of resumption cues is stronger for long-term retention than for immediate recall.

### 5.4. Participants Favor Explicit over Implicit Cues

Looking at the subjective helpfulness ratings, we observe that explicit cues appear to be more helpful in supporting participants resuming the learning tasks. Especially the *History Cue* was acknowledged as being helpful, while the *Image Cue* received the lowest helpfulness rating of all four cue designs. The viewing duration of the different cues indicates that our participants examined the content of the *WordCloud Cue* (on average eleven but up to 74 s) and *History Cue* (on average 19.8 s) in detail. Participants noted that the *WordCloud Cue* could be more helpful if the words were arranged according to a certain logic. While the visualization in this evaluation contained words of the prior lessons, colors and sizes could be adapted according to certain criteria such as correctness in participants' answers, frequency of occurrence in the lesson, or importance for the language in general. The *Image Cue* was viewed only briefly, and participants stated in the interviews that they did not perceive this type as very helpful. In the overall subjective helpfulness ratings, the *WordCloud Cue* and especially the *History Cue* ranked higher when compared to the *Half-Screen Cue* and *Image Cue*, suggesting participants' preference for explicit cues over implicit. Nonetheless, because many participants did not perceive the implicit cues as actual memory cues, we suggest that future work takes a deeper look into the effectiveness of implicit cues with revised designs.

## 6. Revised Implementation

The revised implementation followed two main goals: (1) transferring the evaluation of task resumption cues from the lab into the real world and (2) improving the task resumption cue designs based on feedback collected in the first study.

We decided to implement an Android application for the in-the-wild user study to reach a bigger user group. In this follow-up project, we iterated on the design of the task resumption cues. We embedded them in a similar mobile language learning Android application we developed (for Android 8 and above, Kotlin version 1.3) called *CzechWizard*, which teaches Czech at the beginners' level with a focus on vocabulary and short sentences (published as a closed test in the Google Play Store). We decided to use Czech for the same reasons as Polish before—it relies on the same base alphabet as German without being too similar, making it neither difficult nor easy to learn. Further, Czech is not commonly taught in schools, thus, making it possible to find participants with no knowledge of the language. Using a second language, we aim to diversify our results, simultaneously iterating on the lessons to tackle the problem from the lab evaluation, where participants reported tasks as too difficult.

The app provided a content overview with lessons on topics such as gender, the use of simple verbs such as “be”, or plural use. Each lesson again contained multiple “blocks” including content on the topic of the lesson. We derived several tasks, such as (1) relating vocabulary to images, (2) translating words using multiple-choice answer formats (with and without providing a visual representation of the vocabulary items), and (3) sentence building by selecting the correct words from a list of options. All learning contents were stored as a JSON object within the application. The blocks in each lesson were building on each other and increased in difficulty. Further, words learned in the early sessions were later used to build sentences. The app validated the correctness of the users' answers and presented visual feedback using the commonly applied color scheme of green (correct) and red (incorrect). In case of incorrect answers in multiple-choice answer formats, the correct answer from the set was additionally highlighted in green.

For the later analysis of the data, we logged all user interactions with the application including events such as initial login, app opened/closed, lesson opened/completed, words shown, questions shown, cues shown/answered (in case of the interactive cue), and questionnaires opened/finished. All events were saved with the user id, time stamp, and correctness if applicable, focussing on assessing the task completion time and the error rate in the interaction. The data and progress were stored temporarily locally and uploaded to a Google Firestore Database (<https://firebase.google.com/docs/firestore>, accessed on 17 October 2021) whenever an internet connection was available.

### 6.1. Revised Cue Designs

Similar to the cue set of the first evaluation in this chapter, the revised cue designs include two implicit and two explicit task resumption cues (cf. Figure 8). Since we cannot guarantee that the *Half-Screen Cue* would work reliably for all potentially interrupting applications on devices with different operating systems, we do not include this design in our further evaluation. Instead, we explore a new modality for an implicit cue, a tactile vibration pattern. Further, we extended the *Image Cue* from a single image to a more subtle but also more pervasive color and icon scheme. The *WordCloud Cue* remains mainly the same with minor adaptations in the display of the words. Lastly, we changed the *History Cue* to not contain an overview of prior lessons but to include interactive tasks on content of the prior lessons. In the following, we will outline the cue designs and the implemented changes compared to the earlier cue versions in more detail.

#### 6.1.1. Vibro-Tactile Cue

Figure 8a depicts the idea behind the *Vibro-Tactile Cue*, an implicit cue aiming to create a subtle association between the lesson and a vibration pattern. Whenever a user enters a lesson, the device issues a vibration. This type of subtle tactile feedback has been rec-

ommended as unobtrusive alternative to other modalities as it requires less attention and cognitive capacities. The vibration mechanism aims at creating an association between the stimulus and learning activity and directs the users' focal attention (cf. [48–51]). We specifically defined a vibration wave pattern that is not commonly used in other applications using the Google Vibrator Service (Vibrator Service: <https://developer.android.com/reference/android/os/VibrationEffect>, accessed on 17 October 2021).

### 6.1.2. Color and Icon Cue

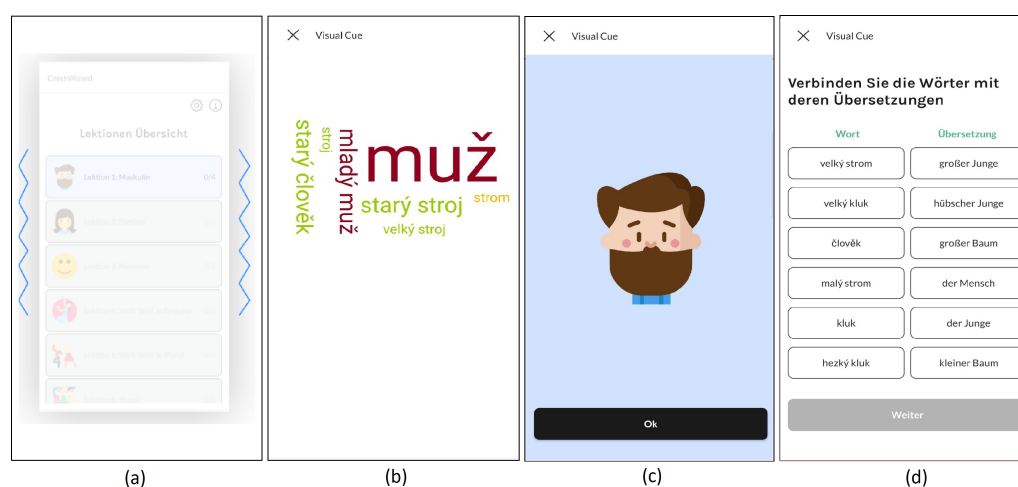
The second implicit cue design is the *Color and Icon Cue* (see Figure 8b). This cue shows an icon and color specifically associated with the lesson's theme the user is currently working on. The design extends the *Image Cue* presented in the first study. Like this cue, an image is chosen from the previous lesson and supplemented by a color. Groups of similar lessons received colors of a related color palette. This cue is inspired by the work of Yatid and Takatsuka [52] who used colors for categorical associations to the context.

### 6.1.3. WordCloud Cue

The *WordCloud Cue* remains similar to the design presented in the first part of this chapter (see Figure 8). The explicit cue generates an overview of words learned in the prior lessons. In contrast to the earlier design, the size of the words in the word cloud now indicate the learning progress. Words displayed in a larger font have been answered correctly more often than words in smaller font. Prior work of Ardissono et al. [53] showed that such type of a visualization model can help users to quickly find the relevant information when accessing a large amount of data. Further, for vocabulary learning, the cloud represents a summary of learning content.

### 6.1.4. Interactive Test Cue

While the *History Cue* received positive feedback in the first study, we decided to further improve the cue by adding interactivity. Since the *WordCloud Cue* already presents a passive visual summary of prior content, this second explicit cue, coined *Interactive Test Cue*, asks users to answer short questions (see Figure 8d). In particular, a screen presents six L1 words and their L2 translations, asking the user to match them by selection. Similar to the *WordCloud Cue*, the recognition tasks are generated from the vocabulary of the last lesson the user learned.



**Figure 8.** The four revised designs for task resumption cues for mobile language learning, (a) *Vibro-Tactile Cue*, (b) *Color and Icon Cue*, (c) *WordCloud Cue*, and (d) *Interactive Test Cue*.

## 7. In-the-Wild User Study

### 7.1. Study Design & Apparatus

To test the four resumption cue designs (independent variable), we embedded them in a mobile learning application compatible with Android 8 or higher. The cues were presented in a counterbalanced order independent from the current task to create a within-subject study design. Further, we included a no-cue condition to work as a baseline. After the interruption and the cue display, we measured the error rate and answer duration for the first task the user performed (dependent variables).

We postulate the following hypothesis:

**Hypothesis 1 (H1).** *Showing a cue after an interruption leads to lower task completion times and error rates as compared to showing no cue in the first five tasks after the interruption occurred.*

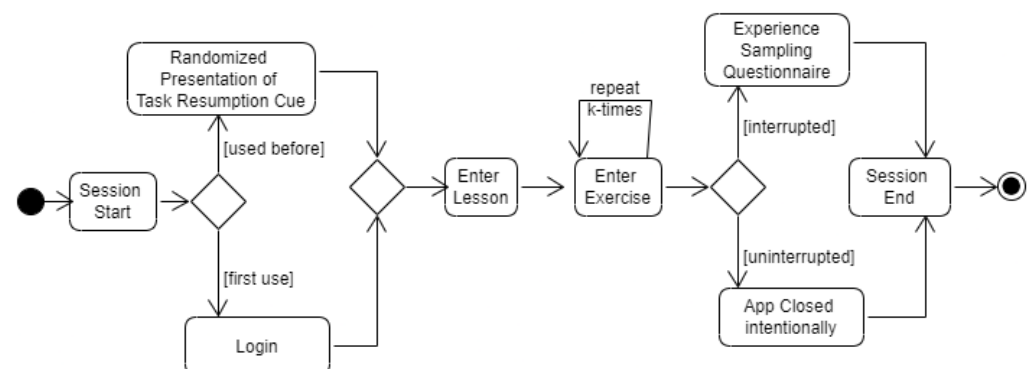
### 7.2. Procedure

After their recruitment, we informed the participants via email about the study procedure. They received an information sheet outlining our university's data protection policies and provided informed consent for the study. A first online questionnaire assessed general demographic information as well as the participants' language proficiency, in particular regarding Slavic languages. At the end of the survey, participants were guided through the installation process of the application downloadable through the Google Play Store. We encouraged the participants to use the app as naturally as possible in their everyday life while the app was logging their usage interactions (anonymized) with it (including but not limited to touch interactions, correctness of the tasks, and reported learning context).

As outlined in Figure 9, the app functionality was slightly adapted for this evaluation. When starting the app for the first time, users were directed to a login screen and could start learning. They entered lessons and solved the included exercises. Independently from whether they were interrupted or ended the session voluntarily, they were presented with a task resumption cue upon starting the app the next time. Additionally, when learning was stopped due to an interruption, an Experience Sampling Questionnaire (ESQ) was shown to collect data on the users' learning context.

After using the application for the study duration, a second questionnaire asked for the usability of the application and feedback on the task resumption cue designs.

Since the users' experience of the cue helpfulness can deviate from the effects observed through the objective measures, combining objective and subjective metrics can reveal further insights. Standardized questionnaires such as the System Usability Scale (SUS) [54] or specifically designed survey or interview questions can generate insights into the users' experience when interacting with task resumption cues in mobile learning applications.



**Figure 9.** A visualization of the app functionality revised for the field study. After entering the app, the user is presented with a randomized cue before starting a lesson and the respective exercises.

### 7.3. Sample

Participants were recruited through our university's mailing list as well as communication and social media channels. In total, 19 people started the study; however, two discontinued their participation after the first phase, excluding their data from this study. Therefore, our final sample size was 17 (twelve identifying as female, four as male, and one as non-binary). Their age ranged from 19 to 29 ( $M = 24$ ,  $SD = 3.17$ ), and they reported no experience with Slavic languages. Ten participants held an A-Level Diploma or equivalent, four a bachelor's degree, and three a master's degree or higher. Fourteen were currently enrolled in a study program; three were employed in full-time jobs. As we conducted this study during the COVID-19 pandemic. Thus, the results may be limited in generalizability as users' routines, mobility, and smartphone usage can deviate due to lockdown and work-from-home phases. We will discuss this issue in the section on limitations. For their participation, every participant received a 20 Euro voucher or study credit points as compensation.

### 7.4. Results

As this study took place in the wild, we asked participants to confirm the appearance of the cues during their learning sessions. The questionnaire first asked participants if they had actually noticed the cue by showing pictures of them. In total, five people stated to have not noticed the vibration pattern or at least did not notice them as a task resumption cue. Furthermore, each one person did not notice the image and *WordCloud Cue*, and two people failed to notice the *Interactive Test Cue* (or notice them as task resumption feature, respectively).

#### 7.4.1. SUS Questionnaire

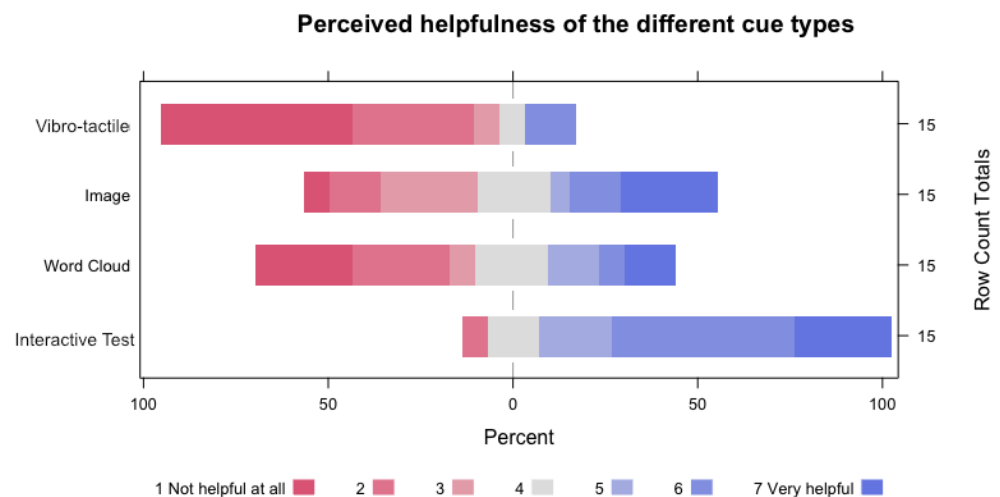
Before reporting the experiences with the task resumption cues, participants completed an adapted version of the System Usability Scale (SUS) questionnaire introduced by Brooke [54]. By using the SUS we aimed to gather data about usability issues with our application in general that might have biased the participants' rating of the task resumption cue feature. The SUS scores ranged from 57.5 to 97.5, with a mean of 82.21 ( $SD = 12.21$ ). Since 70 marks the threshold for acceptable usability, our application's rating of 82 can be considered excellent (cf. [55]) and we do not expect any influences of the usability on participants' cue assessment.

Overall, participants ranked the helpfulness of the *Image Cue* and *Interactive Test Cue* above average ( $M_{\text{image}} = 4.35$ ,  $SD_{\text{image}} = 1.94$ ;  $M_{\text{test}} = 5.59$ ,  $SD_{\text{test}} = 1.29$ , 7-point Likert scale from 1 = *not at all* to 7 = *very helpful*, cf. Figure 10), while the *Vibro-Tactile Cue* and *WordCloud Cue* received lower ratings ( $M_{\text{vibro}} = 2.18$ ,  $SD_{\text{vibro}} = 1.62$ ;  $M_{\text{cloud}} = 3.35$ ,  $SD_{\text{cloud}} = 2.03$ ).

For the in-the-wild usage, we further inquired if the resumption cues bothered or annoyed the participants while using the learning application. No cue was perceived as annoying (7-point Likert scale from 1 = *not at all annoying* to 7 = *very annoying*), with the *Image Cue* being the least annoying ( $M = 1.65$ ,  $SD = 1.13$ ), followed by the *Interactive Test Cue* ( $M = 2.24$ ,  $SD = 1.55$ ), and *WordCloud Cue* ( $M = 2.65$ ,  $SD = 1.75$ ). The *Vibro-Tactile Cue* got the highest rating of all cues, however, on average still remained below a neutral rating ( $M = 3.24$ ,  $SD = 2.36$ ).

We further asked participants if they would have preferred the presentation of the cues at a different point in time as well as to state feedback for the different cues individually. Participants often did not recognize the vibration as a task resumption cue, thus considering the *Vibro-Tactile Cue* useless. Four participants reported they could imagine it being an indicator for correct or incorrect answers and one person mentioned it could help to use it in the middle of a session to keep the user attentive. Two participants stated here that they'd rather prefer to not use the cue at all. As suggestions for improvement they noted to increase the frequency and improve the timing. In contrast, P13 praised the *Vibro-Tactile Cue* for being least intrusive, stating it "[...] *made the lesson stand out in a subtle way*".





**Figure 10.** Perceived helpfulness of the different cue types on a scale from 1 = *not at all* to 7 = *very helpful* for the revised cue designs .

None of the participants had any suggestions for improvement for the timing of the *Image Cue*. It was perceived as a helpful reminder of the last learned lesson (P4, P6, P7, P9, P11, P13, P17). Participants suggested to improve the choice of icons (P1) or add a keyword (P3), especially if the last lesson is further in the past and the memory fades.

For the *WordCloud Cue* the opinions were very diverse. Two participants suggested to remove it all-together, one person mentioned they wanted to see such cues more frequently throughout the app, while one user perceived the timing of this cue as “*arbitrary*”. Three participants considered the cloud as a nice overview (P3, P10, P11) and also helpful to track the individual progress (P12). Referring to words that were shown less frequently, P5 mentioned that “[...] *their smaller size caught my attention and I think it helped me to memorize them better.*” In contrast, P13 perceived this cue as “*demotivating*”, as small words reflect on low performance.

The *Interactive Test Cue* received the most positive ratings and feedback of all cues. Participants described the cue in a very positive way, with all but two considering it useful and helpful (P7 & P17 who did not perceive the cue), while P3 considered it a “[...] *a little long*”, P8 reported that it was badly timed when the interruption occurred at the beginning of a learning session. As the cue asked for translations of words of the specific lesson, it would then include words that are unknown to the user at this point of time. Thus, users suggested to use this cue rather as a quick repetition at the end of the lesson, even without an interruption happening.

As overall conclusion, the participants perceived the task resumption cues in general as helpful to recover from interruptions during mobile learning ( $M = 4.06$ ,  $SD = 1.21$ ; 7-point Likert scale from 1 = *not at all* to 7 = *very helpful*) and would like to see such cues implemented in mobile learning apps (15 participants in favor, two oppose). Two participants further stated that such task resumption cues as evaluated in this app could also be beneficiary for reading longer digital texts (e.g., online articles or e-books) (P1, P7).

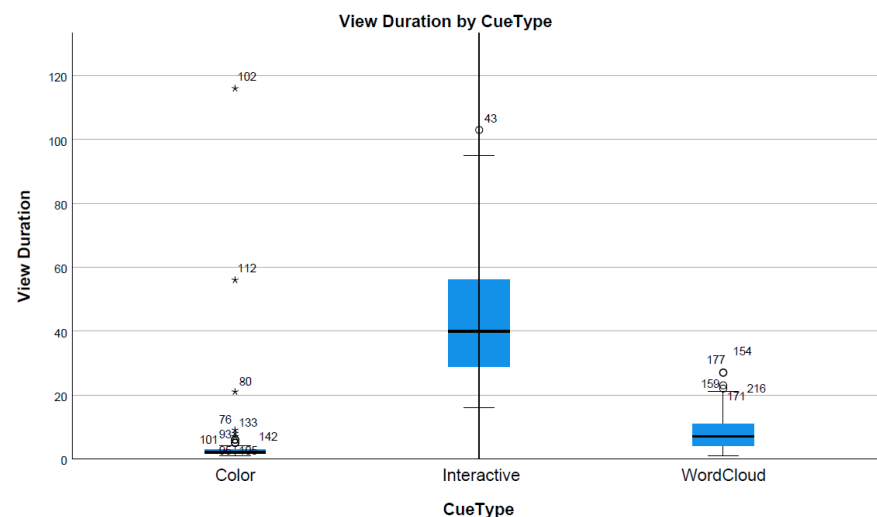
#### 7.4.2. Sessions and Interruptions

The usage of the app, called *CzechWizard*, varied among our participants. For the number of tasks answered we observed a range between 169 and 821 tasks during the course of this study ( $M = 519.44$ ,  $SD = 236.85$ ). The majority of interruptions were detected as result of a screen-off event. The Experience Sampling Questionnaire meant to assess the learning context was majorly dismissed by the participants and only answered in 163 cases. Of these 163, 87 represent learning at home, 36 on public transportation, 22 on the go, 15 at work, and three at the university.

### 7.4.3. Cue Viewing and Interaction

We cleaned our data set of aborted cues, learning sessions that did not exceed 10 s, and cue viewing duration that exceeded five minutes. In our final data set, we recorded 8276 solved learning tasks ( $N = 16$ ) and 209 detected interruptions. The four task resumption cue designs were counter-balanced, due to the removal of certain instances we included 44 *Vibro-Tactile Cue*, 46 *Color and Icon Cue*, 42 *WordCloud Cue*, and 36 *Interactive Test Cue*. In 41 cases, no cue was shown as a control condition.

Figure 11 depicts the viewing duration of the three cue types. As the duration of the *Vibro-Tactile Cue* was fixed in the app, it is not included here. The interaction time with the *Interactive Test Cue* that asked users to match six words and their translations exceeded those of the other cues. In particular, the interaction time with the *Interactive Test Cue* ( $M = 44.88$  s,  $SD = 19.37$ ) was on average five times as high as the viewing duration of the *WordCloud Cue* ( $M = 8.48$ ,  $SD = 6.51$ ) and even ten times as high as the viewing duration of the *Color and Icon Cue* ( $M = 4.71$ ,  $SD = 13.36$ ). Looking at the *Interactive Test Cue* in detail, users aborted or quit the tasks, either immediately or after answering a subset of the questions, in 32 out of 90 cases. Out of the completed set of 58 cues, 35 were answered entirely correct. Overall, the median error rate during the cue tasks remained zero ( $max = 6$ ). Note that all tasks had to be solved correctly in order to continue with the lesson.



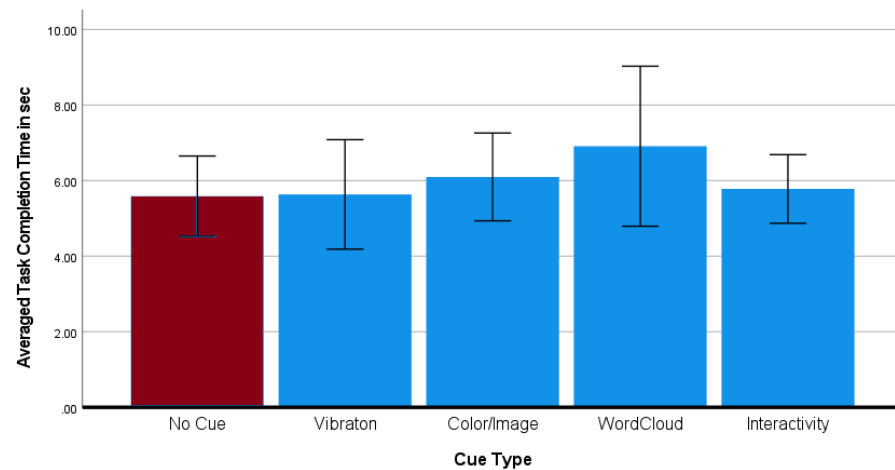
**Figure 11.** The viewing duration in seconds of the three cue types *Color and Icon Cue*, *Interactive Test Cue*, and *WordCloud Cue*.

### 7.4.4. Task Completion Time

In the lab-based user study of the first part of this chapter, we compared the task completion time before and after an interruption, i.e., before and after a resumption cue was presented. In this field study, the majority of interruptions led to the termination of the current learning session. As a result, the task resumption cue was often shown immediately at the beginning of the next learning session. Therefore, we will not compare the tasks before and after the interruption, but only take the first tasks after each resumption cue is shown into the consideration for the analysis ( $N = 1004$ ). Figure 11 visualizes the descriptive statistics of the task completion times. As depicted in Figure 12, the highest average task completion time can be observed after the presentation of the *WordCloud Cue* ( $M = 6.91$  s,  $SD = 3.97$  s), the lowest in the no-cue control condition ( $M = 5.59$  s,  $SD = 1.99$  s).

Due to the removal of outlier values and the cleaning of the data set, we had to perform a computational imputation for missing values. The missing values were replaced by averages across participants. A repeated-measures ANOVA across participants showed no significant difference ( $p > 0.05$ ) among the four different cue types and the no-cue

condition. In other words, users needed on average around the same amount of time to solve a learning task after an interruption no matter if a task resumption cue was shown or not.



**Figure 12.** The task completion time for the first five tasks after an interruption in seconds. The completion times are averaged per participant and visualized to compare the four cue designs and the no cue condition.

#### 7.4.5. Error Rate

The learning content in the application was at a beginners' level and the number of errors made by our participants turned out to be fairly low. Out of 8276 recorded learning tasks, only 110 were incorrectly answered, resulting in an error rate of less than 1%. When we limit the responses to the first five tasks after a cue was presented ( $N = 1004$ ), only 36 were incorrect. We consider this sample insufficient for statistical comparison among five test conditions and will therefore refrain from drawing conclusions on the effectiveness of task resumption cues designs on correctness. We will discuss the feasibility of the applied metrics for measuring the effectiveness of task resumption cues in the limitation section.

## 8. Discussion

### 8.1. Limitations

The evaluation presented above revealed several limitations in our application and also study design.

Firstly, since our user study was performed during the COVID-19 pandemic, our evaluation might include anomalies in the daily routines and smartphone usage of our participants. Lockdown and work-from-home phases could have potentially influenced the use of mobile language learning applications, the occurring interruptions, and users' ability to recall prior learning sessions.

Secondly, the motivation to engage in the learning sessions varied greatly among our participants. We encouraged all study participants to interact with the app frequently but also in the same way they would with any other learning app. We observed similar differences as already reported in Draxler et al. [15]. Some participants learned multiple times per day while others only had very few learning sessions. With this, the number of interruptions and session breaks also varied (min 3, max 29). Without a sufficient number of interruptions, the generalizability of our results in regard to the helpfulness of the task resumption cues is limited.

Thirdly, due to the lack of complexity in the learning content and the users' good overall performance, the explanatory power of our metrics task completion time and error rate about the effectiveness of the task resumption cues is limited. For future work, we recommend to evaluate memory cues in a setting with more complex content.

In addition, we observed several methodological constraints during our user study. One critical issue is to define the difference between an interruption and a new learning session. In the laboratory settings, interruptions were fixed in terms of severity and duration. In users' everyday environment, the only indicator for interruption severity we can gather is the duration between two learning sessions. For the current analysis, we considered every break between learning sessions, be it five minutes or five days, an interruption. However, the memory decay after five days will be significantly worse than after five minutes [7]. Thus, adapting the presentation of task resumption cues and their level of detail in regard to the severity of the interruption might be good to consider for future work.

### 8.2. Divided Opinion on Cue Designs

While the majority of participants agreed on the helpfulness of the *Interactive Test Cue*, their opinions were divided on the other cue designs. In particular, both the *WordCloud Cue* and *Color and Icon Cue* received very positive and very poor feedback. Directly comparing the helpfulness ratings of the first versions of the cues, we found that the perceived helpfulness of the *Color and Icon Cue* increased from a mean value of 2.07 to 4.35. Thus, its seems relatively subtle changes such as the additional coloring in the *Color and Icon Cue* can substantially reduce confusion. On the other hand, the helpfulness of the revised version of the *WordCloud Cue*, which now additionally visualized learning progress, was not rated higher than the original version in the lab study.

### 8.3. Problems and Opportunities for Using Tactile Memory Cues

Since the *Vibro-Tactile Cue* was very subtle and implicit, it was often not perceived by our participants. Some expected that the vibration was a feedback mechanism that implies correct or incorrect answers. Although this was not the case, this idea presents the intriguing opportunity to cue content specifically. For example, a vibration cue during the corrective feedback presentation of a word that has been incorrectly answered multiple times in a row could grab the users' attention. By replaying the *Vibro-Tactile Cue* the next time the same question is posed, we can guide the users' attention and increase their caution and focus when answering this specific task. Thus, we can ensure deeper processing and potentially foster more elaborate rehearsal, leading to better encoding into their LTM.

### 8.4. Task Resumption beyond Micro-Learning

According to our quantitative metrics, we could not observe any effect of the task resumption cues on the learning performance. The course was generally designed to include simple content aimed at teaching beginners' level Czech to avoid differences among participants due to existing language proficiency. The very low overall error rate of less than 1% indicated that users had no difficulties remembering the app's vocabulary and answering the tasks. We suspect that while most participants appreciated the task resumption cues, in particular, the *Color and Icon Cue* and *Interactive Test Cue*, the helpfulness of the cues would be greater for content of greater complexity. The beginners' level Czech course we designed for our experiment had only a few content units building upon each other. Recalling prior content was, therefore, not necessarily required for progressing in the application.

## 9. Implications for Task Resumption Cues in Mobile Learning

The user preferences elicited in the two studies we conducted give important insights into the user acceptance of task resumption cues. Thus, the feedback received can facilitate the future design of cues for mobile learning and other productive activities on mobile devices. In this context, the lab and field studies serve as components of a user-centered iterative design process. By detailing initial considerations and subsequent modifications, we aim to support developers and researchers in their decision-making process. Below, we discuss the evaluated cues, possible modifications, and caveats.

**Explicit cues that clearly refer back to content viewed before an interruption received the most favorable ratings in both studies.** Cues with a history view were previously found to be useful in desktop scenarios such as programming tasks [28]. It is interesting that this strategy also seems to be desirable in mobile learning because the micro-learning exercises typically employed in this context usually require only little contextual information. As mentioned above, the helpfulness of a *History Cue* would probably be greater for more complex content. Nevertheless, and as long as there is no negative effect on task performance, the perceived helpfulness could be taken as reason to include a *History Cue* even when there is no particular need for conveying context. Interactive elements, on the other hand, need to be carefully considered. In our case, we observed a median viewing duration of 40 s for the *Interactive Test Cue*, i.e., a significant delay. However, the *Interactive Test Cue* is aiming to optimize an already existing feature of learning applications—the repetition section—to precisely support task resumption after interruptions. Despite the long duration, it was very positively received by our participants.

**Less structured explicit task resumption cues, in our case, the two variations of *WordCloud Cues*, showed that the perceived helpfulness is limited and that the provided information needs to be well-adjusted to a given context** (e.g., by emphasizing words that were answered incorrectly before). Similarly, the even more implicit color and image cues were rated as very helpful by some users and not at all by others. Again, we assume that the explanatory power in an image will make a difference here. For example, in the work of Liu et al. [56], images of a flower decreased the time spent on secondary tasks; only a short glance was sufficient to participants in their study to become aware of the suspended primary task, although it was not used to reconstruct the full task context. In our studies, the viewing times of *Image Cue* and *Color and Icon Cue* were short, which means that implementing this type of cue would probably not have a detrimental effect on task performance, even if positive effects on tasks are not guaranteed.

Inspired by research conducted in stationary settings (e.g., [33,35]), our first implementation also included a *Half-Screen Cue*. **While *Half-Screen Cue* showed promising effects on task resumption in prior work, the implementation requires redesign before providing any benefit in a mobile learning scenario in the wild.** An important aspect to consider in mobile settings is the fact that several applications concurrently use the small screen space available. Current multi-tasking options for mobile devices exist, but are typically limited to screen-in-screen designs such as music player views hovering on top of an active navigation app. The study by Iqbal and Horvitz [44] showed that when the primary task window was less than 25% visible it took participants significantly longer to return to it as when it was more than 75% visible. However, this work concerned desktop settings and requires future work to assess if smaller ratios (such as leaving the learning app visible like a music player or navigation app), would actually improve task resumption. Overall, the applicability of this approach remains limited to interruption scenarios where the user focus stays on the device—and for many real-life interruptions, this is not the case.

Finally, **the *Vibro-Tactile Cue* provides the most implicit user guidance of the designs we investigated and showed its potential for subtle highlighting by drawing attention to certain lessons**, while related work applied it primarily to draw attention back to the primary task (cf. [47]), we tested the use of simple patterns to trigger the recall of prior memories. In contrast to the other cue designs, this implicit cue did not reflect properties of the learning content, such as grammatical concepts or prior mistakes made for similar questions. Future research could look further into the effect of varying patterns and which information or actions it would make sense to relate them to. Past work has shown that humans can distinguish various vibration patterns, and, for instance, use them to assess urgency [57].

In sum, even though the overall sample size in the lab and field studies was relatively small, we could observe clear preferences for some cue designs and individual preferences for others. **While our quantitative evaluations show no significant decrease in task completion times and error rates due to the cues after an interruption, the partici-**

pant feedback gives clear indications which directions for cue designs are promising in a mobile (learning) context (e.g., cues visualizing activity history), where further modifications are needed (e.g., *Vibro-Tactile Cue*), and which cue designs will likely appeal to a small number of users or use cases only (e.g., *Half-Screen Cue*). We found that some cue designs can be applied in a similar fashion as in stationary settings (e.g., a *History Cue*), while others were previously tested only in specialized settings such as aviation control but are also a good match for learning on mobile devices (e.g., a *Vibro-Tactile Cue*), and some cues will be difficult to re-design for mobile contexts (e.g., a *Half-Screen Cue*).

In general, as stated in the limitation section of the second study, the severity of the interruptions and the complexity of the content influences the need for memory support. In short, easy, or frequent learning sessions users could be supported by implicit tactile or visual cues. At the same time, they would potentially benefit more from explicit and complex cues in long, difficult, or irregular learning sessions. Similar variations in context dependence also occur in other tasks performed on mobile devices. For example, a if someone is interrupted while responding to a chat message about dinner, it will probably be easy to recover the conversation flow without rich cues. This will probably be different for someone returning to the current paragraph in a technical report. which they read for work purposes. These dependencies should be further examined in future work. Finally, giving users the option to adapt the granularity or content of the cues on demand could help increase their acceptance. Through personalization such as bookmarking relevant content before an interruption, the relevance of the later cue can be increased .

## 10. Conclusions and Future Work

This work presents the design of memory cues to support task resumption in mobile learning applications. We derive and iterate on four different cue designs and evaluate user experience in two user studies.

While the quantitative effects of the task resumption cues on learning performance could not be definitely confirmed in our studies, the subjective user feedback revealed great potential of the cues. From the need to break the content down into individual micro-learning units, task resumption cues, if designed properly, can create a seamless connection between individual learning units. Instead of multiple self-contained micro-learning sessions, learners could engage in more complex topics that require longer time and deeper engagement. In particular, interactive cue designs appear promising even beyond the current application of language learning.

We emphasize that research on task resumption cues in mobile and uncontrolled environments is still sparse, and the effectiveness of memory cues might be strongly influenced by the specific interruption and situation of the user. For future work, we propose to evaluate promising cue designs in learning applications containing more complex and coherent learning topics, such as STEM. Further, the concept of memory cues could be transferred to other cognitively demanding tasks that are at risk of interruptions, such as mobile text reading, listening to audiobooks or podcasts, or writing text.

**Author Contributions:** Conceptualization, C.S. and F.D.; methodology, C.S., F.D.; software, V.F. and V.K.; formal analysis, C.S. and F.D.; data curation, V.F. and V.K.; writing—original draft preparation, C.S.; writing—review and editing, C.S. and F.D.; visualization, F.D., V.K. and V.F.; supervision, C.S. and F.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Horizon 2020 grant number 683008 and by Deutsche Forschungsgemeinschaft grant number 251654672.

**Institutional Review Board Statement:** Ethical review and approval were waived for this study. All participants took part in the study voluntarily with the option to withdraw at any point without negative consequences. They received all relevant information about the study and its purpose and all their data was processed anonymously.

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

**Data Availability Statement:** Not applicable.

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

## References

1. Schneegass, C.; Terzimehić, N.; Nettah, M.; Schneegass, S. Informing the design of user-adaptive mobile language learning applications. In Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia, Cairo, Egypt, 25–28 November 2018; pp. 233–238.
2. Demouy, V.; Jones, A.; Kan, Q.; Kukulska-Hulme, A.; Eardley, A. Why and How Do Distance Learners Use Mobile Devices for Language Learning? *EuroCALL Rev.* **2016**, *24*, 10–24. [[CrossRef](#)]
3. Cull, W.L. Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Appl. Cogn. Psychol.* **2000**, *14*, 215–235. [[CrossRef](#)]
4. Bailey, B.P.; Konstan, J.A. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Comput. Hum. Behav.* **2006**, *22*, 685–708. [[CrossRef](#)]
5. Katidioti, I.; Borst, J.P.; van Vugt, M.K.; Taatgen, N.A. Interrupt me: External interruptions are less disruptive than self-interruptions. *Comput. Hum. Behav.* **2016**, *63*, 906–915. [[CrossRef](#)]
6. Kreifeldt, J.G.; McCarthy, M. Interruption as a test of the user-computer interface. In Proceedings of the 17th Annual Conference on Manual Control, Los Angeles, CA, USA, 16–17 June 1981.
7. Schacter, D.L. The seven sins of memory: Insights from psychology and cognitive neuroscience. *Am. Psychol.* **1999**, *54*, 182. [[CrossRef](#)]
8. Bruck, P.A.; Motiwalla, L.; Foerster, F. Mobile Learning with Micro-content: A Framework and Evaluation. In Proceedings of the Bled eConference, Bled, Slovenia, 17–20 June 2012; p. 2.
9. Dinger, T.; Weber, D.; Pielot, M.; Cooper, J.; Chang, C.C.; Henze, N. 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, Vienna, Austria, 4–7 September 2017; p. 28.
10. Page, T. Usability of text input interfaces in smartphones. *J. Des. Res.* **2013**, *11*, 39–56. [[CrossRef](#)]
11. Heil, C.R.; Wu, J.S.; Lee, J.J.; Schmidt, T. A Review of Mobile Language Learning Applications: Trends, Challenges, and Opportunities. *EuroCALL Rev.* **2016**, *24*, 32–50. [[CrossRef](#)]
12. Churchill, D.; Hedberg, J. Learning object design considerations for small-screen handheld devices. *Comput. Educ.* **2008**, *50*, 881–893. [[CrossRef](#)]
13. Trafton, J.G.; Altmann, E.M.; Brock, D.P.; Mintz, F.E. Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *Int. J. Hum.-Comput. Stud.* **2003**, *58*, 583–603. [[CrossRef](#)]
14. Schneegass, C.; Draxler, F. Designing Task Resumption Cues for Interruptions in Mobile Learning Scenarios. In *Technology-Augmented Perception and Cognition*; Dinger, T., Niforatos, E., Eds.; Springer International: Berlin/Heidelberg, Germany, 2021.
15. Draxler, F.; Schneegass, C.; Safranek, J.; Hußmann, H. Why did you stop?—Investigating Origins and Effects of Interruptions during Mobile Language Learning. In Proceedings of the Conference Mensch und Computer, Ingolstadt, Germany, 5–8 September 2021; Gesellschaft für Informatik e.V.: Bonn, Germany, 2021.
16. Cohen, S. Aftereffects of stress on human performance and social behavior: A review of research and theory. *Psychol. Bull.* **1980**, *88*, 82. [[CrossRef](#)]
17. Monk, C.A.; Trafton, J.G.; Boehm-Davis, D.A. The effect of interruption duration and demand on resuming suspended goals. *J. Exp. Psychol. Appl.* **2008**, *14*, 299. [[CrossRef](#)] [[PubMed](#)]
18. Glickman, S.E. Perseverative neural processes and consolidation of the memory trace. *Psychol. Bull.* **1961**, *58*, 218. [[CrossRef](#)] [[PubMed](#)]
19. McGaugh, J.L. Time-dependent processes in memory storage. *Science* **1966**, *153*, 1351–1358. [[CrossRef](#)] [[PubMed](#)]
20. Altmann, E.M.; Trafton, J.G. Memory for goals: An activation-based model. *Cogn. Sci.* **2002**, *26*, 39–83. s15516709cog2601\_2. [[CrossRef](#)]
21. Krause, U.M.; Stark, R. Vorwissen aktivieren. *Handb. Lernstrateg.* **2006**, *1*, 38–49.
22. McDaniel, M.A.; Einstein, G.O.; Graham, T.; Rall, E. Delaying execution of intentions: Overcoming the costs of interruptions. *Appl. Cogn. Psychol. Off. J. Soc. Appl. Res. Mem. Cogn.* **2004**, *18*, 533–547. [[CrossRef](#)]
23. Sasangohar, F.; Scott, S.D.; Cummings, M. Supervisory-level interruption recovery in time-critical control tasks. *Appl. Ergon.* **2014**, *45*, 1148–1156. [[CrossRef](#)] [[PubMed](#)]
24. Cheng, S.; Fan, J.; Dey, A.K. Smooth gaze: A framework for recovering tasks across devices using eye tracking. *Pers. Ubiquitous Comput.* **2018**, *22*, 489–501. [[CrossRef](#)]
25. Kern, D.; Marshall, P.; Schmidt, A. Gazemarks: Gaze-based visual placeholders to ease attention switching. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, GA, USA, 10–15 April 2010; pp. 2093–2102.
26. Mancero, G.; Wong, B.; Loomes, M. Radio dispatchers' interruption recovery strategies. In Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7, Melbourne, Australia, 23–27 November 2009; pp. 113–120.

27. Yeung, W.L.; Li, S.Y. Prototyping the Machine-Human Dialogues in a Smartphone Voice Call Application with Task Resumption Support. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 1788–1793.
28. Parnin, C.; DeLine, R. Evaluating cues for resuming interrupted programming tasks. In Proceedings of the 28th International Conference on Human Factors in Computing Systems—CHI'10, Atlanta, GA, USA, 10–15 April 2010; p. 93. [\[CrossRef\]](#)
29. Laufer, B.; Goldstein, Z. Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Lang. Learn.* **2004**, *54*, 399–436. [\[CrossRef\]](#)
30. Czerwinski, M.; Chrisman, S.; Schumacher, B. Interactive Posters: The effects of warnings and display similarity on interruption in multitasking environments. *SIGCHI Bull.* **1991**, *23*, 38–39. [\[CrossRef\]](#)
31. Czerwinski, M.; Cutrell, E.; Horvitz, E. Instant messaging: Effects of relevance and timing. In *People and Computers XIV: Proceedings of HCI*; Springer: Berlin/Heidelberg, Germany, 2000; Volume 2, pp. 71–76.
32. Oulasvirta, A.; Saariluoma, P. Surviving task interruptions: Investigating the implications of long-term working memory theory. *Int. J. Hum.-Comput. Stud.* **2006**, *64*, 941–961. [\[CrossRef\]](#)
33. Ratwani, R.M.; Andrews, A.E.; McCurry, M.; Trafton, J.G.; Peterson, M.S. Using Peripheral Processing and Spatial Memory to Facilitate Task Resumption. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2007**, *51*, 244–248. [\[CrossRef\]](#)
34. Iqbal, S.T.; Horvitz, E. Conversations Amidst Computing: A Study of Interruptions and Recovery of Task Activity. In *User Modeling 2007*; Conati, C., McCoy, K., Paliouras, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; Volume 4511, pp. 350–354. [\[CrossRef\]](#)
35. Hodgetts, H.M.; Jones, D.M. Contextual cues aid recovery from interruption: The role of associative activation. *J. Exp. Psychol. Learn. Mem. Cogn.* **2006**, *32*, 1120–1132. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Draxler, F.; Schneegass, C.; Niforatos, E. Designing for Task Resumption Support in Mobile Learning. In Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services—MobileHCI'19, Taipei, Taiwan, 1–4 October 2019; Association for Computing Machinery: New York, NY, USA, 2019. [\[CrossRef\]](#)
37. Chen, N.S.; Hsieh, S.W. Effects of short-term memory and content representation type on mobile language learning. *Lang. Learn. Technol.* **2008**, *12*, 93–113.
38. Scott, S.D.; Mercier, S.; Cummings, M.L.; Wang, E. Assisting interruption recovery in supervisory control of multiple UAVs. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, San Francisco, CA, USA, 16–20 October 2006; SAGE Publications Sage CA: Los Angeles, CA, USA, 2006; Volume 50, pp. 699–703. [\[CrossRef\]](#)
39. John, M.; Smallman, H.S.; Manes, D.I. Recovery from Interruptions to a Dynamic Monitoring Task: The Beguiling Utility of Instant Replay. *Proc. Hum. Factors Ergon. Soc. Annu.* **2005**, *49*, 473–477. [\[CrossRef\]](#)
40. Heimerl, F.; Lohmann, S.; Lange, S.; Ertl, T. Word Cloud Explorer: Text Analytics Based on Word Clouds. In Proceedings of the 2014 47th Hawaii International Conference on System Sciences, Waikoloa, HI, USA, 6–9 January 2014; pp. 1833–1842. [\[CrossRef\]](#)
41. Kuo, B.Y.L.; Hentrich, T.; Good, B.M.; Wilkinson, M.D. Tag clouds for summarizing web search results. In Proceedings of the 16th international conference on World Wide Web—WWW'07, Banff, AB, Canada, 8–12 May 2007; p. 1203. [\[CrossRef\]](#)
42. Borojeni, S.S.; Ali, A.E.; Heuten, W.; Boll, S. Peripheral Light Cues for In-Vehicle Task Resumption. In Proceedings of the 9th Nordic Conference on Human-Computer Interaction—NordiCHI'16, Gothenburg, Sweden, 23–27 October 2016. [\[CrossRef\]](#)
43. Clifford, J.D.; Altmann, E.M. Managing multiple tasks: Reducing the resumption time of the primary task. In Proceedings of the Annual Meeting of the Cognitive Science Society, Chicago, IL, USA, 4–7 August 2004; Volume 26.
44. Iqbal, S.T.; Horvitz, E. Disruption and Recovery of Computing Tasks: Field Study, Analysis, and Directions. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems—CHI'07, San Jose, CA, USA, 28 April–3 May 2007; p. 677–686. [\[CrossRef\]](#)
45. Jo, J.; Kim, B.; Seo, J. EyeBookmark: Assisting recovery from interruption during reading. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Korea, 18–23 April 2015; pp. 2963–2966.
46. Toreini, P.; Langner, M.; Maedche, A. Use of Attentive Information Dashboards to Support Task Resumption in Working Environments. In Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications, Warsaw, Poland, 14–17 June 2018. [\[CrossRef\]](#)
47. Smith, C.; Clegg, B.A.; Heggstad, E.D.; Hopp-Levine, P.J. Interruption management: A comparison of auditory and tactile cues for both alerting and orienting. *Int. J. Hum.-Comput. Stud.* **2009**, *67*, 777–786. [\[CrossRef\]](#)
48. Hopp, P.J.; Smith, C.; Clegg, B.A.; Heggstad, E.D. Interruption management: The use of attention-directing tactile cues. *Hum. Factors* **2005**, *47*, 1–11. [\[CrossRef\]](#)
49. Hopp-Levine, P.J.; Smith, C.; Clegg, B.A.; Heggstad, E.D. Tactile interruption management: Tactile cues as task-switching reminders. *Cogn. Technol. Work* **2006**, *8*, 137–145. [\[CrossRef\]](#)
50. Kuznetsov, S.; Dey, A.K.; Hudson, S.E. The effectiveness of haptic cues as an assistive technology for human memory. In Proceedings of the International Conference on Pervasive Computing, Nara, Japan, 11–14 May 2009; pp. 168–175.
51. Pielot, M.; Oliveira, R.D. Peripheral vibro-tactile displays. In Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services, Munich, Germany, 27–30 August 2013; pp. 1–10.
52. Yatid, M.; Takatsuka, M. Understanding the effectiveness of visual cues to support categorical notification. In Proceedings of the 24th Australian Computer-Human Interaction Conference, Melbourne, Australia, 26–30 November 2012; pp. 661–664.



- 
53. Ardissono, L.; Bosio, G.; Segnan, M. A visualization model supporting an efficient context resumption in collaboration environments. In Proceedings of the International Conference on User Modeling, Adaptation, and Personalization, Girona, Spain, 11–15 July 2011; pp. 5–17.
  54. Brooke, J. SUS: A quick and dirty usability. *Usability Eval. Ind.* **1996**, *189*, 4–7.
  55. Bangor, A.; Kortum, P.; Miller, J. Determining what individual SUS scores mean: Adding an adjective rating scale. *J. Usability Stud.* **2009**, *4*, 114–123.
  56. Liu, Y.; Jia, Y.; Pan, W.; Pfaff, M.S. Supporting task resumption using visual feedback. In Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing—CSCW’14, Baltimore, MD, USA, 15–19 February 2014; pp. 767–777. [[CrossRef](#)]
  57. Saket, B.; Prasojo, C.; Huang, Y.; Zhao, S. Designing an effective vibration-based notification interface for mobile phones. In Proceedings of the 2013 Conference on Computer Supported Cooperative Work—CSCW’13, San Antonio, TX, USA, 23–27 February 2013; p. 149. [[CrossRef](#)]