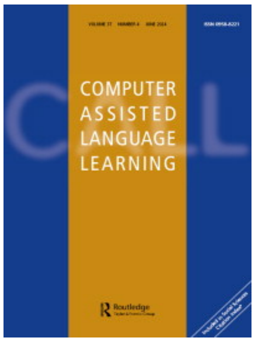


Routledge
Taylor & Francis Group

Computer Assisted Language Learning

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ncal20

# Training finnish morphology with a smartphone application in adult beginner level learners

**Rosa Salmela, Minna Lehtonen, Johanna Köykkä, Stefano Garusi, Matti Laine & Raymond Bertram**

**To cite this article:** Rosa Salmela, Minna Lehtonen, Johanna Köykkä, Stefano Garusi, Matti Laine & Raymond Bertram (01 Jul 2024): Training finnish morphology with a smartphone application in adult beginner level learners, Computer Assisted Language Learning, DOI: 10.1080/09588221.2024.2365664

**To link to this article:** https://doi.org/10.1080/09588221.2024.2365664

COMPUTER ASSISTED LANGUAGE LEARNING
https://doi.org/10.1080/09588221.2024.2365664



# Training finnish morphology with a smartphone application in adult beginner level learners

Rosa Salmela[a,b], Minna Lehtonen[b,c], Johanna Köykkä[d], Stefano Garusi[a], Matti Laine[a] and Raymond Bertram[b]

[a]Department of Psychology, Åbo Akademi University, Turku, Finland; [b]Department of Psychology and Speech-Language Pathology, University of Turku, Turku, Finland; [c]MultiLing Center for Multilingualism in Society across the Lifespan, University of Oslo, Oslo, Norway; [d]Department of Speech and Language Pathology, Åbo Akademi University, Turku, Finland

**ABSTRACT**

The rich inflectional grammar of Finnish is often considered a challenge for second language (L2) learners. However, the inflectional variation is based on relatively systematic morphological and morphophonological rules. Smartphone applications could offer an effective way to train these rules by utilizing adaptive spaced repetition exercises. Currently there is little research on morphology learning *via* Mobile Assisted Language Learning (MALL) applications in morphologically rich languages. The present study introduces a newly developed Finnish Morphology Application (FMA) that aims to teach beginner-level Finnish morphology. In a randomized controlled trial, the test group participants ($n = 33$) practiced Finnish morphology with the FMA during a three-week period, and their improvements were compared to those of an active control group using a general Finnish language learning application ($n = 38$). The practiced morphological operations comprised transparent and semi-transparent inflections. Pre- and posttests included a visual lexical decision task and a morphological production task. In comparison to the control group, the FMA group improved substantially in written morphological production skills, especially in transparent inflections. In visual lexical decision task, there was an improvement in accuracy for trained words in the FMA group, but this improvement applied to both transparent and semi-transparent inflections as well as monomorphemic nouns. In reaction times, there were no differences between the groups. Self-ratings of perceived efficacy of the FMA regarding grammar and vocabulary varied from neutral to positive. The results suggest that the FMA is particularly useful for learning inflectional production skills in Finnish by adult beginner level learners but can also be used to train word recognition.

**ARTICLE HISTORY**
Received 30 May 2023
Accepted 2 June 2024

**KEYWORDS**
Mobile assisted language learning; Finnish morphology; second language; randomized controlled trial; exemplar-based learning

**CONTACT** Rosa Salmela rosa.salmela@abo.fi Department of Psychology, Åbo Akademi University, Finland

Supplemental data for this article can be accessed online at https://doi.org/10.1080/09588221.2024.2365664.

## Introduction

Mobile Assisted Language Learning (MALL) refers to educational usage of mobile technologies (i.e. smartphones and tablets) in language instruction. In the past 15 years, MALL efficacy has been extensively studied. While the results are promising, most studies have focused either on vocabulary learning or on reading and writing, and have been primarily conducted in English as a second language (Chen et al., 2020; Lin & Lin, 2019; Sung et al., 2015). Studies concerning grammar learning are fewer, especially in morphologically rich languages. The Finnish language is characterized by vast inflectional grammar, which is often considered a challenge for second language (L2) learners. In the current study, we investigated whether using a prototype of a newly developed Finnish Morphology Application (hereafter referred to as FMA) improved the acquisition of inflectional morphology in Finnish L2 learners.

### *Finnish morphology*

Finnish is considered a morphologically rich language, as most grammatical relations are expressed *via* adding suffixes to a word stem (e.g. *talo: talo + ssa* 'house: in a house'). One of the greatest challenges in Finnish is related to the correct use of morphosyntax, i.e. which grammatical cases can be used in specific syntactic positions or roles (Martin, 1995). This morphosyntactic challenge is also present for English (see Goldschneider & DeKeyser, 2001, for a review on grammatical morpheme acquisition of L2 English speakers; Hopp, 2020; Jensen et al., 2020), but is amplified in Finnish by its complex case system. To illustrate, whereas English expresses locational relations through prepositions, Finnish uses inflectional case endings for the same purpose (see Table 1). For instance, the Finnish translation of the English sentence 'the cat goes on**to** the table' is 'kissa[(THE CAT)] menee[(GOES)] pöydä+**lle**[(ONTO THE TABLE)]', where the verb 'go' requires the allative case ending for the noun 'table'. Similarly, a sentence like 'the cat comes **from** the table' translates into 'kissa[(THE CAT)] tulee[(COMES)]

**Table 1.** Locative cases in Finnish and examples of transparent and semi-transparent inflections.

| Domain | Case | Transparent inflection (no CG) | Semi-transparent inflection (quantitative CG) | Semi-transparent inflection (qualitative CG) | Grammatical Function |
|---|---|---|---|---|---|
| | Nominative | sohva, 'sofa' | laati**kk**o, 'box' | pa**t**a, 'casserole' | |
| Internal | Inessive | sohva + ssa | laati**k**o + ssa | pa**d**a + ssa | 'in' |
| | Elative | sohva + sta | laati**k**o + sta | pa**d**a + sta | 'out of' |
| | Illative | sohva + an | laati**kk**o + on | pa**t**a + an | 'into' |
| External | Adessive | sohva + lla | laati**k**o + lla | pa**d**a + lla | 'on' |
| | Ablative | sohva + lta | laati**k**o + lta | pa**d**a + lta | 'off' |
| | Allative | sohva + lle | laati**k**o + lle | pa**d**a + lle | 'onto' |

*Notes.* CG = consonant gradation.

pöydä+**ltä**(FROM THE TABLE)', with the verb 'come' requiring the ablative case for the noun 'table'. In Finnish second language education, inflectional morphology is considered as part of grammar, as all inflected forms carry grammatical information (Karlsson, 2002). In the current study, we focus on locative cases, which are used to express a location of an object or direction of an action. They can be divided into internal and external domains, depending on where the object is or action takes place. The general rule is that when an object is inside or an action takes place inside, e.g. a building, the internal locatives are used (e.g. *in a house, out of a house, into a house*). When the object is or action takes place in outer spaces, e.g. in the open air, or on top of something, the external locatives are used (e.g. *on a table, off a table, onto a table*).

Another focus of the current study and challenge for Finnish L2 speakers is related to morphophonology. Morphophonology is a term that refers to phonological alterations in morphemes when they are combined with other morphemes, resulting in allomorphs. Allomorphs of a morpheme share the same grammatical meaning but have different orthograpic and phonological representations. Consonant gradation (CG) is a prevalent morphophonological feature in Finnish that may reduce the transparency of the bound stem. CG changes the stem phonemes/ *k, p, t*/ in the final syllable of a word, when they are preceded by a vowel or a voiced consonant/*l, m, n, r, h*/. The changes can be quantitative or qualitative (Table 1). CG is a relatively common phenomenon in the Finnish language: of the 1000 most common words in the frequency vocabulary, almost one-third are subject to CG (VISK, 2004: § 41). For the sake of clarity, the CG-inflections are subsequently referred to as semi-transparent inflections, as opposed to transparent inflections, where the bound stem does not change.

While the Finnish language is characterized by complex inflectional grammar, it is worth noting that the variation is based on systematic morphological and morphophonological rules (Karlsson, 2002; Martin, 1995). However, the number of the grammatical rules is large, which makes the learning process often laborious. Therefore, the possibility to use MALL can make an important contribution to automatize morphological rules for an L2 speaker. Hence, we developed the FMA, a novel mobile application that aims to teach rules to form transparent and semi-transparent Finnish locative inflections and how to use them in different syntactic positions and roles.

### MALL effects in L2 grammar

Several studies have demonstrated the benefits of integrating MALL into classroom teaching for grammar learning. However, there is only one

MALL study of which we are aware that focuses directly on learning morphology in morphologically rich languages (Eryiğit et al., 2023). This study introduced a gamified application (İTÜRK; Eryigit, 2022) for learning complex morphology of Turkish. The instructed topics in the application were plurality, possession, noun cases and verb tenses. After a three-week training period, the application was qualified as being effective and positively perceived by the L2 students. The findings were based on self-reported data and the experimental design was non-controlled. The authors expressed the intention to conduct a larger study to evaluate the learning outcomes and compare the efficacy of the app to other existing language learning apps.

Other grammar studies in MALL have used languages characterized by relatively simple morphology, such as English and Spanish. These studies have consistently shown promising results. Wang et al. (2021) found that a personalized mobile-assisted learning environment significantly facilitated learning grammar in English as a foreign language for high school students in China. The instructed grammar rules included tense, mood, different types of clauses and direct/indirect statements. An experimental group and a passive control group both used the system for one semester, whereby the experimental group had full access to the learning environment and used it freely to learn grammar, while the control group could use it only to submit weekly grammar assignments. The results showed that at posttest, the experimental group obtained significantly higher English grammar test scores than the control group.

Similarly, Li and Hegelheimer (2013) designed a mobile application, Grammar Clinic, that supports learning grammar in English as a second language. Their training (16 wk) consisted of identifying and correcting various sentence-level errors in the texts. The participants took a pre- and posttest, which measured performance in similar tasks. The results showed that their posttest scores were significantly better than the pretest scores. Similar findings have been reported in other studies, both employing controlled designs (Chu et al., 2019; Ghorbani et al., 2020, for L2 English) and non-controlled designs (Castañeda & Cho, 2016, for L2 Spanish). Rachels and Rockinson-Szapkiw (2018) investigated the impact of Duolingo® on Spanish language achievement and academic self-efficacy of third and fourth grade students, the analysis revealed no significant difference between groups who used the application and who received only traditional instruction. The authors point out the possibility that the absence of random group allocation could be a contributing factor to the lack of observed differences.

In addition to MALL studies, there is a large body of studies in grammar learning *via* Computer Assisted Language Learning (CALL), which encompasses a wider range of computer-based technologies than MALL. Similar to MALL studies, CALL studies have reported positive outcomes for learning various aspects of grammar in English (AbuSeileek, 2009; Kılıçkaya, 2015; Kruk, 2015; Lucas, 2020; Naba'h et al., 2009; Reynolds & Kao, 2021, Torlaković, & Deugo, 2004), Spanish (Collentine & Collentine, 2015), and Dutch (Penning de Vries et al., 2020). The studies were mostly concerned with syntactic and semantic aspects of grammar, such as passive voice, grammatical tenses, plural marking, and adverbs. None of the studies explicitly focused on learning of morphology.

Generally, it has been recognized that while the results are promising, additional empirical studies with more rigorous research designs (randomized controlled trials) are called for to comprehensively investigate and compare the effectiveness of CALL with traditional instructional methods (Cerezo et al., 2014; Grgurović et al., 2013). Overall, the reported results provide evidence supporting the potential usefulness of mobile devices in grammar teaching and learning. However, there are two gaps in this field. First, most MALL studies related to learning of grammar are conducted in English; second, few studies explicitly focus on learning of morphology through MALL. The current study addresses these issues by investigating learning of morphology through MALL in a language rich in morphology, Finnish.

## FMA

FMA is a smartphone application developed by researchers at Åbo Akademi University and the University of Turku. The aim of the application is to teach beginner level Finnish language learners to produce transparent and semi-transparent inflections and use them in a sentence context. It consists of 81 word items that are practiced in nominative and inflected forms. Each item is presented with an audio cue and an image illustrating the item. The learner's task is to correctly identify what is shown in the image, for example, 'What is this?' → 'This is **a cat**' or inflect it, 'Where is the cat?' → 'The cat is **in the basket'**, or → 'The cat is **on the table**' (See Figure 1). The items of the FMA are distributed over 7 study modules: 4 vocabulary modules and 3 grammar modules. Vocabulary modules are presented before the respective grammar modules, in which the participants are asked to inflect the words from the vocabulary modules and use them correctly in a sentence context.

The items are practiced in three conditions: nominative (e.g. kissa, 'cat'), transparent inflections (e.g. metsä: metsästä, 'from the forest') and semi-transparent inflections (e.g. laati**kk**o: laati**k**ossa; 'in a box'). Each condition is practiced in its own study module. To complete the training program, the learner must complete 141 exercises for the nominative and 145

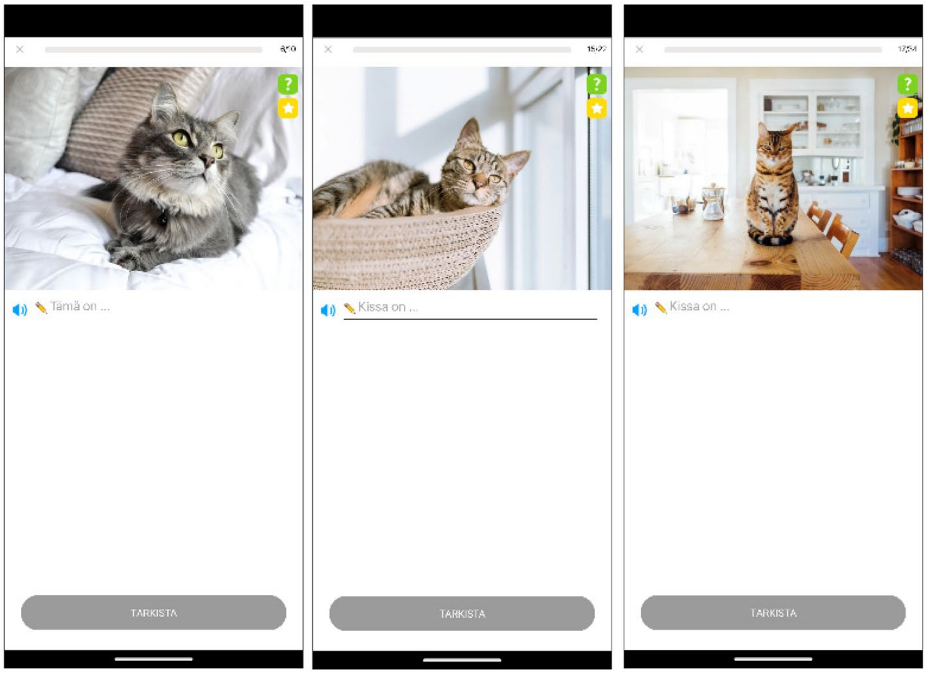


**Figure 1.** An example of three grammatical conditions that are trained in the FMA. Conditions from left to right: (1) nominative: 'This is a cat (kissa)'; (2) transparent inflection: 'The cat is in the basket (kori: korissa)'; (3) semi-transparent inflection: 'The cat is on the table (pöytä: pöydällä)'.

exercises for the inflection modules. The transparent inflections are divided into two separate modules, i.e. internal (53 items), external (39 items), and semi-transparent inflections are presented in a separate module (53 items). After each module, there is a short test. After this, the application starts to repeat the most difficult modules based on registered reaction times and accuracy for each item (for details, see the section *adaptive spacing*).

The items are concrete everyday Finnish words, such as *cat*, *car*, or *table*. In case the learner does not know the right answer merely by looking at the image, they can use an audio cue ('speaker' icon) or a written cue ('question mark' icon). By tapping the speaker icon, the learner hears the correct answer. The help button provides the learner with the right answer in written form. Thus, it is not possible to get stuck in a certain exercise. For the test section, the question mark or speaker icon are not enabled, and the test repeatedly prompts the test items until the learner gives a correct answer. The application is monolingually Finnish, as not all Finnish L2 learners are familiar with English.

## *Spaced repetition*

As a learning method, the FMA utilizes spaced repetition, which has proven to be beneficial to enhance learning, also in the second language

(Kim & Webb, 2022). Spaced repetition refers to repetition of study material with temporal gaps between the repetitions; in other words, the learning sessions are spread over longer periods of time, as opposed to massed repetition, where the same amount of study material is practiced multiple times over a short period of time (see e.g. Kim et al., 2019; Kim & Webb, 2022; Metzler-Baddeley & Baddeley, 2009). Spaced repetition to improve subsequent retrieval has been extensively studied, and its efficiency has been demonstrated both in laboratory settings (Cepeda et al. 2008), clinical interventions (Balota et al., 2006; Green et al., 2014), and classroom teaching (Kapler et al., 2015; Sobel et al., 2011). In the FMA, spaced repetition is used to repeat the most challenging exercises for each learner based on their performance in the previous attempt.

### *Adaptive spacing*

Several studies suggest that the perceived item difficulty and the overall study time available affect how much time learners allocate to studying a certain item (see e.g. Metcalfe & Kornell, 2003; Son & Metcalfe, 2000). Adaptive spacing takes this into account by focusing on items that an individual struggles with the most and adjusts the spacing delays between trials accordingly. Adaptive spacing has been shown to produce significantly higher immediate and delayed recall rates than random spacing (Metzler-Baddeley & Baddeley, 2009). Adaptive spacing has also been implemented in the FMA by using item reaction times and accuracy in conjunction with elapsed time as an indicator of item stability. The forgetting curve of Ebbinghaus is used here to quantify the decay rate of a studied item (Ebbinghaus, 1885/1964). In other words, the retrievability of a given item decreases exponentially as a function of stability. In practice this means that when starting a new session in the FMA, the least stable items will be presented again, making the spaced repetition method adaptive for each learner.

### *Test effect*

According to several studies, the mere act of taking a test enhances the learning outcome and retrieval success, sometimes even more than passive studying or repetition (e.g. Roediger & Karpicke, 2006). This is referred to as a testing effect. The FMA utilizes the test effect by presenting a small test after each study unit. The test module is otherwise identical to study modules, but it has no option to use auditory or written cues. The test is a pass/fail test where the participant needs to get all answers correct before continuing to the following modules. After each item, the correct response is presented to the learner.

### Theoretical underpinning

The theoretical underpinning of the FMA builds upon usage-based language models of language (UBL), which suggest that a language is learned implicitly by encountering it in different usage contexts (Eskildsen, 2009; Tomasello, 2009). The major concept of usage-based approaches regarding this study is exemplar-based learning (Ellis & Wulff, 2015; Javadi & Kazemirad, 2020) which posits that 'whenever the learner encounters an exemplar of a construction, the language system starts to compare this exemplar with previous encounters of the same or similar exemplar in order to retrieve the correct interpretation. Based on exemplar theory, constructions emerge over time as the learner's language system, and this system processes exemplar after exemplar, identifies the existing regularities and then makes the corresponding abstractions' (Javadi & Kazemirad, 2020, p. 475). The FMA aligns with this framework as the morphosyntax is trained without explicit grammar instruction. Instead, the learner grasps new grammatical units from the cloze-tasks, by trial and error creates linguistic analogies of these forms, and finally starts to produce new expressions based on these analogies. In other words, the learning takes place by identifying the regularities in the exemplars.

### Technology component

The FMA is embedded as a native application for iOS and Android platforms. It is available for free in app stores currently by the project name LexLearn. Following download, the user can access the FMA after authentication with an email address and a password. Initially, users are presented with a list of study modules, with only the first module enabled at the start. Subsequent modules become accessible sequentially as users successfully complete each preceding study module and test module. When the entire list of study modules is completed, the FMA starts to use the sampled data from participants' onset typing times and accuracy for each item, re-activating the most challenging modules and/or items that were trained the longest time ago. The user interface (UI) is presented in Figure 2. For the functionalities of the interactive elements in each cloze-task exercise, see Figure 3. During the user interaction, the FMA registers the completed units and modules and provides a timestamp for each of these sequence events in a password-secured database. In addition, the system stores data related to onset typing times and users' responses for each item.

### The current study

The aim of the present study was to develop a Finnish morphology application and test its effectiveness in L2 learning of morphology (i.e.

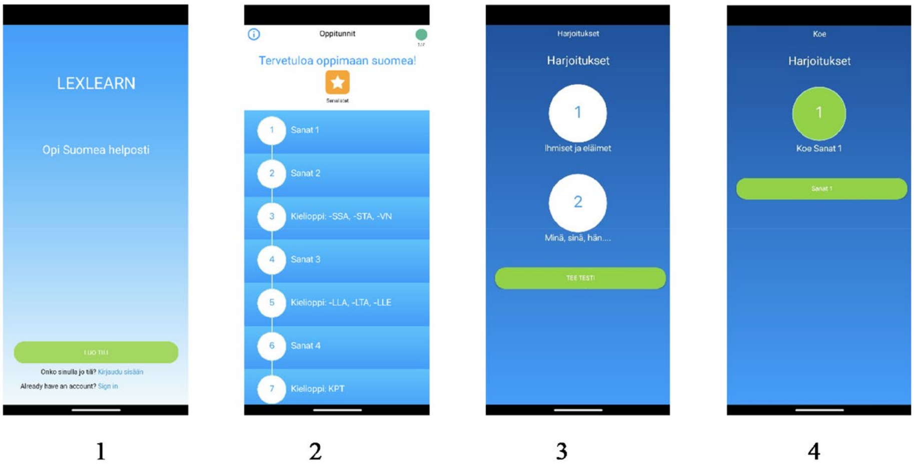


**Figure 2.** The user interface of the study modules 1–7 and test sections in the FMA. After logging in (Pane 1), the user starts from the module 1 (Pane 2). After completing the study module 1, (Pane 3), they will be directed to the test section (Pane 4). After passing the test section, the user will be redirected to navigation menu and the next module will be activated (Pane 2).

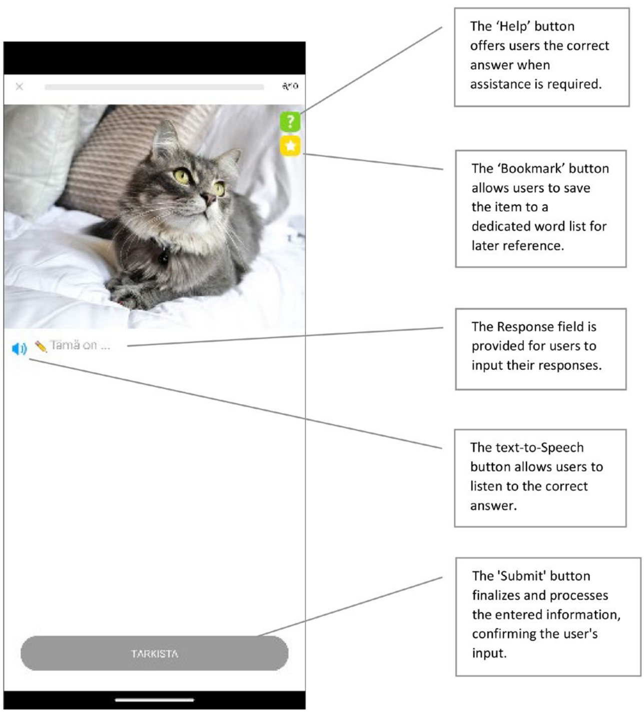


**Figure 3.** Functionalities of the interactive elements in a cloze-task exercise.

transparent and semi-transparent inflections) at the receptive and productive level. Based on previous mobile-assisted grammar learning and adaptive training studies, our hypothesis was that training with the FMA will improve Finnish L2 speakers' morphological skills, particularly in production, which is explicitly practiced in the application, but also in receptive skills, as they were also practiced by reading self-written words.

## Method

### *Participants*

A total of 173 participants were recruited from Finnish L2 institutions in the Turku region of Finland. The participants were non-native beginner level Finnish learners. The data were collected in four rounds: October 2019, March 2020, October 2021, and March 2022. Of the 173 enrolled participants, 53 terminated the study prior to pretest and 29 withdrew during the training period. Of the remaining 91 participants, 20 participants were excluded for one of the following reasons: a diagnosed language disorder ($n = 2$), inactivity during the training period ($n = 13$, reported training less than 15 min per day), or not appearing at the posttest ($n = 5$). Thus, the final sample consisted of 71 participants (45 females and 26 males). We collected background information regarding age, gender, exposure to Finnish, education level, motivation to learn Finnish, motivation to use MALL, and mobile-use acceptance. Based on these factors, we formed two groups, i.e. experimental group training morphology with the FMA, and an active control group, using an alternative Finnish language learning application without a specific focus on morphology. The two groups were largely similar regarding these aspects. Relevant participant characteristics of both the experiment and the control group are listed in Table 2. The sample included 30 native languages, the most common one being German ($n = 16$). The language background of the participants is listed in detail in Appendix 1 (supplementary material).

### *Design and procedure*

This intervention study was a single-blind[1] randomized controlled trial divided into four main stages over a 5-week period (see Figure 4). The first stage was recruitment, which was conducted by visiting classrooms at beginner level Finnish courses. The pretests took place approximately one week after recruitment. The participants were tested in an IT class. One session took 1–1.5 h. Subsequently from each pair of participants, one was randomly allocated to the training group and the other to the active control group. The pairing was based on age, gender, and exposure to Finnish. This was done to attain the optimal balance between the two groups, as the sample size in each round was relatively small. A week after the pretest, both groups were sent an email containing a link to

**Table 2.** Descriptive data of the two participants groups; mean (SD).

| | FMA group ($n = 33$) | Active control group ($n = 38$) |
|---|---|---|
| Age | 23.90 (3.07) | 26.58 (6.05) |
| Gender, female/male | 20/13 | 25/13 |
| Exposure[a] | 3.33 (3.91) | 5.19 (8.37) |
| Kielo Score[b] | 16.1 (6.9) | 15.2 (5.6) |
| Motivation to learn Finnish[c] | 3.67 (3.54) | 3.54 (1.22) |
| MALL motivation before[d] | 4.39 (0.83) | 4.31 (0.92) |
| MALL motivation after[e] | 3.20 (1.26) | 3.26 (1.25) |
| Mobile-use acceptance[f] | 4.42 (0.90) | 4.39 (0.67) |
| Education level (*n*)[g] | | |
| 1 Primary/Lower Secondary | 0 | 1 |
| 2 Higher Secondary | 8 | 10 |
| 3 Vocational School | 1 | 2 |
| 4 University of Applied Science (Bachelor) | 2 | 2 |
| 5 University (Bachelor or Master) | 22 | 22 |

[a]Length of studying Finnish in months.
[b]Score in General Language Test Kielo on scale 0–30.
[c]Self-rating based on a 5-point Likert scale: How motivated are you to learn Finnish? 1 = not motivated at all, 5 = extremely motivated.
[d]Self-rating based on a 5-point Likert scale: How motivated were you to use the application in the beginning of the training period? 1 = not motivated at all, 5 = extremely motivated.
[e]Self-rating based on a 5-point Likert scale: How motivated were you to use the application in the end of the training period? 1 = not motivated at all, 5 = extremely motivated.
[f]Self-rating based on a 5-point Likert scale: How comfortable do you feel using new smartphone applications in general? 1 = very uncomfortable, 5 = very comfortable.
[g]One participant did not report educational level.

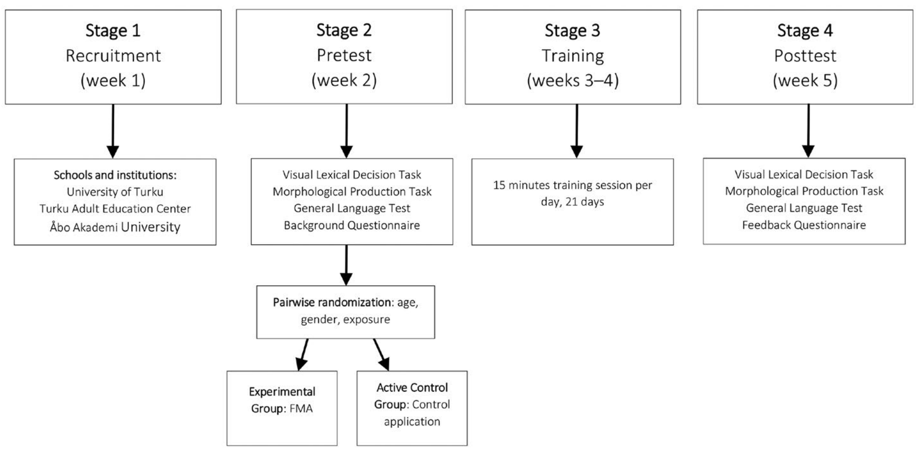


**Figure 4.** The four stages of the data collection.

their respective applications. Specifically, the experimental group received a link to download the FMA, while the control group received a link to download the Suomipassi application. In addition to the link, both groups were given identical instructions to practice for 15 min per day over a period of three weeks using the provided application. The participants were blinded to condition, i.e. they did not know if they belonged to the experiment group or the control group. The length of the training period was 21 days for each participant. During the training period, the students continued their studies in their respective courses, which were basic level courses of Finnish (A1–A2). The participants received a weekly email reminder to prompt to use their respective application. Concurrently, the FMA collected data of user behavior, i.e. how many study modules they had completed per day. To assess the training times in the control group for which such automatic user behavior tracking was not available, both groups were also advised to record their training times in a hand-written log.[2] In addition, they were asked to evaluate their daily use of the application and user experience in a post-survey. After the training period, the participants completed the posttest during week 5. The reward for each participant who completed the whole study was two film tickets or a gift card worth 10–20 euros.

### *Target application contents*

The target group performed the training session with the FMA. For the description, see the section 'FMA'.

### *Contents of the application used by the control group*

The active control group performed the training session with an application called *Suomipassi* ('Finnish Language Passport'; Suomipassi, 2020), which is a smartphone application developed by the University of Turku Language Center. It teaches the learner basic phrases and pronunciation in Finnish (e.g. 'Excuse me, is this seat free?'). The learner listens to exemplar phrases, repeats them orally, and is encouraged to apply these expressions in real-life situations. The learners can earn stamps every time they use the phrases in real life outside the house, however, such real-life usage was not required in this study. There are 18 modules related to different situations or topics like 'in the post office' or 'buying paying', and the users can flexibly choose which sentences they wish to practice within a module without a strict sequential order. The sentences can be translated to nine different languages and can be expressed in colloquial Finnish as well. Although morphology is present in the *Suomipassi* phrases, the application does not focus on teaching morphological rules in any systematic way. *Suomipassi* was selected to be the application used by the control group, as it is an attractive, non-commercial mobile application, and learning everyday phrases was considered to be a motivating task for Finnish language learners. As was the case for the FMA, the theoretical underpinning of the Suomipassi app follows largely the UBL model, more specifically exemplar-based learning. Furthermore, the theoretical underpinning can be identified as rooted in task-based language learning, given that learners were encouraged to utilize the phrases in real-life conversations.

### Pre- and posttest tasks

To assess the morphological training outcomes for both receptive and productive skills, we composed two pre- and posttest tasks, namely a Visual Lexical Decision task (VLD) and a Morphological Production (MP) task. In addition, the general level of proficiency in Finnish was assessed with a shortened version of Kielo (Tani, 2008), a general language test in Finnish. The tasks were performed in the above-mentioned order. The tasks were identical in the pretest and posttest and were administered on a PC in a quiet room in small groups. One participant completed the posttest separately in a quiet room alone with the researcher.

#### VLD

In this task, participants are presented with a string of letters one by one on a computer screen. The participants are asked to judge as quickly as possible whether the target string is a real Finnish word or not by a button press. A fixation point appeared before each stimulus for 500 ms. The stimulus remained on the screen until the participant responded or until the item timed out (4000 ms). A set of 165 nouns were selected from a popular A1-level Finnish textbooks series (Suomen Mestari 1-2) in Finnish L2 instruction, in order to maximize the familiarity of the words for the participants. The target words were divided over three conditions: nominative (e.g. *lääkäri*, 'doctor'), transparent locatives (e.g. *aamu: aamu + lla*, 'in the morning') and semi-transparent locatives (e.g. *ilta: illa + lla*; 'in the evening'). There were 55 items per condition. To test generalization effects, 25 items in each condition were trained items that appeared in the FMA training modules, and 30 items were untrained items that did not appear in the training modules. Both within the trained group and untrained group of items, the transparent and semi-transparent conditions were matched on lemma frequency, length and bigram frequency. Frequencies were taken from a 22.7 million-word newspaper database (WordMill; Laine & Virtanen, 1999). Given the restrictions during item selection—items had to be depictable and appear within A1-level textbooks—we did not manage to match all conditions on lemma frequency and length, but we accounted for this in our statistical analyses by including these variables as control variables in our models. The VLD also included 145 pseudowords, most of them mimicking the three target noun conditions. Thus, approximately 1/3 of the pseudowords did not have any apparent sublexical structure, 1/4 were semi-transparent pseudoword inflections, 1/4 were transparent pseudoword inflections and other pseudowords contained a noun stem and a pseudo-suffix. The pseudowords were similar in length and bigram frequency as the real words. In Table 3, a summary of lexical characteristics of the VLD words are listed. The VLD task was performed by using the open-source experimental software PsychoPy (Peirce et al., 2019). Accuracy rate and lexical decision time were employed as dependent variables. The accuracy score was defined by counting the number of correctly identified items out of a total of 310 items.

### *MP*

In the MP, the participants were presented pictures of non-specific objects linked to pseudowords like 'mosa' (see Figure 5, left panel) and were instructed to either reproduce them in the uninflected nominative form or to inflect them and include a locative case that follows Finnish grammar rules. The target cases were nominative, transparent inflections in a locative case, and semi-transparent inflections in a locative case, i.e. pseudowords with stems that contained CG phonemes. The target pseudoword was provided in the nominative. The different conditions were matched on word length and bigram frequency. To elicit the nominative form, the participant was prompted to embed the target item in a sentence by answering the question *Mikä tämä on?* ('what is this?'); 'Tämä on *mosa*', ('This is a mosa', see Figure 5, left panel). For locatives, a familiar object (i.e. *kissa,* 'cat') was either in, on, moving away from or

**Table 3.** Summary of lexical characteristics of the VLD items. Lemma frequency is scaled to 1 million and subsequently logarithmically transformed, bigram frequency is scaled to 1000.

| | | Loglem | Length | Bigram |
|---|---|---|---|---|
| Trained items | Nominative | 7.02 | 6.47 | 7.63 |
| | Transparent | 7.09 | 8.42 | 8.18 |
| | Semi-transparent | 6.44 | 8.49 | 8.27 |
| Untrained items | Nominative | 6.79 | 8.07 | 9.01 |
| | Transparent | 7.44 | 7.13 | 8.27 |
| | Semi-transparent | 7.62 | 7.40 | 8.00 |

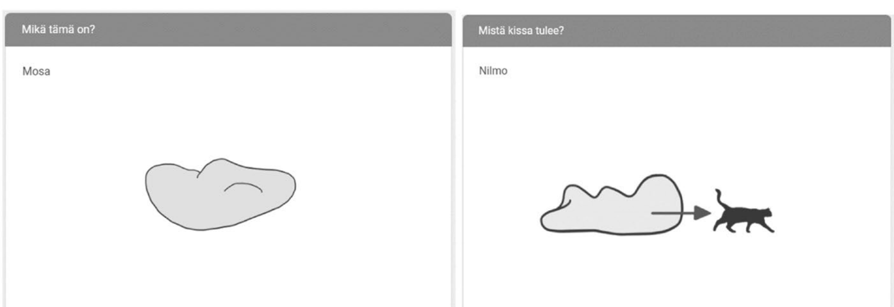


**Figure 5.** Example of the MP task. At the left, only a nonspecific object (mosa) is depicted to elicit the nominative of the pseudoword. At the right a nonspecific object is depicted next to an arrow and a familiar object to elicit a locative case (Kissa tulee nilmosta, 'The cat is coming from the nilmo').

moving towards a non-specific object. To elicit the inflected forms, the participant was prompted to embed the target item in a sentence by answering the question *Mistä kissa tulee?* ('Where is the cat coming from?'); *Kissa tulee nilmosta* ('The cat is coming from the *nilmo*', see Figure 5, right panel). The participants wrote only the target word (*mosa*, *nilmosta*), not the whole sentence. A practice session of 10 items preceded the actual experiment. After the practice session, 90 questions (30 for nominatives, 30 for transparent inflections and 30 for semi-transparent inflections) were presented in a random order. The task was unspeeded and the written answers were registered *via* Google Forms. Scoring involved assigning 1 point for a correct response and 0 points for an incorrect response. A response was counted as incorrect if there were any mistakes, i.e. typographical or grammatical errors. The maximum score on the test was 90 points. The lexical characteristics of the pseudoword conditions were matched for length and average bigram frequency across conditions. Table 4 lists the characteristics of the MP items.

### General language proficiency

General language proficiency was assessed *via* pen and paper by a subset of tasks included in Kielo (Tani, 2008), a standardized test material package intended to determine the level of Finnish L2 learners' language skills and to monitor their development. The first 26 exercises (tasks 1–3) were selected for the test. The questions addressed everyday topics, such as going to the store or understanding the purpose of various signs and advertisements. Scoring was conducted according to the instructions provided in the material package for each task. The maximum score for the test was 30.

### Pre- and post-survey

In the pre-survey, we collected information regarding the participants background factors, e.g. age, gender, native language and familiarity of using digital technology in daily life. In the post-survey, we collected information about the participants' user experience regarding their respective applications. The post-survey included 10 questions about user experience: App enjoyment, comprehension of UI, technical performance, desire to continue, ease of exercises, motivation before training,

**Table 4.** Characteristics of the pseudowords in the MP task.

| Condition | Length | | Bigram frequency | |
|---|---|---|---|---|
| | M | SD | M | SD |
| Nominative | 4.97 | 0.48 | 6.48 | 2.49 |
| Transparent | 4.93 | 0.75 | 6.48 | 2.42 |
| Semi-transparent | 4.67 | 0.75 | 6.32 | 2.87 |

motivation after training, and perceived efficacy of learning Finnish grammar, vocabulary, and communication.

### *Pre-registration*

The current study was pre-registered on AsPredicted.com at https://aspredicted.org/blind.php?x=GVL_1FJ.

### *Statistical analyses*

Data were analyzed with linear mixed effect models (LMM) using the lme4 package (Bates et al., 2015) in the R statistical software (Version 4.21; R Core Team, 2021). Separate models were built for the following measures: MP accuracy, VLD accuracy and VLD reaction time (RT). In the MP analysis, we used a generalized linear mixed effect model with accuracy rate as a dependent variable. Proficiency and the interaction of morphological condition * group * time point were added as fixed effects.

In VLD, a generalized linear mixed effect model was used for the accuracy rate and a linear mixed effect model for RTs. Proficiency, word length, lemma frequency and an interaction of morphological condition * group * time point were added as fixed effects. We made separate models for words that were trained in the FMA and words that were untrained to examine whether the FMA group's performance improved not only for trained but also untrained words. For the RT data, responses below 250 ms and incorrect responses were removed and log transformations were made to normalize the data. Further data trimming (exclusion of the RTs over 2.5 SD) was not conducted, as the comparison of non-filtered to filtered models indicated a better $R^2$ value for the unfiltered model.

For all models, condition (nominative, transparent inflection, semi-transparent inflection), group, and time point were fitted as a treatment-coded fixed effect variable. Proficiency was a centered composite variable based on the mean of the standardized values of the exposure score, self-rating in Finnish, and Kielo pretest score. Participants and items were included as random intercepts in all models. For the sake of simplicity, only significant effects are reported in the text. The analysis, scripts, and the datafiles are available at https://osf.io/csg6n/?view_only=67cea79851d4499399df4ace99f20d4e.

Degrees of freedom are difficult to determine in the LMM analyses, due to which no exact *p*-value can be calculated (Baayen et al., 2008). Statistical significance at the .05 level is indicated by values of $|t \text{ or } z| > [\pm\ 1.96]$.

For the user feedback, the Kruskall-Wallis test was used to analyze whether there was a difference within the experienced learning efficacy as a function of three linguistic domains (grammar, vocabulary, and communication skills). Post hoc tests were performed using Dunn's multiple comparisons test. The difference between motivation ratings before and after training was analyzed with the paired Wilcoxon Rank-Sum test.

## Results

We present the analyses for three outcome variables: MP accuracy, VLD accuracy and VLD RTs. The VLD results are presented separately for the trained and untrained words. Observed means and standard deviations of the pre- and posttest measures for each condition are presented in Table 5.

## Effects of FMA training

### *Morphological production task*

Generalized linear mixed effect model for posttest accuracy in the MP task revealed a main effect of proficiency: the higher the proficiency, the higher the overall accuracy in both groups. In addition, there was a main effect of time point: both groups performed better in the posttest than in the pretest. Moreover, there was a main effect of group: the FMA group performed better than the active control group. There was also a main effect of condition: nominatives elicited higher accuracy rates than

**Table 5.** Accuracy rates (% correct) and RT (in ms) for the outcome measures (means and SDs).

| | | | FMA ($n = 33$) | | | | Control ($n = 38$) | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Pre | | Post | | Pre | | Post | |
| Measure | Outcome variable | Condition | M | SD | M | SD | M | SD | M | SD |
| MP | Accuracy | Nominative | 0.81 | 0.40 | 0.86 | 0.35 | 0.74 | 0.44 | 0.85 | 0.36 |
| | | Transparent | 0.48 | 0.50 | 0.88 | 0.32 | 0.42 | 0.49 | 0.65 | 0.48 |
| | | Semi-transparent | 0.28 | 0.45 | 0.5 | 0.50 | 0.19 | 0.40 | 0.28 | 0.45 |
| VLD[1] | Accuracy: trained items | Nominative | 0.75 | 0.43 | 0.92 | 0.28 | 0.75 | 0.43 | 0.78 | 0.41 |
| | | Transparent | 0.72 | 0.45 | 0.93 | 0.26 | 0.77 | 0.42 | 0.80 | 0.40 |
| | | Semi-transparent | 0.72 | 0.45 | 0.89 | 0.31 | 0.73 | 0.45 | 0.71 | 0.45 |
| | Accuracy: untrained items | Nominative | 0.71 | 0.45 | 0.65 | 0.48 | 0.73 | 0.45 | 0.72 | 0.45 |
| | | Transparent | 0.69 | 0.46 | 0.72 | 0.45 | 0.75 | 0.43 | 0.75 | 0.44 |
| | | Semi-transparent | 0.70 | 0.46 | 0.70 | 0.46 | 0.73 | 0.45 | 0.71 | 0.45 |
| VLD[2] | RT: trained items | Nominative | 1127 | 560 | 946 | 438 | 1128 | 565 | 999 | 496 |
| | | Transparent | 1495 | 715 | 1202 | 578 | 1490 | 745 | 1180 | 600 |
| | | Semi-transparent | 1495 | 688 | 1249 | 582 | 1479 | 674 | 1293 | 637 |
| | RT: untrained items | Nominative | 1306 | 586 | 1216 | 558 | 1343 | 663 | 1166 | 551 |
| | | Transparent | 1276 | 644 | 1200 | 575 | 1349 | 666 | 1151 | 542 |
| | | Semi-transparent | 1296 | 620 | 1260 | 615 | 1370 | 664 | 1163 | 561 |

[1,2]VLD sample size: $n = 32$ (FMA), $n = 36$ (Control); Data loss due to technical problems.

transparent inflections, which in turn elicited higher accuracy rates than semi-transparent inflections.

There was also an interaction between time point and morphological condition: more errors were made in producing transparent and semi-transparent forms in comparison to nominative forms at both time points, but the gap between nominatives and the transparent and semi-transparent condition had reduced in the posttest in comparison to the pretest. Most importantly, there was an interaction between group, time point and condition. This indicates that the FMA group improved more than the control group on both the transparent and semi-transparent inflections, but not on the nominatives (see Figure 6). Table 6 lists all the fixed effects.

## Summary of the results on the MP task

The aim of the MP task was to assess how a three-week training with the FMA affects morphological production in the target group. The FMA worked as expected, i.e. training with the FMA improved the target group's morphological production skills, whereas the morphological skills of the active control group improved less. In addition, when looking at both

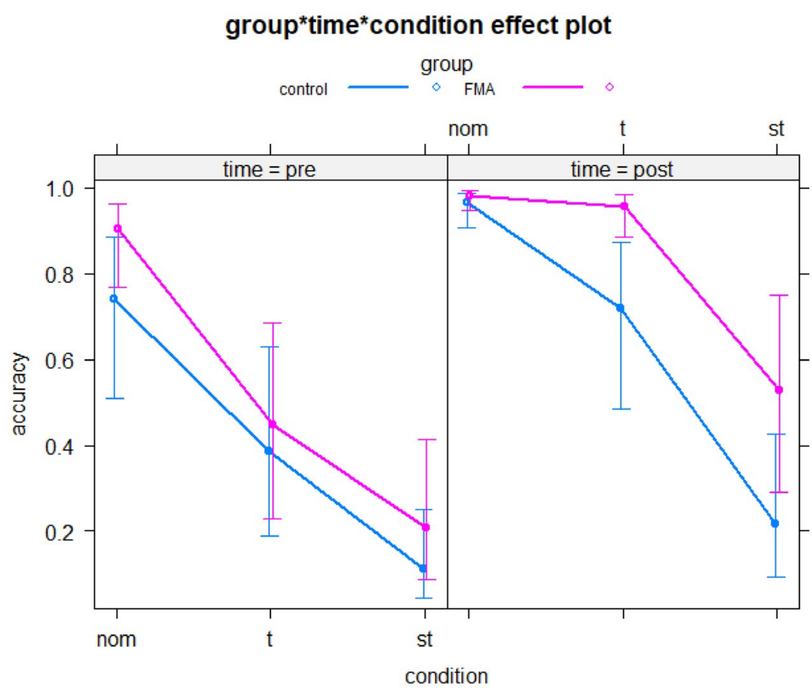


**Figure 6.** The interaction between group, time point and morphological condition. As illustrated, the FMA group improved more on the transparent (t) and semi-transparent inflections (st) than the control group, leading to a bigger group difference at the posttest (right panel) than at the pretest (left panel).

**Table 6.** Fixed Effects of the Accuracy Rates in the MP Task.

| | Accuracy | | | |
|---|---|---|---|---|
| *Predictors* | *Odds Ratios* | *Std. Error* | *CI* | *Statistic* |
| (Intercept) | 2.79 * | 1.42 | 1.02 – 7.57 | 2.01 |
| Proficiency | 2.98 *** | 0.78 | 1.78 – 4.98 | 4.17 |
| Time [post] | 9.77 *** | 2.32 | 6.14 – 15.55 | 9.61 |
| Group [FMA] | 3.27 ** | 1.28 | 1.52 – 7.06 | 3.02 |
| Condition [t][a] | 0.22 * | 0.14 | 0.06 – 0.77 | −2.38 |
| Condition [st][b] | 0.04 *** | 0.03 | 0.01 – 0.15 | −4.93 |
| Time [post] * group [FMA] | 0.60 | 0.23 | 0.28 – 1.27 | −1.34 |
| Time [post] * condition [t] | 0.41 *** | 0.11 | 0.25 – 0.69 | −3.39 |
| Time [post] * condition [st] | 0.23 *** | 0.06 | 0.13 – 0.39 | −5.46 |
| Group [FMA] * condition [t] | 0.39 *** | 0.08 | 0.25 – 0.60 | −4.33 |
| Group [FMA] * condition [st] | 0.64 | 0.15 | 0.40 – 1.01 | −1.92 |
| (time [post] * group [FMA]) * condition [t] | 11.04 *** | 4.73 | 4.77 – 25.54 | 5.61 |
| (time [post] * group [FMA]) * condition [st] | 3.23 ** | 1.37 | 1.41 – 7.40 | 2.77 |

[a]*t* = transparent inflection.
[b]st = semitransparent inflection.

pre- and posttest scores, all participants made more errors with morphologically complex forms in comparison to the nominatives, but training with the FMA reduced this effect in the target group to a larger extent than in the control group. Taken together, the results suggest that morphologically complex forms were initially more difficult to produce, but that training with the FMA helped to improve morphological production skills significantly in a relatively short period.

## VLD

### Accuracy

***Accuracy for trained items.*** For the trained items, the model for recognition accuracy revealed a main effect of participants' proficiency: the more proficient the participants were in Finnish, the better their recognition accuracy. There was also a main effect of time point: both groups' accuracies were higher in the posttest. The model also indicated a main effect of word length and lemma frequency: for shorter and more frequent words, the accuracy rates were higher. In addition, an interaction was found between group and time point: the recognition accuracy of the FMA group improved more during the training period than that of the control group across all morphological conditions. Table 7 lists the fixed effects and Figure 7 illustrates the interaction between group and time point.

***Accuracy for untrained items.*** For the untrained items, the model for recognition accuracy in the VLD task revealed again a main effect of participants' proficiency: the more proficient the participants were in Finnish, the better their recognition accuracy. In addition, there was an interaction between time point and group: overall, the FMA group performed slightly worse in the posttest than the control group, which mainly stems from the

**Table 7.** Fixed Effects of the Accuracy Rates in the VLD Task for trained items.

| | Accuracy | | | |
|---|---|---|---|---|
| *Predictors* | *Odds Ratios* | *std. Error* | *CI* | *Statistic* |
| (Intercept) | 0.39 | 0.29 | 0.09 – 1.66 | −1.27 |
| Proficiency | 1.71 *** | 0.21 | 1.34 – 2.18 | 4.28 |
| Time [post] | 1.28 * | 0.16 | 1.00 – 1.64 | 1.97 |
| Group [FMA] | 0.96 | 0.19 | 0.65 – 1.42 | −0.21 |
| Condition [t][a] | 0.82 | 0.22 | 0.48 – 1.40 | −0.72 |
| Condition [st][b] | 0.77 | 0.21 | 0.45 – 1.30 | −0.99 |
| Length | 1.12 * | 0.05 | 1.02 – 1.23 | 2.34 |
| Logarithmic lemma frequency | 1.26 ** | 0.10 | 1.09 – 1.47 | 3.09 |
| Time [post] × group [FMA] | 3.43 *** | 0.71 | 2.30 – 5.14 | 6.01 |
| Time [post] × condition [t] | 0.97 | 0.17 | 0.68 – 1.37 | −0.18 |
| Time [post] × condition [st] | 0.72 | 0.12 | 0.51 – 1.01 | −1.93 |
| Group [FMA] × condition [t] | 0.71 | 0.12 | 0.50 – 1.00 | −1.95 |
| Group [FMA] × condition [st] | 0.90 | 0.16 | 0.64 – 1.26 | −0.64 |
| Time [post] × group [FMA]) × condition [t] | 1.47 | 0.43 | 0.83 – 2.61 | 1.32 |
| Time [post] × group [FMA]) × condition [st] | 1.17 | 0.32 | 0.68 – 2.01 | 0.55 |

[a] *t* = transparent inflection.
[b] *st* = semitransparent inflection.

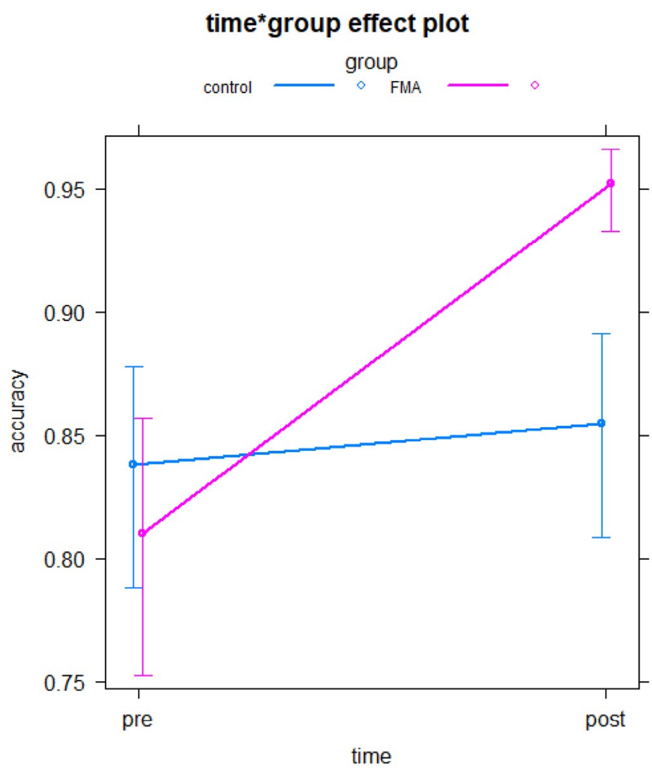


**Figure 7.** The interaction between time point and training group in VLD accuracy for trained items. As illustrated, the FMA group improved more than the control group, leading to a difference between groups at the posttest.

slightly worse performance on the nominative forms in the posttest compared to the pretest (Figure 8). Related to this, the results indicated an interaction between time, group, and condition: the FMA group seemed to improve more in transparent inflections in comparison to nominatives than the

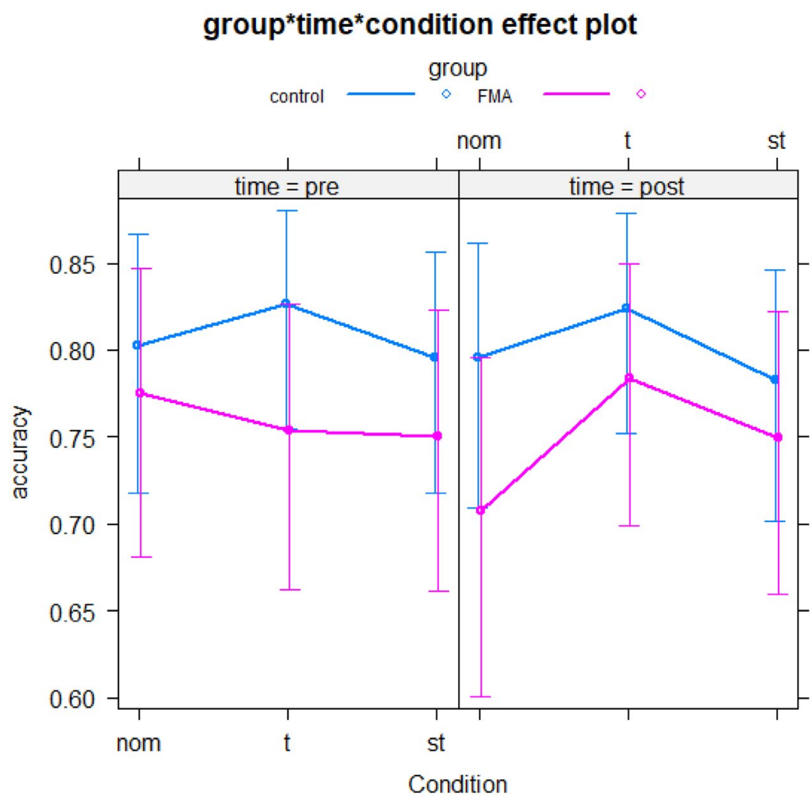


**Figure 8.** The interaction between group, time point and morphological condition for accuracy in untrained words.

control group did. However, the significant interaction seems primarily to be due to the decline in performance for the nominative baseline in the posttests (Figure 8). The statistics for each variable are presented in Table 8.

### *Reaction times*

***RTs for trained items.*** For trained items, there was a main effect of time point: both groups' RTs speeded up between pre- and posttests. In addition, morphological condition had a main effect: transparent inflections and semi-transparent inflections were recognized more slowly than the nominatives. The model indicated also a main effect of word length and lemma frequency: the shorter and more frequent the word was, the faster the reaction times. In addition, the results showed an interaction between time point and condition: the difference between nominatives and transparent inflections was smaller in the posttest than the pretest. No significant effects between groups were found in reaction times. Table 9 lists the fixed effects and Figure 9 illustrates the interaction between time point and condition.

***Reaction times in untrained items.*** In the untrained items, there was only a main effect of time point: both groups' reaction times became faster between

**Table 8.** Fixed Effects of the Accuracy Rates in the VLD Task for untrained items.

| | Accuracy | | | |
|---|---|---|---|---|
| *Predictors* | *Odds Ratios* | *Std. Error* | *CI* | *Statistic* |
| (Intercept) | 0.65 | 0.56 | 0.12 – 3.52 | −0.51 |
| Proficiency | 1.70 *** | 0.23 | 1.31 – 2.22 | 3.96 |
| Time [post] | 0.96 | 0.11 | 0.77 – 1.19 | −0.39 |
| Group [FMA] | 0.85 | 0.17 | 0.56 – 1.27 | −0.81 |
| Condition [t][a] | 1.17 | 0.32 | 0.69 – 1.98 | 0.57 |
| Condition [st][b] | 0.96 | 0.26 | 0.57 – 1.61 | −0.17 |
| Length | 1.13 | 0.09 | 0.96 – 1.32 | 1.44 |
| Lemfreq [log] | 1.13 | 0.09 | 0.97 – 1.31 | 1.53 |
| Time [post] × group [FMA] | 0.73 * | 0.12 | 0.54 – 1.00 | −1.97 |
| Time [post] × condition [t] | 1.03 | 0.16 | 0.76 – 1.40 | 0.19 |
| Time [post] × condition [st] | 0.97 | 0.15 | 0.71 – 1.31 | −0.22 |
| Group [FMA] × condition [t] | 0.76 | 0.12 | 0.56 – 1.04 | −1.73 |
| Group [FMA] × condition [st] | 0.92 | 0.14 | 0.67 – 1.24 | −0.56 |
| (Time [post] × group [FMA]) × condition [t] | 1.63 * | 0.36 | 1.06 – 2.53 | 2.21 |
| (Time [post] × group [FMA]) × condition [st] | 1.46 | 0.32 | 0.95 – 2.25 | 1.72 |

[a]*t* = transparent inflection.
[b]*st* = semitransparent inflection.

**Table 9.** Fixed Effects of the RT in the VLD Task for trained items.

| | Log(RT) | | | |
|---|---|---|---|---|
| *Predictors* | *Estimates* | *Std. Error* | *CI* | *Statistic* |
| (Intercept) | 6.86 *** | 0.10 | 6.67 to 7.05 | 69.50 |
| Proficiency | −0.00 | 0.05 | −0.09 to 0.09 | −0.06 |
| Time [post] | −0.11 *** | 0.02 | −0.15 to −0.07 | −5.46 |
| Group [FMA] | 0.01 | 0.06 | −0.11 to 0.13 | 0.16 |
| Condition [t][a] | 0.16 *** | 0.04 | 0.10 to 0.23 | 4.64 |
| Condition [st][b] | 0.13 *** | 0.04 | 0.06 to 0.20 | 3.57 |
| Length | 0.04 *** | 0.01 | 0.03 to 0.05 | 6.97 |
| Lemfreq [log] | −0.02 * | 0.01 | −0.04 to −0.00 | −2.50 |
| Time [post] × group [FMA] | −0.04 | 0.03 | −0.10 to 0.01 | −1.51 |
| Time [post] × condition [t] | −0.11 *** | 0.03 | −0.17 to −0.06 | −3.95 |
| Time [post] × condition [st] | −0.03 | 0.03 | −0.09 to 0.03 | −1.06 |
| Group [FMA] × condition [t] | −0.00 | 0.03 | −0.06 to 0.05 | −0.17 |
| Group [FMA] × condition [st] | 0.00 | 0.03 | −0.05 to 0.06 | 0.12 |
| (time [post] × group [FMA]) × condition [t] | 0.06 | 0.04 | −0.02 to 0.14 | 1.45 |
| (time [post] × group [FMA]) × condition [st] | 0.01 | 0.04 | −0.07 to 0.10 | 0.34 |

[a]*t* = transparent inflection.
[b]*st* = semitransparent inflection.

pre- and posttest measurements (OR=−0.13, 95% CI [−0.17, −0.09]). In addition, word length had a main effect: longer words were recognized more slowly than shorter ones (OR = 0.04, 95% CI [0.03, 0.06]).

## Summary of the results on the VLD task

The aim of the VLD task was to measure how a three-week training with the FMA affects word recognition accuracy and speed in trained and untrained words in the three morphological conditions. Regarding accuracy, the results suggest that the FMA training improved recognition of the trained words in all conditions. However, there was no additional morphological effect, i.e. no differences in recognition accuracy improvement between conditions. For those words that were not trained in the FMA, instead, the analyses indicated an improvement in recognition accuracy for transparent inflections in the

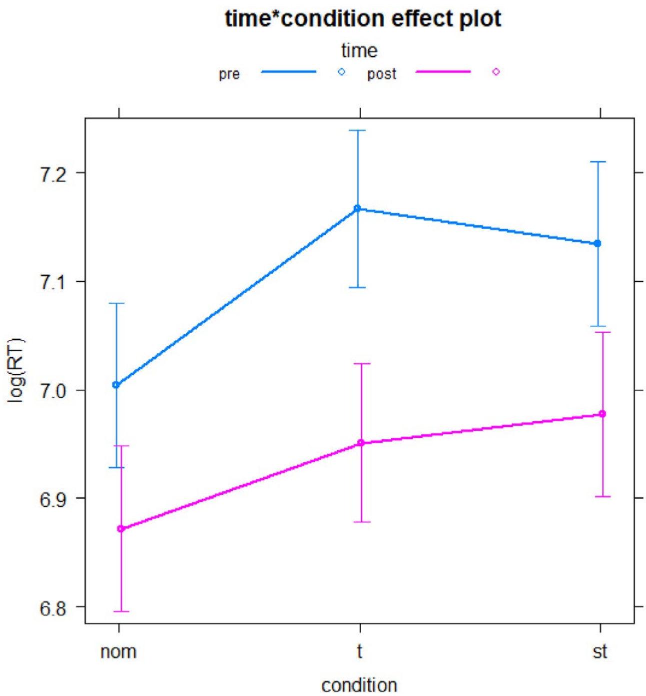


**Figure 9.** Interaction between time point and condition for reaction times in the trained items. The difference between nominatives and transparent inflections was smaller in the posttest than the pretest.

FMA group in comparison to nominatives, suggesting that the training helped to recognize words that had a transparent inflection. However, the significant interaction seemed primarily due to decreased accuracy rates for the nominative baseline in the posttests (i.e. the accuracy decreased from 71% to 65%). In RTs there were no significant between-group effects, only a main effect of proficiency and morphological condition. Overall, the results of the VLD task suggest that training with the FMA improved the recognition accuracy in those specific words that were trained in the FMA, both nominatives and inflections alike, but the effect did not generalize to other, untrained words.

## User experience

As a part of the post-survey, we collected data on the participants' user experience and perceived effectiveness of the FMA. We charted ten aspects regarding the user experience on a 1–5 Likert scale, such as app enjoyment and perceived efficacy of learning grammar. On average, the ratings varied from neutral to positive. For the FMA, the highest rating was for Comprehension of UI, which scored 4.4 on average ($SD = 0.9$). Moreover, the average rating for learning Finnish grammar and vocabulary was considered positive with a mean of 4.0 ($SD = 1.0$) and 4.3 ($SD = 0.8$), respectively. For the FMA, the lowest rating was reported for

**Table 10.** The Average User Experience Ratings and Standard Deviations on a Scale 1–5 for FMA and Control Group.

| User Experience | FMA | | Control | |
|---|---|---|---|---|
| | M | SD | M | SD |
| App Enjoyment | 3.5 | 1.3 | 3.1 | 1.2 |
| Comprehension of UI | 4.4 | 0.9 | 4.4 | 1.0 |
| Technical performance | 3.3 | 1.4 | 4.5 | 0.7 |
| Desire to continue | 3.3 | 1.4 | 3.1 | 1.2 |
| Ease of exercises | 3.5 | 0.9 | 3.2 | 1.0 |
| Motivation before | 4.3 | 0.9 | 4.3 | 0.8 |
| Motivation after | 3.2 | 1.2 | 3.0 | 1.2 |
| Learning Finnish grammar | 4.0 | 1.0 | 2.7 | 1.2 |
| Learning Finnish vocabulary | 4.3 | 0.8 | 4.0 | 0.8 |
| Learning of communication in Finnish | 3.0 | 1.2 | 3.6 | 1.1 |

learning to communicate ($M = 3.0$; $SD = 1.2$). The Kruskall-Wallis test showed that there was a statistically significant difference in the perceived efficacy between learning communication skills and grammar ($Z = -3.53$, $p = 0.001$) and between learning communication skills and vocabulary ($Z = -5.12$, $p < 0.001$), but not between grammar and vocabulary ($Z = -1.59$, $p = 0.112$). This suggests that, according to the participants, the FMA was perceived to be more useful for learning grammar and vocabulary than for learning communication skills. For both groups, motivation was slightly higher before the training period than in the end of it: for the FMA group, motivation decreased from $M = 4.3$ ($SD = 0.9$) to $M = 3.2$ ($SD = 1.2$), and for the control group, it decreased from $M = 4.3$ ($SD = 0.8$) to $M = 3.0$ ($SD = 1.2$). Both findings were statistically significant with Wilcoxon signed Rank test ($V = 450.5$, $p < 0.001$; and $V = 450.5$, $p < 0.001$ for the FMA and control group, respectively). The observed average ratings for all 10 topics are presented in Table 10. The questions for the ratings are presented in Appendix 2 (supplementary material).

## General discussion

The aim of this study was to develop a FMA and test its effectiveness for L2 morphological learning. We hypothesized that training with the FMA would improve Finnish L2 speakers' morphological skills in transparent and semi-transparent forms especially in the domain of production but also in reception. The target group ($n = 33$) practiced Finnish morphology with the FMA during a three-week period. The training consisted of producing inflected forms *via* answering questions like 'where is the cat?' or 'where did the cat come from/go to?'. The practiced morphological operations comprised transparent and semi-transparent inflections. The morphological improvements of the FMA group were compared to those of an active control group. The active control group used a general Finnish language learning application, Suomipassi ($n = 38$), that did not focus on morphology instruction. Pre- and posttests included MP and VLD tasks.

First, training with the FMA for three weeks had a positive effect on morphological production skills, for both transparent and semi-transparent inflections. That is, the difference between the accuracy rates of nominatives and both types of inflections were smaller in the posttest compared to the pretest, and more so for the FMA group than for the control group. Since the production task consisted of pseudowords, it can be concluded that the training effect reflects specifically the learning of morphological rules, not memorization of certain inflectional forms. This suggests that even though morphological complexity is a challenge for L2 learners, a structural and well-planned morphological training program as implemented in the current FMA can improve the production of morphologically complex forms in Finnish.

In the VLD, we aimed to find out whether recognition accuracy and recognition speed of morphologically complex words can be improved with FMA training. We tested this with words that were included in the training materials of the FMA group and with completely new words that were not presented during the training period. For reaction times, both groups were faster in the posttest than in the pretest, but there were neither statistically significant difference between groups for the trained words nor for the untrained words. For recognition accuracy, the FMA group improved significantly on the trained words, i.e. those specific words they had practiced during the 3-week training period, unlike the control group. This improvement was independent of morphological structure, as it held for both the nominative and the inflected forms. For the untrained words, there was no clear improvement for any of the conditions in accuracies. Thus, in word recognition, morphological training did not lead to better performance of untrained inflected words. The discrepancy between MP and VLD results may be attributed to the fact that unlike the MP, the VLD task was speeded, and perhaps did not provide the participants enough time to conduct still-not-automatized morphological parsing. In addition, the training in the FMA focused on production and the training exercises were largely similar to the MP tasks. Taken together, the results suggest that the FMA enhances rule-based learning for production and also improves word recognition for those specific words that have been trained in the application.

In both tasks, there was a main effect of time point of testing: both groups performed better in the posttests than in the pretests. In the MP task this was reflected by higher accuracy in the posttest for both groups, in VLD in shorter reaction times. This is likely caused by a test-retest effect and/or general improvement of Finnish skills over time.

The best performance in both tasks was elicited by nominatives, followed by transparent inflections, and then semi-transparent inflections. This is an expected finding, as more complex morphological rules must be applied to process semi-transparent word forms than transparent or nominative word forms.

While the primary goal of this study was to assess the objective pre-post learning outcomes, we also collected data on the participants' user experience and the perceived efficacy of the application. For the FMA, the overall user ratings regarding e.g. app enjoyment, UI comprehension, and technical performance averaged from neutral to positive. Regarding the efficacy of learning grammar, vocabulary, and communication skills, the participants reported that the application was perceived more effective for grammar and vocabulary learning than for learning general communication skills. This is logical as the application did not consist of communication exercises, but only exercises training Finnish morphosyntax with Finnish words.

Past studies on grammar in MALL have mostly concentrated on enhancing grammatical learning in languages with relatively simple morphology (Castañeda & Cho, 2016; Chu et al., 2019; Ghorbani et al., 2020; Li & Hegelheimer, 2013; Rachels & Rockinson-Szapkiw, 2018; Wang et al., 2021). There is also a large body of evidence on improving grammar proficiency *via* CALL methods (AbuSeileek, 2009; Collentine & Collentine, 2015; Kılıçkaya, 2015; Kruk, 2015; Lucas, 2020; Naba'h et al., 2009; Penning de Vries et al., 2020; Reynolds & Kao, 2021, Torlaković, & Deugo, 2004), but only few studies extend to morphological processing (Eryiğit et al., 2023). To our knowledge, no previous randomized controlled trial design study has been used to test a mobile application designed to train learning of morphology. This study shows that also morphological rules can be effectively trained with MALL, even in morphologically rich languages. In addition, the findings support that the theory of exemplar-based learning applies not only to traditional teaching but also in the context of MALL.

## Limitations, caveats, and future directions

The overall motivation level of the participants slightly decreased during the intervention, indicating that a mildly positive motivation in the beginning diminished to a neutral level during the training. However, the motivation ratings for the control app users decreased in a similar manner. For this reason, we believe that the decline in motivation is not specifically associated with features of the FMA but more likely linked to general user fatigue, which may occur with repetitive mobile exercises over time, as observed in previous studies (Hanson & Brown, 2020).

It should also be noted that a large part of the data collection (2019–2022) took place during the COVID-19 pandemic. Research suggests that this pandemic may have elicited negative implications for mobile learning attitudes, possibly attributed to the increased use of new applications and software during remote studies (Stockwell & Wang, 2023). It is possible that this also negatively affected our user experience ratings and attrition rate, with 34 out of 120 participants (28%) withdrawing during the training period, and an additional 13 participants not being active enough to be included in the data analysis. The overall attrition rate was 39% and might have introduced selection bias (Wolke et al., 2009). Nevertheless, it is worth noting that the drop-out rate was about equal for the FMA and the control group. In general, challenges in participant engagement are not completely uncommon in the field of CALL and MALL (Botero & Questier, 2016; Nielson, 2011). According to Botero and Questier (2016), a high attrition rate may be associated with extra-curricular learning, which can be perceived as unmotivating without adequate external support. For a more successful use of CALL and MALL applications, recommendations include curricular integration, mentoring, and scaffolding (for similar findings, see Nielson, 2011).

The foremost reason for using active controls that trained with another Finnish language learning program was to control for various non-specific effects (engagement with an app, participant expectancies and demand characteristics) that could have threatened the internal validity of our intervention experiment. In our study, we used a single-blind design, emphasizing the importance of participants being unaware of whether they belonged to the experimental or the active control group. To prevent the participants from realizing this, we selected a control application, *Suomipassi*, designed to improve general Finnish language skills without specifically targeting morphology, which was the focus of our study. It is important to note that our intention was not to directly compare the performance of these two applications in terms of morphological learning. Instead, we intended to assure that a general approach to learning the Finnish language *via* MALL does not lead to the same kind of specific improvements in morphological knowledge as the FMA does. Note though that the FMA application only covers a specific subdomain of morphology and utilizes a specific approach to learn Finnish morphology. It is evident that future research and applications should extend the scope and test alternative methods for learning Finnish morphology. As there are no other Finnish morphology applications available, we hope the current application will serve as an initial step and inspiration for future efforts.

It is also worth noting that the improvement in morphological production skills may not be exclusively attributed to the use of MALL, but that similar improvements could have been achieved through more traditional methods of morphology training as well. Exploring the differences in learning outcomes between content-wise identical MALL and traditional methods presents an intriguing issue both in the field of MALL and morphological learning and could be explored *via* an additional control group in the study design. However, we leave further exploration of this issue for future studies. What is important for now is that self-training through an app specifically targeting morphology can effectively enhance morphological skills in Finnish. This finding is particularly significant given the widespread use of mobile devices nowadays.

According to previous studies, various moderator variables, such as age or mobile use acceptance can affect the results in MALL (Sung et al., 2015). In the future, it would therefore be interesting to assess how such moderator variables affect the usefulness of the FMA, an effort that would require access to a larger sample size. Finally, although the focus of this study was to improve the use of locative cases and consonant gradation in Finnish, it is worth emphasizing that there are other morphological domains (derivations, compounding) and morphophonological operations (e.g. backward consonant gradation, vowel alterations) in Finnish that could greatly benefit from additional practice. It is good to notice, however, that the FMA is currently tested only for a relatively small subset of morphological operations and it concerns a concrete, spatially grounded aspect of their usage. The more abstract use of inflections, e.g. 'the car broke down **in seconds**', was not assessed, and it is theoretically possible that the impact is limited only to semantically concrete operations. However, given the high regularity of Finnish morphology we anticipate that the same principles, i.e. adaptive spaced repetition and exemplar-based learning, would aid the learning of other morphological phenomena as well. The goal is to upgrade the FMA in the future by including such a wider range of morphology.

## Summary and conclusions

This study investigated whether three weeks of MALL training with the FMA improved morphological processing of adult beginner level learners in comparison to an active control group. Pre- and posttests were conducted in the form of MP and VLD. The results showed that the FMA improved the production of both transparent and semi-transparent inflectional forms. When it comes to word recognition, accuracy improved more in the FMA group than in the control group, but only for those specific words that were trained during the training period, with no difference in improvement between the morphological conditions. It can be concluded that the FMA supports adult beginner level learners' morphological production skills, and that this improvement is rule-based, as it applies to novel pseudowords, for which no lexical representation can exist. In word recognition, the FMA enhances the vocabulary knowledge for those specific words that were trained during the training period, but there was no enhanced effect for the inflectional conditions. A stronger morphological effect in production than in recognition was expected, as the FMA explicitly trained productive skills. These measured learning effects are supported by the fact that the perceived efficacy for grammar and vocabulary learning was also rated positively in the post-survey feedback.

To conclude, the present study suggests that MALL can improve morphological production skills in the context of a morphologically complex language like Finnish. The FMA is thus a valuable tool that can be utilized to improve morphological learning. It can be used to study Finnish independently, but also be implemented into Finnish as L2 study curriculum. Given the major challenge to master the vast Finnish morphology and morphosyntax, regular use of an application like the FMA can significantly accelerate the acquisition of the language.

## Notes

1. NB. The term single-blind is used to distinguish our design from the double-blind design, where neither the participants nor the researchers are aware of the participants' study group.
2. However, the return rate was too low to use hand-written logs for assessing training times.


## Acknowledgements

We thank Tiina Manninen and Eveliina Sipakko for assistance in data collection, Satu Savo for help in data coding, and Ali Moazami Goodarzi for assistance in statistical analyses. The user interface of the FMA application was designed by UX designer Tiina Viljakainen (Taiste, Finland). Part ofthe photographs in the FMA application were taken by photography students from Turku University of Applied Sciences, and part of them were retrieved from Unsplash (2023). We would like to thank all the Finnish L2 teachers at various institutions in Turku who kindly allowed us to visit their classes for participant recruitment. Additionally, we would like to express our gratitude to the creators and owners of Suomipassi (Centre for Language and Communication Studies, University of Turku) for letting us use it as a control application in this study.


## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Ethics statement

This study was carried out with written informed consent from all participants. The protocol was approved by the Institutional Review Board of the Åbo Akademi University Psychology Department. A pre-registration was conducted *via* AsPredicted (https://aspredicted.org/blind.php?x=GVL_1FJ).

## Funding

This study was funded by Åbo Akademi University Minority Research Profile. Prof. M. Laine was supported by the Academy of Finland [grant #323251].

## ORCID

Rosa Salmela http://orcid.org/0000-0002-4781-8649
Minna Lehtonen http://orcid.org/0000-0002-6137-8854
Matti Laine http://orcid.org/0000-0003-4529-4965
Raymond Bertram http://orcid.org/0000-0002-2709-3644

## References

AbuSeileek, A. F. (2009). The effect of using an online-based course on the learning of grammar inductively and deductively. *ReCALL*, *21*(3), 319–336. https://doi.org/10.1017/S095834400999005X

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Balota, D. A., Duchek, J. M., Sergent-Marshall, S. D., & Roediger, H. L.III (2006). Does expanded retrieval produce benefits over equal-interval spacing? Explorations of spacing effects in healthy aging and early stage Alzheimer's disease. *Psychology and Aging*, *21*, 19–31. https://doi.org/10.1037/0882-7974.21.1.19

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Botero, G. G., & Questier, F. (2016). What students think and what they actually do in a mobile assisted language learning context: New insights for self-directed language learning in higher education. CALL communities and culture–short papers from EUROCALL, 150–154. https://research-publishing.net/publication/chapters/978-1-908416-44-5/553.pdf.

Castañeda, D. A., & Cho, M. (2016). Use of a game-like application on a mobile device to improve accuracy in conjugating Spanish verbs. *Computer Assisted Language Learning*, *29*(7), 1195–1204. https://doi.org/10.1080/09588221.2016.1197950

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science*, *19*, 1095–1102. https://doi.org/10.1037/0033-2909.132.3.354

Cerezo, L., Baralt, M., Suh, B.-R., & Leow, R. P. (2014). Does the medium really matter in L2 development? The validity of CALL research designs. *Computer Assisted Language Learning*, *27*(4), 294–310. https://doi.org/10.1080/09588221.2013.839569

Chen, Z., Chen, W., Jia, J., & An, H. (2020). The effects of using mobile devices on language learning: A meta-analysis. *Education Tech Research Dev*, *68*(4), 1769–1789. https://doi.org/10.1007/s11423-020-09801-5

Chu, H.-C., Wang, C.-C., & Wang, L. (2019). Impacts of concept map-based collaborative mobile gaming on English Grammar Learning Performance and Behaviors. *Journal of Educational Technology & Society*, *22*(2), 86–100. https://www.jstor.org/stable/26819619

Collentine, J., & Collentine, K. (2015). Input and output grammar instruction in tutorial CALL with a complex grammatical structure. *CALICO Journal*, *32*(2), 273–298. https://www.jstor.org/stable/calicojournal.32.2.273 https://doi.org/10.1558/cj.v32i2.24548

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology*. Dover. (Original work published 1885). https://doi.org/10.5214/ans.0972.7531.200408

Ellis, N. C., & Wulff, J. (2015). Usage-based approaches to SLA,(w:) VanPatten B., Williams J.(red.), Theories in second language acquisition.

Eryigit. (2022). Automated gamification method for learning morphologically rich languages, WO/2023/063907. https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2023063907&_cid=P21-LT5RHG-05799-1.

Eryiğit, G., Bektaş, F., Ali, U., & Dereli, B. (2023). Gamification of complex morphology learning: The case of Turkish. *Computer Assisted Language Learning*, *36*(8), 1421–1449. https://doi.org/10.1080/09588221.2021.1996396

Eskildsen, S. (2009). Constructing another language – Usage-based linguistics in second language acquisition. *Applied Linguistics*, *30*(3), 37. https://doi.org/10.1093/applin/amn037

Ghorbani, N., Ebadi, S., & Popescu, M. (2020). Exploring learners' grammatical development in mobile assisted language learning. *Cogent Education*, *7*(1), 99. https://doi.org/10.1080/2331186X.2019.1704599

Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the "Natural Order of L2 Morpheme Acquisition" in English: A meta-analysis of multiple determinants. *Language Learning*, *51*, 1–50. https://doi.org/10.1111/1467-9922.00147

Green, J., Weston, T., Wiseheart, M., & Rosenbaum, R. (2014). Long-term spacing effect benefits in developmental amnesia: Case experiments in rehabilitation. *Neuropsychology*, *28*, 685–694. https://doi.org/10.1037/neu0000070

Grgurović, M., Chapelle, C., & Shelley, M. (2013). A meta-analysis of effectiveness studies on computer technology-supported language learning. *ReCALL*, *25*(2), 165–198. https://doi.org/10.1017/S0958344013000013

Hakulinen, A., Vilkuna, M., Korhonen, R., Koivisto, V., Heinonen, T. R., & Alho, I. (2004). *VISK = Iso suomen kielioppi* [The Comprehensive Grammar of Finnish]. Suomalaisen Kirjallisuuden Seura. Online version. Retrieved April 03, 2023, from http://scripta.kotus.fi/visk.

Hanson, A., & Brown, C. M. (2020). Enhancing L2 learning through a mobile assisted spaced-repetition tool: An effective but bitter pill? *Computer Assisted Language Learning*, *33*(1–2), 133–155. https://doi.org/10.1080/09588221.2018.1552975

Hopp, H. (2020). Morphosyntactic adaptation in adult L2 processing: Exposure and the processing of case and tense violations. *Applied Psycholinguistics*, *41*(3), 627–656. https://doi.org/10.1017/S0142716420000119

Javadi, Y., & Kazemirad, F. (2020). Usage-based approaches to second language acquisition: Cognitive and social aspects. *Journal of Language Teaching and Research*, *11*(3), 473–479. https://doi.org/10.17507/jltr.1103.16

Jensen, I. N., Slabakova, R., Westergaard, M., & Lundquist, B. (2020). The Bottleneck Hypothesis in L2 acquisition: L1 Norwegian learners' knowledge of syntax and morphology in L2 English. *Second Language Research*, *36*(1), 3–29. https://doi.org/10.1177/0267658318825067

Kapler, I. V., Weston, T., & Wiseheart, M. (2015). Spacing in a simulated undergraduate classroom: Long-term benefits for factual and higher-level learning. *Learning and Instruction*, *36*, 38–45. https://doi.org/10.1016/j.learninstruc.2014.11.001

Karlsson, F. (2002). *Finnish: An essential grammar*. Routledge.

Kılıçkaya, F. (2015). Computer-based grammar instruction in an EFL context: Improving the effectiveness of teaching adverbial clauses. *Computer Assisted Language Learning*, *28*(4), 325–340. https://doi.org/10.1080/09588221.2013.818563

Kim, A. S. N., Wong-Kee-You, A. M. B., Wiseheart, M., & Rosenbaum, R. S. (2019). The spacing effect stands up to big data. *Behavior Research Methods*, *51*(4), 1485–1497. https://doi.org/10.3758/s13428-018-1184-7

Kim, S. K., & Webb, S. (2022). The effects of spaced practice on second language learning: A meta-analysis. *Language Learning*, *72*, 269–319. https://doi.org/10.1111/lang.12479
Kruk, M. (2015). Practicing the English present simple tense in Active Worlds. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, *5*(4), 52–65. https://doi.org/10.4018/IJCALLT.2015100104
Laine, M., & Virtanen, P. (1999). *WordMill lexical search program*. Center for Cognitive Neuroscience, University of Turku.
Li, Z., & Hegelheimer, V. (2013). Mobile-assisted grammar exercises: Effects on self-editing in L2 writing. *Language Learning & Technology*, *17*, 135–156.
Lin, J., & Lin, H. (2019). Mobile-assisted ESL/EFL vocabulary learning: A systematic review and meta-analysis. *Computer Assisted Language Learning*, *32*(8), 878–919. https://doi.org/10.1080/09588221.2018.1541359
Lucas, M. (2020). Cross-linguistic influence and plural marking: Enhancing error recognition and written accuracy among Japanese EFL learners through web-based awareness-raising. *Language Awareness*, *29*(2), 155–177. https://doi.org/10.1080/09658416.2020.1737096
Martin, M. (1995). Matkailijana Morfologiassa [Travelling in Morphology]. In E. Aalto, & M. Suni, (Ed.), *Kohdekielenä suomi. Näkökulmia opetukseen* [Finnish as a Target Language: Aspects of Teaching] (pp. 51–71). Korkeakoulujen kielikeskuksen selosteita n: O 1 Jyväskylän yliopisto.
Metcalfe, J., & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology*, *132*(4), 530–542. https://doi.org/10.1037/0096-3445.132.4.530
Metzler-Baddeley, C., & Baddeley, R. J. (2009). Does adaptive training work? *Applied Cognitive Psychology*, *23*, 254–266. https://doi.org/10.1002/acp.1454
Naba'h, A. A., Hussain, J., Al-Omari, A., & Shdeifat, S. (2009). The effect of computer assisted language learning in teaching English grammar on the achievement of secondary students in Jordan. *The International Arab Journal of Information Technology*, *6*(4), 431–439.
Nielson, K. B. (2011). Self-study with language learning software in the workplace: What happens? *Language Learning & Technology*, *15*(3), 110–129.
Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*, 195–203. https://doi.org/10.3758/s13428-018-01193-y
Penning de Vries, B. W., Cucchiarini, C., Strik, H., & Van Hout, R. (2020). Spoken grammar practice in CALL: The effect of corrective feedback and education level in adult L2 learning. *Language Teaching Research*, *24*(5), 714–735. https://doi.org/10.1177/1362168818819027
Rachels, J., & Rockinson-Szapkiw, A. J. (2018). The effects of a mobile gamification app on elementary students' Spanish achievement and self-efficacy. *Computer Assisted Language Learning*, *31*(1-2), 72–89. https://doi.org/10.1080/09588221.2017.1382536
R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Reynolds, B. L., & Kao, C. W. (2021). The effects of digital game-based instruction, teacher instruction, and direct focused written corrective feedback on the grammatical accuracy of English articles. *Computer Assisted Language Learning*, *34*(4), 462–482. https://doi.org/10.1080/09588221.2019.1617747
Roediger, H. L., III., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x
Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2011). Spacing effects in real-world classroom vocabulary learning. *Applied Cognitive Psychology*, *25*, 763–767. https://doi.org/10.1002/acp.1747

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology*, *26*, 204–221. https://doi.org/10.1037/0278-7393.26.1.204

Stockwell, G., & Wang, Y. (2023). Exploring the challenges of technology in language teaching in the aftermath of the pandemic. *RELC Journal*, *54*(2), 474–482. https://doi.org/10.1177/00336882231168438

Sung, Y.-T., Chang, K.-E., & Yang, J.-M. (2015). How effective are mobile devices for language learning. *Educational Research Review*, *16*, 68–84. https://doi.org/10.1016/j.edurev.2015.09.001

Suomipassi. (2020). = https://suomipassi.utu.fi/.

Tani, H. (2008). Kielo. Materiaalia aikuisten maahanmuuttajien suomen kielen taidon kartoitukseen ja kehityksen seuraamiseen. http://www.oph.fi/download/46536_kielo.

Tomasello, M. (2009). The usage-based theory of language acquisition. In *The Cambridge handbook of child language* (pp. 69–87) Cambridge Univ. Press.

Torlaković, E., & Deugo, D. (2004). Application of a CALL system in the acquisition of adverbs in English. *Computer Assisted Language Learning*, *17*(2), 203–235. https://doi.org/10.1080/0958822042000334244

Unsplash. (2023). = license description. Retrieved March 10, 2023, from https://unsplash.com/license.

Wang, X., Chen, J., & Zhang, T. (2021). Facilitating English Grammar learning by a personalized mobile-assisted system with a self-regulated learning mechanism. *Frontiers in Psychology*, *12*, 624430. https://doi.org/10.3389/fpsyg.2021.624430

Wolke, D., Waylen, A., Samara, M., Steer, C., Goodman, R., Ford, T., & Lamberts, K. (2009). Selective drop-out in longitudinal studies and non-biased prediction of behaviour disorders. *The British Journal of Psychiatry*, *195*(3), 249–256. https://doi.org/10.1192/bjp.bp.108.053751