Language Learning & Technology February 2017, Volume 21, Number 1
http://llt.msu.edu/issues/february2017/cornillieetal.pdf pp. 121–145

EXAMINING FOCUSED L2 PRACTICE: FROM IN VITRO TO IN VIVO

Frederik Cornillie, KU Leuven, imec
Wim Van Den Noortgate, KU Leuven, imec
Kris Van den Branden, KU Leuven
Piet Desmet, KU Leuven, imec

Behaviour-tracking technology has been used for decades in SLA research on focused practice with an eye toward elucidating the nature of L2 automatization (e.g. DeKeyser, 1997; Robinson, 1997). This involves longitudinally capturing learners' judgments or linguistic production along with their response times in order to investigate how specific skills become automatic over time. However, previous research in this area has been conducted mostly in laboratories (i.e., in vitro), sometimes with artificial languages, thereby compromising ecological validity of the findings.

Building on this work, this article reports on a one-month study in which learners' (*N* = 126) behaviour was tracked while they practised two constructions of English grammar (varying in complexity) using mini-games that involved some time pressure and were embedded in meaning-focused reading and discussion activities in class. Feedback was randomly varied between participants. Multilevel statistical analyses of accuracy and response time suggest that practice helped to develop automaticity, and that rule complexity and metalinguistic feedback played a role. The methodological innovation of this study consists of the application of in vitro experimental research techniques in in vivo L2 learning contexts and of the use of statistical mixed effects models to account for the complexity of real-life tracking data.

Language(s) Learned in this Study: English

Keywords: Behaviour-tracking Technology, Computer-assisted Language Learning, Second Language Acquisition, Grammar, Syntax

APA Citation: Cornillie, F., Van Den Noortgate, W., Van den Branden, K., & Desmet, P. (2017). Examining focused L2 practice: From in vitro to in vivo. *Language Learning & Technology*, *21*(1), 121–145. Retrieved from http://llt.msu.edu/issues/february2017/cornillieetal.pdf

Received: December 3, 2015; **Accepted:** July 20, 2016; **Published:** February 1, 2017

INTRODUCTION

When embedded within meaning-oriented second language (L2) use, focused practice can be a focus-on-form technique that helps to automatize knowledge about lexical, morphosyntactic, phonological, and pragmatic aspects of the L2 in procedural memory. This could result in effortless and target-like performance in the L2 while freeing up attentional resources for higher-order skills during complex tasks (Segalowitz & Hulstijn, 2005). Hence, research on meaningful focused practice for L2 learning is relevant both theoretically and pedagogically. However, experimental research into the effects of focused practice on the development of L2 automaticity faces a number of methodological challenges. First, in order to carefully manipulate features of the learning environment and measure learner performance accurately, the available studies have typically taken place in laboratory settings (i.e., in vitro settings), sometimes with artificial languages (e.g. DeKeyser, 1997; Robinson, 1997). In order not to compromise ecological validity, it is essential that this strand of research be replicated in real L2 learning settings such as classrooms (i.e., in vivo settings). When moving research on focused L2 practice out of the lab, however, a second challenge arises, namely to provide L2 learners with feedback that is consistent and individualized. While this is possible in laboratory research, it is impractical or undesirable—if not infeasible—in classroom contexts. A third challenge for both pedagogy and ecologically valid research alike is to move beyond mechanical L2 practice, by coupling automatization through repetition with meaningful information processing in highly contextualized L2 use (DeKeyser, 1998; Segalowitz & Hulstijn, 2005). This move entails a challenge for instructional design. In sum, the mission for second language acquisition (SLA) research on focused L2 practice is to bridge controlled experimental trialling and L2 learning in real classrooms guided by state-of-the-art pedagogical principles.

Claims have been made that computer-assisted language learning (CALL) holds great promise for the future of research on practice, as it allows for massive and fine-grained data collection on L2 performance in longitudinal experimental research designs, potentially in externally valid ways (DeKeyser, 2007a). Moreover, in the human–computer interactions of tutorial CALL (Heift & Schulze, 2015), feedback can be delivered much more consistently than is possible in face-to-face settings, and ad infinitum. So, research on practice through tutorial technology might offer an answer to the three methodological challenges outlined above: careful manipulation and control of the conditions of practice in ecologically valid settings, the delivery of consistent feedback, and—pending instructional designs that couple form focus with meaning focus—the embedding of focused practice in meaningful L2 use.

This article reports on an empirical study on focused L2 grammar practice with CALL materials in real contexts that was aimed at overcoming the three methodological challenges listed above. It details the design of the practice environment and presents results of analyses of process data from one month of practice (accuracy rates and response times), which were captured by means of behaviour-tracking technology. The current study is part of a larger research project that also investigated the effects of practice on transfer to other L2 tasks as measured by pre- and post-tests. The latter results are presented briefly in Cornillie, Van den Branden, and Desmet (2015). The current manuscript presents a more detailed description of the experimental procedure used in this research project and reports on in-depth analyses of the practice data.

BACKGROUND RESEARCH

Defining the Scope of Focused L2 Practice

The term *focused practice* refers to activities in a L2 that focus on specific linguistic constructions and involve a great deal of repetition, feedback, and often time pressure with the goal of developing declarative knowledge about these constructions as well as skills in the L2 (DeKeyser, 2007b). It is frequently associated exclusively with narrow forms of drilling, but in fact covers a wide range of activities that involve different kinds of L2 processing. Activities for focused practice vary, first of all, with respect to the concurrence (or dissociation) of form and meaning processing. In this respect, focused practice may be mechanical (when learners do not need to process the meaning of the utterance to complete the task), it may be meaningful (when the task requires them to comprehend the L2 on both a structural and semantic level), or it may be communicative (when learners convey personal meaning rather than reproduce prefabricated and predictable responses; see DeKeyser, 1998; Paulston & Bruder, 1976). Further, while the term *practice* is sometimes reserved for output practice, focused practice can involve receptive or productive skills. Next, practice can be oral or written and can focus on various formal aspects of the L2 (phonological, morphological, syntactic, or lexical form). Thus, focused practice can take many forms, which are likely to induce different cognitive processes. We define *focused practice* in its broadest possible sense, with the key feature that some repetition of linguistically related constructions is inherent in the activity.

Even though activities for focused practice are quite varied, critiques typically focus on one specific manifestation, namely productive oral pattern drills that are mechanical in nature and involve sentences that share only some grammatical pattern rather than a meaningful context. As is well known, such drills were popularized by audiolingualism and have been shown to be ineffective and sometimes even disadvantageous for the development of communicative L2 competence—that is, the ability to express personal meaning fluently and accurately (for a review see Wong & VanPatten, 2003). While the effects of mechanical practice are clear, little is known about the usefulness of meaningful and communicative focused practice for L2 learning.

Framing Focused L2 Practice: Skill Acquisition Theory

A theoretical framework aimed at explaining the phenomena associated with focused practice, including the role of feedback, is skill acquisition theory (SAT) (DeKeyser, 2007b). SAT is derived from a general theory of human development known as adaptive control of thought (Anderson, 1992) and posits that the development of a skill is the gradual process of moving through a series of stages that differ with respect to the effort used and type of knowledge relied on to perform the skill. The typical trajectory in L2 skill learning comprises three stages. Initially, through explicit instruction, declarative knowledge is developed. Declarative knowledge is available to awareness, can be verbalized, and typically comprises rule-based knowledge about grammatical constructions. This stage is followed by a first phase of practice, in which declarative knowledge is active in working memory and applied consciously and with great effort to concrete L2 items. This practice phase assumedly results in the development of *procedures*, or condition-action pairs which encode the rules in behaviour and comprise knowledge on what needs to be done under specific circumstances. The main advantage of proceduralization is efficiency of retrieval. Knowledge about a particular L2 construction becomes available as a ready-made chunk in memory, ready to be called upon when the conditions for its use reoccur, and quickly retrievable. The final stage consists of automatizing procedural knowledge through continued practice, which requires increasingly fewer conscious cognitive resources, and sometimes even leads to loss of initial declarative knowledge. In sum, L2 learning is seen as a transition from controlled to automatic L2 processing.

SAT posits that automatization through practice is manifested in learner behaviour by a decrease in response times and error rates. Characteristic of automatization is that both these evolutions are non-linear. They follow a curve that drops steeply in the beginning and decreases more gradually in later stages to finally flatten out, and which can be mathematically expressed by means of a power law—hence the term *power law of practice* (DeKeyser, 2007b).

While automatization through focused practice is thought to culminate in the development of accurate, quickly retrievable, and robust knowledge that is functionally equivalent to implicit L2 knowledge (developed through implicit learning), declarative knowledge continues to play a role in later phases. When automatic processing fails, declarative knowledge may need to be recalled in working memory to provide additional conscious support that may interact with automatic processes, for instance through corrective feedback (CF). For some learners at some stages of development, limited CF in the form of knowledge-of-results may suffice, such as signals of communication breakdown in face-to-face settings (i.e., requests for clarification; e.g. Lyster & Ranta, 1997), or disconfirmations of grammatical accuracy in computerized practice (Schulze, 2003). Such prompts may remind these learners of initially taught declarative information. For other learners, and perhaps for more complex grammar rules, declarative information may need to be repeated, rephrased, or elaborated upon, calling for more elaborate CF that comprises metalinguistic clues or explanations.

Process-focused Intervention Studies on L2 Grammar Practice

Although empirical interest in automaticity in L2 grammar learning has intensified over the past decade (e.g., De Jong & Perfetti, 2011; Jiang, 2007; Lim & Godfroid, 2015; Rodgers, 2011; Serafini, 2013), there are few process-focused intervention studies on how L2 automatization can be stimulated through practice. Such studies take a longitudinal and fine-grained approach: rather than comparing the difference between scores on pre- and post-tests, they measure and timestamp responses to individual practice items over a certain period of time. They may also manipulate features of task design (such as the type of instruction or feedback) to investigate how these variables affect learner behaviour.

In an attempt to empirically test SAT in L2 learning, DeKeyser (1997) conducted a computerized experiment on the effects of meaningful practice on automatization of grammar rules for comprehension and production skills in an artificial yet natural language-like L2. Over a period of 8 weeks, both skills developed through gradual decreases in error rate and response time, which followed a power law. Consistent with another prediction of SAT, the effects of practice were skill-specific, as comprehension practice did not transfer well to production skills and vice-versa.

From an exemplar-based (rather than rule-based) perspective on automatization, Robinson (1997) investigated the learning of the constraints on dative alternation in English L2 under four conditions that varied with respect to the presence and implementation of focus-on-form. No evidence was found of automatization in the form of faster response times to grammaticality judgments as a function of increasing frequency of presentation. This may be due to the relatively short length of practice (30 minutes). The study did find that learners who were provided with declarative information prior to practice were significantly more accurate than the uninstructed learners in judging the grammaticality of unencountered sentences.

De Jong and Perfetti (2011) examined the effects of speech repetition on oral fluency development in English L2 classrooms. They found that students who repeated their speeches became more fluent than students who spoke about different topics and concluded that speech repetition had helped proceduralization of linguistic knowledge, evidencing a restructuring of underlying cognitive mechanisms.

An outstanding theoretical issue that has not yet received an empirical response but has important consequences for the teaching of grammar concerns the question whether the effects of focused practice with feedback depend on the complexity of rules. One position is that simple rules make the best candidates for explicit instruction (DeKeyser, 1998). In contrast, Hulstijn and de Graaff (1994) argue that the advantage of explicit instruction is greater for complex rules than for simple rules, because learners are likely to pick up simple formal phenomena spontaneously (i.e., through implicit learning). Further, Hulstijn (2007) proposes that in the foreign language classroom, grammar rules need to be as short and simple as possible, as human beings can only handle a limited amount of declarative knowledge at a time. The findings of a meta-analysis by Spada and Tomita (2010) do not support the hypothesis that type of instruction (explicit vs. implicit) interacts with the cognitive complexity of grammatical structures. So far, no process-focused intervention studies on L2 practice have evaluated the impact of rule complexity.

Taking Intervention Studies on L2 Grammar Practice from in vitro to in vivo: A Place for CALL?

With the exception of De Jong and Perfetti (2011), process-focused intervention studies targeting the effects of L2 grammar practice on automatization have taken place in laboratory settings. Robinson (1997) pointed at a major weakness in this research. On the one hand, he noted that the use of artificial language and decontextualized stimuli rather than extended text is necessary in order to meticulously measure the effects of practice on L2 development. On the other hand, he argued that these methodological choices are exactly the ones that may interfere with learners' interest or may cause shallow processing, hindering the transferability of such research to real L2 learning. Similarly, DeKeyser (1997, 1998) has insisted on several occasions that practice which is not genuinely meaningful does not qualify as productive for L2 development, and that the findings of lab-based studies should therefore not be used as recommendations for L2 pedagogy.

Further, DeKeyser (1998) has made pleas for research that "combines the degree of control of a psycholinguistic experiment with the validity of research on real second language learning, and [...] that, on top of that, takes a process, that is, a developmental, longitudinal, perspective" (p. 60). Although the phenomenon of L2 automatization can be examined in cross-sectional studies, its developmental character aligns better with within-subject designs, in particular longitudinal ones (for comments, see Lim & Godfroid, 2015; Rodgers, 2011). These challenges can be tackled through CALL, as technology can vary the conditions of practice (e.g., feedback) and measure learner behaviour longitudinally, potentially in meaningful language tasks (see concluding remarks in DeKeyser, 2007a).

The methodological aim of the current study was to combine controlled experimental trialling with ecologically valid, meaning-oriented L2 practice supported by CALL materials and to evaluate whether tracking behaviour technology and analytics could inform SLA-focused research in this context. We approached this aim by means of a mystery text, read and discussed in class for its meaning, of which the content served as the basis for focused grammar practice with mini-games in class and in learners' homes. The system manipulated the type of CF and the complexity of the grammatical constructions, and logged learners' practice behaviour in order to measure their performance and control for the effect of time on task in the analyses. In other words, the study implemented features of laboratory-based, in vitro research on focused L2 practice in the in vivo contexts of real L2 learning in classrooms with meaning-focused practice materials. The research questions were as follows:

1. What are the effects of focused practice on automatization of grammatical knowledge in the ecologically valid setting of language classrooms and home contexts?
2. How is practice affected by metalinguistic information provided in feedback as well as by rule complexity?

METHOD

Participants and Research Context

The data were collected in the spring of 2014. The participants were 126 Dutch-speaking learners of English from eight classes in the fifth and sixth year of general secondary education in Flanders, Belgium. Typically, these learners are between 16 and 18 years old, and have completed three to four years of formal instruction in English, which is meant to correspond roughly to the intermediate level (B1) of the Common European Framework of Reference for Languages.

The study took place in an in vivo context: the data were collected through web-based behaviour-tracking technology while learners practised English grammar on the computer both in class and at home. In contrast with typical experiments in the psycholinguistic lab, the learners were left largely to their own devices as for the time they invested in the study. Although the first author and participating teachers stimulated practice as much as possible, no attempts were made to obtain an equal or minimal number of practice sessions for all participants. Moreover, in contrast with how participants are usually motivated in laboratory studies, the learners were not paid or rewarded for their participation. In some classes, however, learners received grades from their teachers on the basis of their performance in practice and, in one class, on the basis of their result on a reading comprehension test related to the materials. This was done with a view to making the learners take the study seriously. Other than this, it may be argued that the learners' motivation for practice was largely self-determined—in any case not markedly different from the motivation involved in typical coursework.

Materials and Procedure for Instruction and Practice

Grammatical Constructions

Instruction and practice focused on *quantifiers* in English (QNT) and *verbs with two objects* (V2O; see Table 1), known more commonly as *dative alternation* (Carroll & Swain, 1993). These grammatical constructions were chosen for three reasons: they are frequent in English; the constraints on their use do not apply in Dutch; and ungrammatical instances are frequent in the interlanguage of Dutch-speaking learners, particularly for QNT (Tops, Dekeyser, Devriendt, & Geukens, 2001). Furthermore, the two constructions differ in complexity, which could determine how learners may benefit from instruction and practice. Additionally, acquisition of the principles underlying V2O is considered difficult for L1 and L2 learners alike (e.g., Gropen, Pinker, Hollander, Goldberg, & Wilson, 1989) and, given the high productivity of V2O in English use, it is a typical example of a learnability issue (i.e., it is puzzling that L1 learners demonstrate knowledge of constructions to which they have not been exposed in the input). Hence, for L2 learners in particular, acquisition of V2O is considered to require negative evidence—for instance, in the form of systematic practice with CF (Carroll & Swain, 1993). V2O was also selected because L2 learners acquire it late and because it is typically not instructed in L2 English curricula (Ellis, 2009). The participants of this study had not received any overt instruction on V2O prior to the experiment. Thus, their knowledge of this construction was likely to be implicit only.

For the instruction of QNT, we adopted a conservative approach for the constructions *less/least + uncountable noun* versus *fewer/fewest + countable noun*. In present-day English, especially in informal registers, the quantifier *less* is becoming increasingly more frequent in combination with countable nouns (Leech & Svartvik, 1994). Therefore, communicative English L2 grammars often relax the rules for this distinction. However, we chose to instruct QNT as typical in course books supported by the Flemish curriculum for English, as these focus on formal registers of the L2 with which learners are less familiar on the basis of out-of-class learning. We did, however, tell learners that *less* and *least* were becoming more frequent in English with countable nouns, but that they could not go wrong if they stuck to the rules. Teaching exceptions to these rules might have caused confusion, and is likely to have compromised our interest in the relative effectiveness of practice for constructions varying in complexity.

Instruction for V2O was based on the grammar explanations used in the study on CF by Carroll and Swain (1993), which state that the syntactical alternation between the double object construction (DOC) and the prepositional construction (PC) is governed by two constraints: a morphophonological constraint and a semantic one. The participants of the current study were instructed that the DOC was only possible when two conditions were met: (a) the sentence had to include either a short (i.e., one-syllable) verb or a longer verb (two syllables and more) that had initial stress, and (b) the sentence had to express transfer of possession. They were also told that all other cases require use of the PC. While recent descriptive research shows that the morphophonological constraint does not hold entirely (Anttila, Adams, & Speriosu, 2010), and that the distribution of the DOC and PC is much more probabilistic in actual use (Bresnan & Ford, 2010), we stuck to the simplified rules by Carroll and Swain (1993) for pedagogical reasons. Evidently, pedagogical grammars strive for different objectives than descriptive grammars, and we considered that as long as the general tendencies are reflected in the rule, simplifications of the probabilistic regularities are not too problematic.

In correspondence with how complexity was operationalized in the meta-analysis by Spada and Tomita (2010), rule instruction for QNT and V2O differed in cognitive complexity. While for QNT, learners only had to perform one step to arrive at a solution (i.e., evaluate whether a given quantifier occurred in combination with a countable noun in the plural), our rule system for V2O required learners to apply several steps, involving a semantic and morphophonological component. Table 1 presents an overview of the prototypes for each of the two grammatical constructions, along with sample sentences used in the practice materials.

Table 1. *Overview of the Constructions*

Grammatical Construction	**Prototype**	**Sample Sentence from Practice Materials**
QNT	*many*, *few*, *fewer*, *fewest* + *countable noun*	*Charley has fewer shares in the company.*
	**much*, *little*, *less*, *least* + *countable noun*	**Copies of Coca-Cola use less ingredients.*
V2O	$V^{monosyllabic}$ + NP + NP	*John Pemberton taught me some tricks on how to make Coca-Cola.*
	$V^{polysyllabic}$, initial stress + NP + NP	*Father promised me the rights to the name 'Coca-Cola'.*
	*$V^{polysyllabic}$, final stress + NP + NP	**Pemberton revealed me the secret formula.*
	*$V^{no\ transfer\ of\ possession}$ + NP + NP	**Legend says that Charley stirred his father the first brew of Coca-Cola.*

Note. * *denotes ungrammaticality.*

Procedure and Practice Materials

Figure 1 provides an overview of the experimental procedure. Learners were taught by the first author, who was not their regular teacher and was introduced as a researcher. Other than the difference in teacher and use of the mini-games designed for this study, the experiment was not fundamentally different from regular class time. In Phase 1, after being invited to the study, participants filled out consent forms and completed pre-tests.

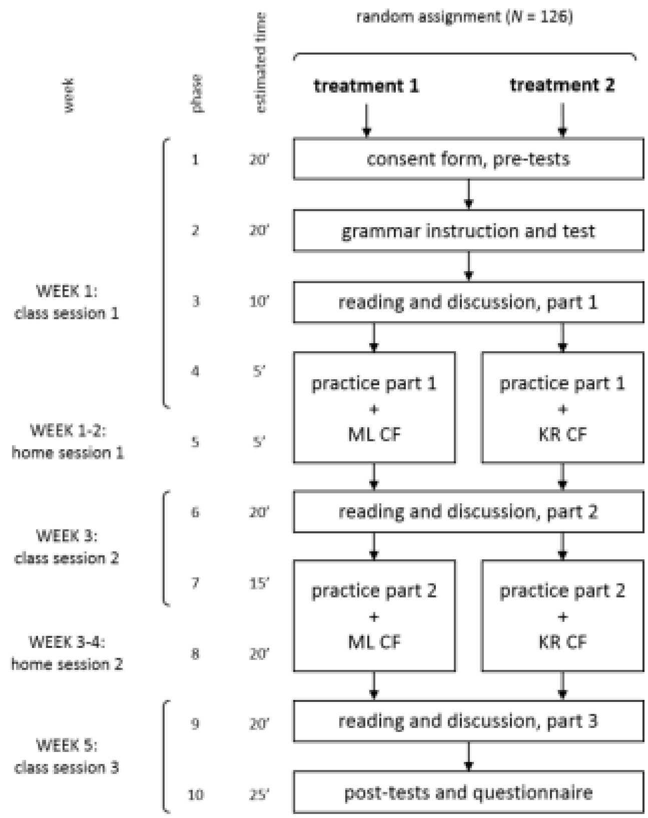

Figure 1. Experimental procedure.

Phase 2 provided explicit rule explanation. This step comprises the declarative knowledge building phase. Rule instruction was done inductively in order to engage the learners more actively, in accordance with recommendations from Ranta and Lyster (2007, pp. 150–151), but relied on decontextualized exemplars (available here). The metalinguistic terminology used was consistent with the learners' course books, and parts of the rule instruction were repeated or elaborated in the learners' mother tongue on an as-needed basis. After the inductive and collaborative identification of rules, the first author provided explicit instruction supported by schemata projected for the entire class (available here). Next, learners took a metalinguistic knowledge test. They were allowed to take notes during the entire rule instruction phase, but were asked to return their sheets with exemplars and any notes before taking the test.

Following this phase, learners read the first section of a mystery text in the L2 written by the first author (available here) and based on the early history of Coca-Cola (Pendergrast, 1997), answered comprehension questions, and discussed the text in class. The function of reading comprehension was to introduce the background context of the linguistic constructions that were to be practised, so that during practice, learners were more likely to also process the constructions for their meaning. Hence, this reading task immediately preceded the first practice activities.

In Phase 4 and Phase 5, learners practised the grammar constructions by means of web-based mini-games. The learners were told that their task was to solve the mystery introduced in the text, and that the mini-games were intended to involve them as the detective in the story, interrogating witnesses and potential suspects by means of a special device called a *tele-interrogator*. The utterances of the interviewees were presented in written form only and were drawn from the next chapter of the mystery text, thus offering learners clues that could help resolve the mystery. Although the format of the mini-game was interrogation, involving a focus on meaning, the learners were required to judge whether the sentences were well-formed according to grammar instruction given earlier. Learners used the J- and F-keys on the computer keyboard to indicate whether they considered the sentences grammatical or ungrammatical, respectively. This format was designed for three reasons. First, it is easy to automatically score and provide feedback on. Second, save from its potentially stronger meaning-focus, the format closely mirrors grammaticality judgment tasks, which were used in pre- and post-testing, allowing to measure near transfer in terms of task type. And third, in contrast to other examples of mini-games, the format was very limited with respect to interactivity. In order to perform the task, learners did not need to take into account any elements on the screen other than the sentences and the CF—such as obstacles to be evaded or targets to be hit. While this might have been a more stimulating task for the adolescent learners, more advanced element interactivity might have constituted a dual task condition, detracting from the learning content and hindering transfer. This decision was taken in line with recent research which showed that game interactivity may induce extraneous cognitive load and hamper L2 learning (deHaan, Reed, & Kuwada, 2010).

Figure 2. Tutorial version of the mini-game.

Learners practised with two versions of the mini-games. They first practised with a tutorial version, in which immediate CF was given and which lacked time pressure and rewarding. For the learners in one group, this feedback consisted of knowledge-of-results only (KR CF), visualized as a green checkmark or a red cross, with audio support, and a statement that their response was incorrect. In the other group this statement was expanded with a metalinguistic explanation displayed on top of the screen (ML CF; see Figure 2). In addition, ungrammatical sentences were highlighted in red for both groups. The tutorial mini-games were intended to give learners the opportunity to apply declarative knowledge about the grammatical structures, fine-tune it through interaction with CF, and proceduralize knowledge. In accordance with SAT, learners need to get ample time in the proceduralization phase; they should not be rushed (DeKeyser, 1998, p. 55). Therefore, no time pressure was included in the tutorial games. According to DeKeyser (2007a) however, proceduralization does not take long: "being required to use a rule a limited number of times to process a set of sentences is all it takes" (p. 290). Thus, learners practised each grammatical construction once through 12 sentences (4 grammatical and 8 ungrammatical). The items were offered in 4 sets in counterbalanced order; the order of items within each set was randomized, but equal for all participants. After the completion of these 12 items, learners were shown an overview of their responses with KR CF or ML CF, depending on the treatment condition.

Figure 3. Full version of the mini-game.

At the end of the first classroom session, learners practised with the full version of the mini-games (see Figure 3). This version included time pressure, immediate KR CF, and points awarded for correct responses, and a game design feature intended to provide a humorous effect when they failed. This meant that their interrogation device was damaged with each incorrect response (supported by animations and sound effects), and broke down if they eventually made five mistakes, ending the practice session. When learners made fewer than five mistakes, the practice session was stopped after one and a half minutes. Subsequently, the system displayed an overview of the items that were answered incorrectly, allowing participants to further develop and fine-tune declarative knowledge. For the first treatment group, this screen comprised metalinguistic explanations (see Figure 4); the other group only saw a list of incorrect responses marked as either grammatical or ungrammatical. So, both groups saw KR CF in the speeded mini-game, but once the practice session ended, learners in the first treatment group got delayed ML CF in addition. These practice materials consisted of 24 items for each grammatical construction (8 grammatical and 16 ungrammatical), which were again offered in 4 sets in counterbalanced order and looped when learners had completed all 24 items before the one-and-a-half-minute practice session had ended.

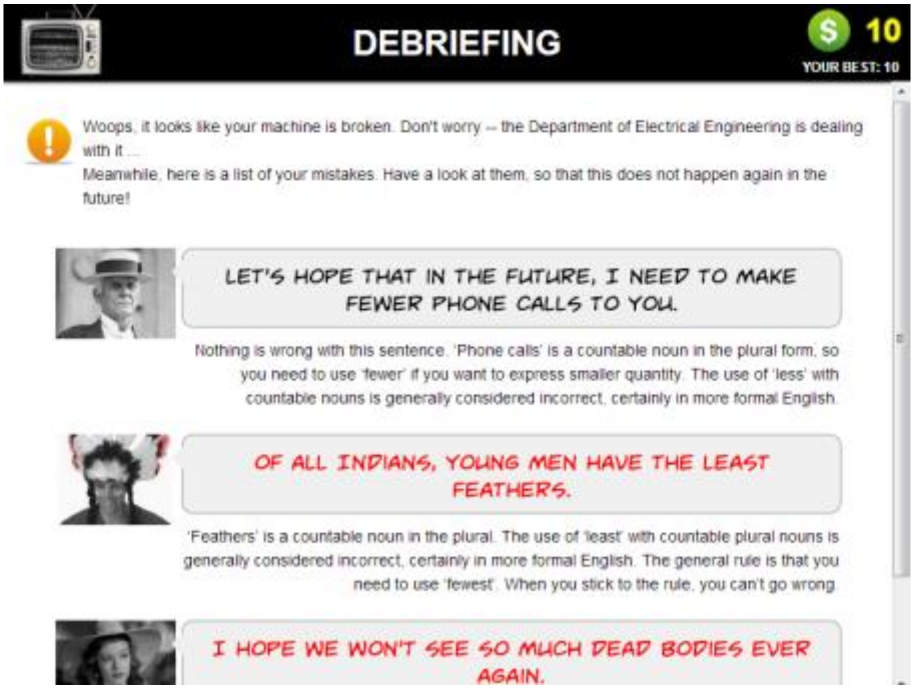

Figure 4. Overview of incorrect responses with ML CF.

The time pressure in the mini-games was intended to stimulate learners to automatize their receptive knowledge of the constructions and was rather mild. The time allowed to judge each sentence was based on sentence length and earlier research with grammaticality judgment tasks (e.g. Gutiérrez, 2013), with the baseline ranging between 3.00 and 6.48 secs depending on sentence length plus 10 seconds and divided by the square root of a number between 1 and 5. This number was incremented for each series of four consecutive correct responses, which increased time pressure and multiplied the learner's score for subsequent correct responses.

After the first class session, learners got opportunities for voluntary practice with the mini-games at home or at school. The materials for QNT and V2O were interwoven in order to provide equal exposure to both constructions. The learners used anonymous codes to sign in to the website, and the system continued to log their behaviour. Learners were asked to practise at least 20 minutes at home before the next class session. In order to engage them, a leaderboard was shown, anonymously comparing the learner's personal highest score for each grammatical construction with the five highest scores, and the participating teachers regularly reminded them of their assignment.

In the first part of the second classroom session, two weeks later, learners were engaged in reading comprehension and discussion of the next episode of the mystery text (Phase 6). In this phase, the text served a double purpose. First, as in the first session, it was intended to make learners focus on meaning during the subsequent practice phase. Second, the text now included one grammatical sentence for each verb used in the practice material for V2O. In line with recommendations from DeKeyser (1998), this was meant to help learners further automatize their knowledge of V2O while reading and discussing the text. Attempts failed to equally include items for the QNT constructions, as the text started to feel artificial. The first author did try to elicit learners' use of the constructions for QNT (as for DOC) during the meaning-oriented discussions in class. In order to keep learners' attention focused on meaning as much as possible during the reading and discussion phases, it was decided to not enhance the text visually (e.g., highlighting of the linguistic phenomena), and the first author tried to avoid pronouncing the constructions with particular stress during class discussion. Following discussion, the learners again practised the constructions for QNT and V2O, but with new practice items that were based on the final part of the story. At the end of this class session, learners were again asked to practice 20 minutes at home before the next session.

In the third classroom session, two weeks later, learners read and discussed the conclusion of the mystery text, which again included examples for V2O. Post-tests and a motivation questionnaire concluded the procedure. Analysis of the pre-tests, metalinguistic knowledge test, post-tests, and questionnaire data is outside the scope of this article.

Research Design, Data Preparation, and Analyses

The study used a mixed factorial experimental design (i.e., with both within- and between-subject factors). One within-subject factor was the grammatical construction. The software was programmed to present the two constructions in interleaved ways so as to obtain a balanced distribution of the constructions for each participant. As a between-subject factor, we had the type of CF that was provided during practice (knowledge-of-results vs. metalinguistic). With a view to obtaining a true experiment and increasing statistical power, participants were assigned at random to one of these treatment conditions within classes. In classroom-based SLA research, this is typically impossible, necessitating the use of quasi-experimental designs instead.

Tracking technology was implemented in the mini-games in order to measure learners' behaviour in practice in fine-grained ways. Individual responses were recorded and timestamped in view of constructing accuracy rates and response times as measures of the degree to which knowledge of the constructions became automatized.

After the treatment, the data were downloaded from the web server, and prepared and analysed in R, a programming environment for statistical computing. The data were first inspected for unexpected values. Out of a total of 92,700 responses, the response times of 8 responses were removed because the values were abnormally high. Then, for each practice session, averages were computed of accuracy and response time. Although all practice sessions dealt with the same constructions, the actual instances of these constructions (i.e., sentences) differed for each practice period (Weeks 1–2 and Weeks 3–4), as they drew on different story content. The sessions were coded accordingly and were then numbered in order of their start time, per practice period, and by grammatical construction. A session number within a practice period thus covers one session for QNT and one for V2O. This variable constitutes the second within-subject variable in our research design, though it was not experimentally manipulated. The Appendix shows a sample of the prepared data. As learners could practise as much as they wanted, the number of practice sessions was likely to vary between participants.

To model the effects of session number, construction, and feedback type on accuracy and response time, and in order to account for the unbalanced nature of the data (due to differences in how frequently learners practised), the data were analysed through multilevel modelling. The two practice periods were analysed separately, as they covered different story content, resulting in four different analyses (two with accuracy as response variable, and two for response time).

For each of the four analyses, model selection involved comparing the fit of a linear model with the fit of a polynomial (quadratic) model, modelling two possible patterns of learning trajectories. This resulted in the following model with a linear trend over sessions:

$$Y_{ij} = \beta_0 + \beta_1 Session_{ij} + \beta_2 Feedback_j^{ML} + \beta_3 Construction_{ij}^{V2O} + \beta_4 (Construction^{V2O} * Feedback^{ML})_{ij} + u_{0j} + e_{ij}$$

where the response variable Y at measurement i for participant j is either the mean accuracy or mean response time of a practice session as predicted by the fixed factors session number, feedback type and grammatical construction (including the interaction between the latter two factors). $Feedback^{ML}$ and

Construction^{V2O} are equal to 1 if the type of feedback is ML and the type of construction is V2O, respectively, otherwise they are equal to 0. Therefore, β_2 refers to the increase of the expected value of variable *Y* when the type of feedback is ML, β_3 to the increase when the type of construction is V2O, and β_4 to the additional increase (on top of $\beta_2 + \beta_3$) when, at the same time, feedback is ML and construction is V2O. In addition, the intercept of the model was allowed to vary randomly between participants. As the participants were allowed to practise outside of classroom time, the sessions are likely to be distributed unequally. Consequently, residuals within and between persons, e_{ij} and u_{0j}, are assumed to be independently and identically normally distributed with zero means and with variances σ_{e^2} and σ_{u^2}. For the model with a quadratic trend over time, we added an additional effect to the model: the effect of *Session*^{2}. The model parameters were estimated using the full maximum likelihood procedure, which, unlike the generally preferable restricted maximum likelihood (REML) procedure, allows an analysis of variance comparison of the fit of models that have different fixed factors (Maindonald & Braun, 2010). Model fit was compared using Akaike's Information Criterion (AIC), where lower values indicate a better fit. After model selection, we re-estimated the model parameters using REML to obtain unbiased variance estimates.

RESULTS

Table 2 comprises the descriptive statistics for the behaviour-tracking data per practice period and session. The data show that attrition was quite high: roughly half of the participants had dropped out by the eighth session in the first period, and after seven sessions in the second period. Less than a quarter of the participants were still active in the tenth session of both periods. The number of learners was roughly equal in both feedback groups.

Figures 5–8 plot mean accuracy and response time for both practice periods over time, per feedback type (KR vs. ML) and construction (QNT vs. V2O). The plots show that the data tend to follow a curve that rises or drops quite rapidly in the beginning, followed by more gradual increases or decreases that level off at the end. This applies particularly to response time. The tails of the plots are less consistent with this pattern, which can be explained by participant attrition, resulting in a less reliable signal. Further, in the second session of the first practice period, accuracy drops to about the chance level (50%) for three of the four conditions. This may be due to the introduction of time pressure in the second practice session. A similar explanation can account for the steep decay in response time in the second practice session.

Table 2. *Descriptive Statistics for Behaviour-tracking Data*

Practice period		Session number 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 (Weeks 1–2)	*N* (all)	126	125	118	107	100	88	78	65	46	29	23	10	8	3	2	2
	N (ML)	63	62	59	53	49	43	39	33	22	14	12	8	7	2	2	2
	N (KR)	63	63	59	54	51	45	39	32	24	15	11	2	1	1		
	Accuracy	.60 (0.18)	.54 (0.22)	.60 (0.21)	.61 (0.22)	.65 (0.23)	.68 (0.20)	.68 (0.21)	.68 (0.20)	.67 (0.21)	.68 (0.21)	.68 (0.24)	.72 (0.21)	.67 (0.24)	.60 (0.22)	.60 (0.18)	.67 (0.21)
	Response Time	8.29 (4.05)	4.79 (1.60)	3.96 (1.38)	3.42 (1.32)	3.24 (1.35)	2.86 (1.27)	2.83 (1.37)	2.63 (1.32)	2.58 (1.38)	2.63 (1.64)	2.29 (1.95)	2.36 (1.12)	1.89 (1.41)	1.64 (0.39)	3.32 (2.98)	2.47 (0.92)
2 (Weeks 3–4)	*N* (all)	119	113	96	80	71	64	55	43	32	23	20	13	12	6	6	4
	N (ML)	60	59	50	41	36	29	27	19	13	9	8	6	5	3	3	2
	N (KR)	59	44	46	39	35	32	28	24	19	14	12	7	7	3	3	2
	Accuracy	0.65 (0.23)	0.69 (0.22)	0.71 (0.23)	0.74 (0.21)	0.76 (0.20)	0.77 (0.18)	0.80 (0.18)	0.80 (0.16)	0.80 (0.22)	0.79 (0.20)	0.82 (0.18)	0.84 (0.14)	0.84 (0.12)	0.85 (0.13)	0.91 (0.07)	0.87 (0.11)
	Response Time	3.73 (1.61)	2.98 (1.43)	2.71 (1.38)	2.41 (1.37)	2.14 (1.19)	2.04 (1.18)	1.85 (1.08)	1.77 (1.04)	1.64 (0.98)	1.69 (1.06)	1.48 (0.98)	1.37 (0.91)	1.37 (1.18)	1.44 (1.30)	1.45 (1.07)	1.72 (1.16)

Practice period		Session number 17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1 (Weeks 1–2)	*N* (all)	1	1	1	1	1	1										
	N (ML)	1	1	1	1	1	1										
	N (KR)																
	Accuracy	.82 (0.19)	.80 (0.16)	.60 (0.32)	.71 (0.30)	.85 (0.16)	.98 (N/A)										
	Response Time	1.74 (1.08)	1.57 (0.87)	1.11 (0.19)	1.13 (0.20)	1.38 (0.71)	0.85 (N/A)										
2 (Weeks 3–4)	*N* (all)	3	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1
	N (ML)	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	N (KR)	1	1	1	1												
	Accuracy	.93 (0.05)	.93 (0.08)	.96 (0.03)	.96 (0.03)	.94 (0.03)	.92 (0.03)	.96 (0.01)	.96 (0.02)	0.98 (0.03)	.97 (0.04)	.96 (0.01)	.99 (.01)	.96 (0.06)	.99 (0.01)	.96 (0.02)	1.00 (N/A)
	Response Time	1.77 (1.22)	1.89 (1.50)	1.68 (1.25)	1.49 (1.06)	2.06 (0.22)	1.93 (0.25)	1.67 (0.32)	1.64 (0.23)	1.63 (0.39)	1.81 (0.52)	1.48 (0.22)	1.39 (0.20)	1.22 (0.10)	1.27 (0.26)	1.28 (0.15)	1.53 (N/A)

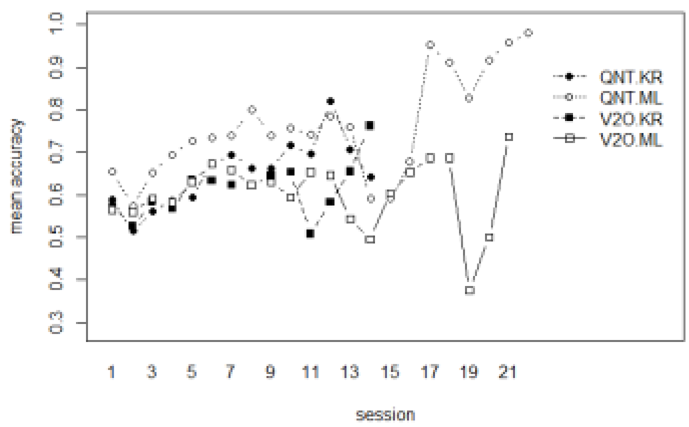

Figure 5. Mean accuracy for first practice period.

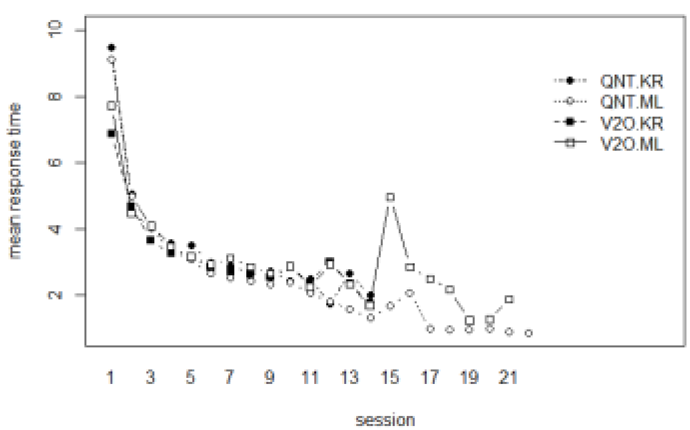

Figure 6. Mean response time for first practice period.

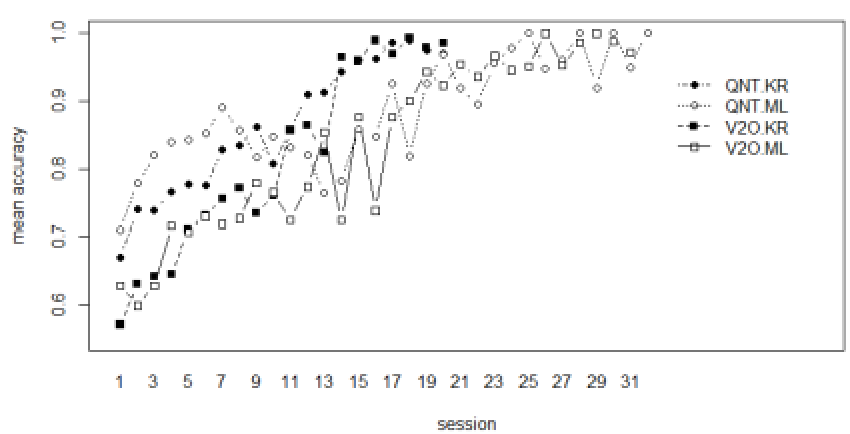

Figure 7. Mean accuracy for second practice period.

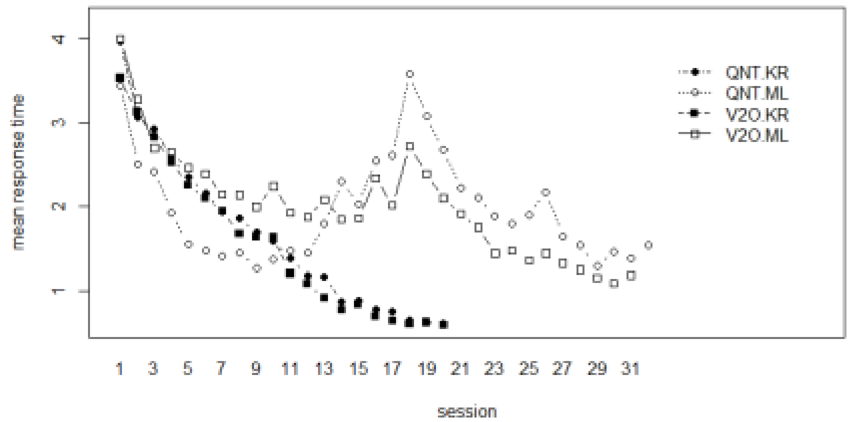

Figure 8. Mean response time second practice period.

For model selection and analysis, we decided to only use the first 11 sessions in each practice period, as the number of participants in the subsequent session drops to under 20, yielding potentially unreliable measurements. The initial analyses for the first practice period did not reveal a difference in fit between the linear and quadratic model for accuracy (see Table 3). As time pressure was introduced in the second session, it can be argued that the data from this session onward are incomparable to the data gathered in the first practice session. We therefore decided to discard the first session. Removal of this session yielded a better fit of the quadratic model for the accuracy data from the first practice period. The other models equally improved when a quadratic function was fitted to the data.

Table 3. *Fit of the Linear versus Quadratic Models*

Practice period	Response variable	Model	*AIC*	Δ (-2 log likelihood)	*p*
1	Accuracy	Linear	-910.79	462.40	.130
		Quadratic	-911.09	463.54	
	Response Time	Linear	7733.97	-3859.99	<.001
		Quadratic	7391.41	-3687.70	
1 (without tutorial session)	Accuracy	Linear	-765.09	389.54	<.001
		Quadratic	-784.97	400.48	
	Response Time	Linear	4923.04	-2454.52	<.001
		Quadratic	4834.97	-2409.49	
2	Accuracy	Linear	-843.66	428.83	<.001
		Quadratic	-853.02	434.51	
	Response Time	Linear	4143.75	-2064.88	<.001
		Quadratic	4088.32	-2036.16	

As multilevel modelling assumes normal distributions of the residuals at both levels, the residuals of the four statistical models were inspected visually using QQ plots and histograms. This did not reveal any abnormalities.

Table 4. *Fixed Effects*

Practice period	Response variable	Parameter	Estimate	*SE*	Degrees of freedom	*t*-value	*p*-value
1 (without tutorial session)	Accuracy	(intercept)	.42	0.03	1363	14.73	<.001
		session	.06	0.01	1363	7.20	<.001
		session²	0	0	1363	-4.69	<.001
		feedback^{ML}	.08	0.03	123	3.19	.002
		construction^{V2O}	-.01	0.01	1363	-1.06	.287
		feedback^{ML}: construction^{V2O}	-0.07	0.02	1363	-3.77	<.001
	Response Time	(intercept)	6.15	0.19	1363	32.51	<.001
		session	-0.79	0.06	1363	-14.00	<.001
		session²	0.04	0	1363	9.62	<.001
		feedback^{ML}	-0.15	0.18	123	-0.83	.407
		construction^{V2O}	-0.22	0.08	1363	-2.74	.006
		feedback^{ML}: construction^{V2O}	0.35	0.12	1363	3.02	.003
2	Accuracy	(intercept)	.64	0.02	1255	28.88	<.001
		session	.04	0.01	1255	6.62	<.001
		session²	0	0	1255	-3.37	<.001
		feedback^{ML}	.05	0.03	117	1.90	.061
		construction^{V2O}	-.09	0.01	1255	-6.99	<.001
		feedback^{ML}: construction^{V2O}	-.05	0.02	1255	-2.68	.007
	Response Time	(intercept)	4.18	0.16	1255	26.93	<.001
		session	-0.53	0.04	1255	-14.07	<.001
		session²	0.03	0.01	1255	7.65	<.001
		feedback^{ML}	-0.40	0.19	117	-2.08	.039
		construction^{V2O}	-0.13	0.07	1255	-1.80	.072
		feedback^{ML}: construction^{V2O}	0.81	0.10	1255	7.82	<.001

Table 4 comprises the effects of the fixed factors and the interaction between feedback and construction for all four models; the corresponding curves with 95% confidence bands are displayed in Figures 9–12. These reveal a significant main effect of session number on accuracy and response time in both practice periods. Further, there was a significant main effect of feedback on accuracy in the first period and on response time in the second period. Results also show a main effect of construction on response time in the first period and on accuracy in the second period. Finally, the interaction between feedback and construction was significant in all four models. The model plots show that ML feedback affected accuracy rates positively for QNT, but not for V2O. Moreover, in comparison with KR feedback, ML feedback resulted in faster response times for QNT and slower response times for V2O. This effect was stronger in the second practice period.

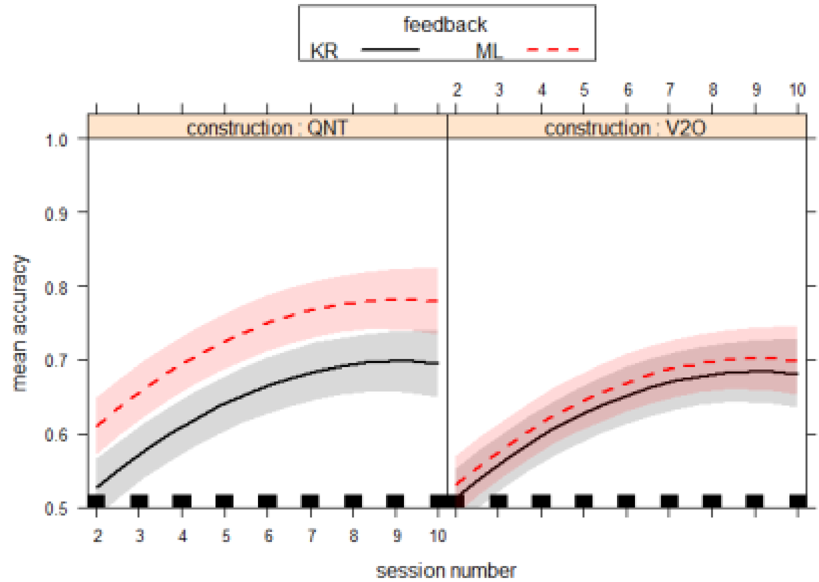

Figure 9. Effect plots for accuracy (first practice period).

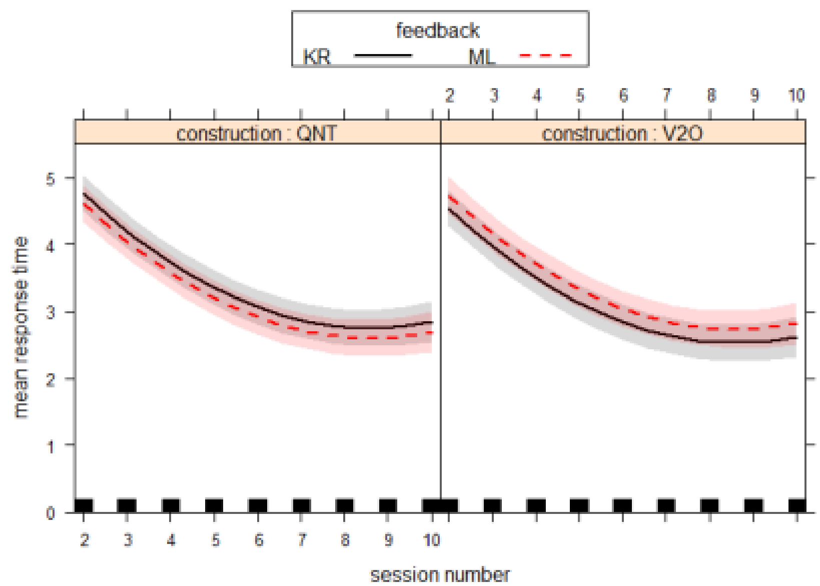

Figure 10. Effect plots for response time (first practice period).

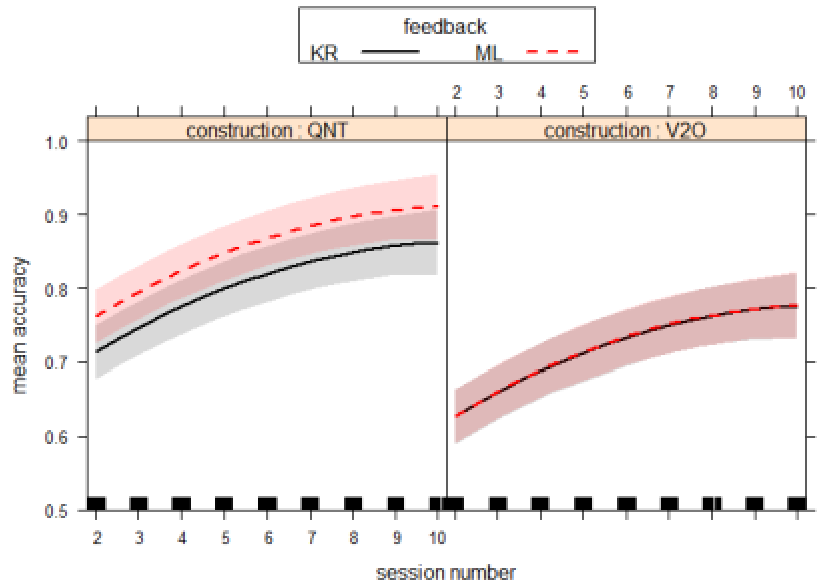

Figure 11. Effect plots for accuracy (second practice period).

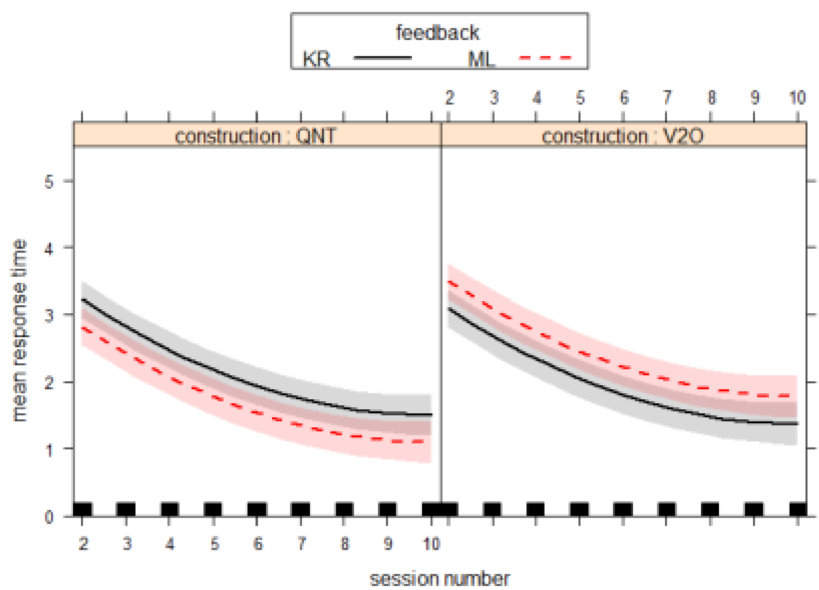

Figure 12. Effect plots for response time (second practice period).

DISCUSSION

As for the first research question, there is some evidence that focused practice helped learners develop automaticity in the ecologically valid context of L2 classrooms and learners' homes. With increasing practice, learners became more accurate and quicker at judging the grammaticality of sentences on QNT and V2O. The curves of accuracy and response time fitted a quadratic function better than a linear one, and were thus characteristic of automatization: in the initial stages of learning, a rapid increase in accuracy and decrease in response time, followed by more gradual changes that flatten out towards the end. This is consistent with the theoretical model of skill acquisition as well as with the learning curves observed in previous research on automatization of L2 grammar (e.g., DeKeyser, 1997). Evidence for automatization can also come in the form of a change in the coefficient of variation (a measure of the relative variability of response time; e.g. Lim & Godfroid, 2015) or as increased fluency on oral tasks (e.g. De Jong & Perfetti, 2011). First results of the analyses of the post-tests in this study show that repeated practice was associated with small accuracy gains on an oral elicited imitation task (Cornillie et al., 2015). Whether practice also benefited fluency on this task merits further investigation.

The second research question concerns the impact of metalinguistic information in feedback as well as rule complexity on learner performance in practice. Results show that the effects of metalinguistic feedback and rule complexity are interdependent. Whereas accuracy scores for QNT and V2O were similar in the first practice period, they were significantly higher for QNT in the second practice period. Moreover, there is an interaction effect of rule complexity and feedback on accuracy. The plots show higher accuracy curves of QNT for learners in the ML group than for their peers in the KR group; the availability of ML feedback in contrast did not improve accuracy for practice of V2O. This indicates that practice with CF was more beneficial for the grammatical construction with a simpler rule system. These findings are in line with DeKeyser's (1998) position that explicit instruction and practice are more beneficial for simple rules. Interestingly, the effect of ML feedback on accuracy was less pronounced in the second practice period. This may be due to the fact that oral grammar instruction, and the accompanying development of explicit knowledge, took place before the first practice period.

ML feedback also impacted response time in the second practice period, and similarly as for accuracy, this effect depended on rule complexity (in both periods). In comparison with their peers who only received KR feedback, learners who had ML information at their disposal responded more quickly to QNT items, but more slowly to V2O items. This interaction effect was stronger in the second practice period. Recall that our more complex rule system of V2O involved more steps in order to arrive at a solution. If learners in the ML group relied on declarative knowledge before responding to V2O items, the relative complexity of the rule system may have resulted in longer thinking, or perhaps even doubt, as time went on. In contrast, it is striking that the response time curves for the learners in the KR group have a comparable intercept for both grammar constructions. Given that ML information was not available to these learners during practice, it is likely that they memorized exemplars instead of relying on the more effortful strategy of retrieving rule-based knowledge.

CONCLUSION

Technological evolutions are creating new opportunities for research on L2 practice. Teaching institutions are adopting online CALL platforms powered by learning analytics. These platforms track learners' behaviour in detail, on multiple devices, and seamlessly while they travel between the classroom, their homes, and other connected spaces and report this information to teachers via dashboards. This implies that teachers can assign focused practice activities to learners outside of the classroom—saving classroom time for more communicative tasks—without losing sight of what their learners do. Researchers can analyse the data collected in such ecologically valid settings to investigate theoretical issues in SLA. Insights of such studies can generate design recommendations that are valid for teaching practice.

The methodological innovation of this study consists of the use of CALL and learning analytics to combine in vitro research techniques that allow for experimental control and measurement with L2 learning and teaching in in vivo settings (classrooms and learners' homes) through learning materials informed by prevailing pedagogical principles, such as the embedding of forms for focused practice in meaning-oriented tasks. On a theoretical plane, the study shows that in this ecologically valid context, learners automatized L2 grammar knowledge, evidenced by learning curves of accuracy and response time that followed a quadratic function. It also suggests that practice and metalinguistic feedback are more beneficial for grammar constructions with low cognitive complexity. Thus, this contribution has demonstrated that relatively controlled experimental research on focused L2 practice can be carried out in real instructed L2 learning contexts while contributing to theoretical issues in SLA. By moving from in vitro to in vivo, we hope to break a lance for the continuation of research on focused L2 practice, with clear implications for teaching.

Implementation of this study, however, required compromises in terms of both experimental control and ecological validity. First, as a feature of our in vivo research environment, the number of sessions was not controlled or manipulated: participants were allowed to practise as much or as little as they wanted. This resulted in high attrition rates, which, although amenable to multilevel statistical modelling, may limit the power of our study.

Weak points from the perspective of external validity concern the design of the focused practice activities, the simplification and complexity of grammar instruction, and the length of the treatment. First, the interactivity of the mini-games was limited and possibly not very stimulating. Next, in order to allow for experimental control and automated feedback, the format of the practice activities was constrained to performing grammaticality judgments in conditions that afforded meaning focus, but did not require it. Time pressure and competition may have resulted in learners' tuning out of meaning, focusing on linguistic content rather than on story. Future redesigns ought to necessitate meaning focus, for instance by further developing the detective task.

Additionally, with a view to reducing cognitive complexity of grammar instruction, the rules for both QNT and V2O were limited in scope (number of structures covered) and reliability (the extent to which grammar rules hold true; see Hulstijn & de Graaff, 1994). While this may also happen in real L2 teaching, this design choice deviates from curricula that prioritize real-world use of English. Moreover, with V2O we also selected a construction that is not likely to be taught in English curricula because of its cognitive complexity.

Finally, the length of practice was limited to four weeks. DeKeyser (2007a) notes that "no experiment lasting only a few weeks is representative of the long-term dynamics of real-world language learning" (p. 301). Future studies could therefore provide teachers with the technology needed to create similar practice materials for longer-term use in L2 classrooms. Alternatively, considering the increasing availability of mobile devices and the low cost of collecting data in the cloud, massive online experiments could be set up in which the practice behaviour of large numbers of L2 learners is captured in informal contexts and over larger stretches of time.

APPENDIX. Sample of the Logging Data (Aggregated on the Level of Practice Session)

user key	period	session number	tutorial	construction	feedback	start	mean accuracy	mean response time	number of completed items
4648711539916800	Weeks 1–2	1	TRUE	QNT	ML	2014-01-22 09:19:19	.92	10.305	12
4648711539916800	Weeks 1–2	1	TRUE	V2O	ML	2014-01-22 09:22:43	.67	4.340	12
4648711539916800	Weeks 1–2	2	FALSE	QNT	ML	2014-02-04 21:13:02	1.00	4.368	20
4648711539916800	Weeks 1–2	2	FALSE	V2O	ML	2014-02-04 21:15:02	.58	5.139	12
4648711539916800	Weeks 1–2	3	FALSE	QNT	ML	2014-02-04 21:21:33	.95	2.032	42
4648711539916800	Weeks 1–2	3	FALSE	V2O	ML	2014-02-04 21:23:42	.44	5.955	9
…									
4648711539916800	Weeks 3–4	1	FALSE	QNT	ML	2014-02-05 08:56:28	1.00	1.181	75
4648711539916800	Weeks 3–4	1	FALSE	V2O	ML	2014-02-05 08:59:21	.55	7.205	11
4648711539916800	Weeks 3–4	2	FALSE	QNT	ML	2014-02-05 09:01:43	.99	1.018	85
4648711539916800	Weeks 3–4	2	FALSE	V2O	ML	2014-02-05 09:04:13	.44	8.734	9
4648711539916800	Weeks 3–4	3	FALSE	QNT	ML	2014-02-05 09:06:24	.99	0.933	93
4648711539916800	Weeks 3–4	3	FALSE	V2O	ML	2014-02-08 17:16:22	.62	1.596	13
…									
4649432020680704	Weeks 1–2	1	TRUE	QNT	KR	2014-01-31 10:44:25	.50	4.532	12
4649432020680704	Weeks 1–2	1	TRUE	V2O	KR	2014-01-31 10:46:07	.50	3.892	12
4649432020680704	Weeks 1–2	2	FALSE	QNT	KR	2014-01-31 10:47:31	.55	3.676	11
4649432020680704	Weeks 1–2	2	FALSE	V2O	KR	2014-01-31 10:49:32	.50	4.321	10
4649432020680704	Weeks 1–2	3	FALSE	QNT	KR	2014-01-31 10:50:39	.58	3.778	12
4649432020680704	Weeks 1–2	3	FALSE	V2O	KR	2014-01-31 10:51:56	.69	4.000	16
…									
4649432020680704	Weeks 3–4	1	FALSE	QNT	KR	2014-02-13 14:50:40	.64	4.037	14
4649432020680704	Weeks 3–4	1	FALSE	V2O	KR	2014-02-13 14:52:03	.50	3.418	10
4649432020680704	Weeks 3–4	2	FALSE	QNT	KR	2014-02-13 14:53:09	.58	2.856	12
4649432020680704	Weeks 3–4	2	FALSE	V2O	KR	2014-02-13 14:54:14	.44	2.257	9

ABOUT THE AUTHORS

Frederik Cornillie is a research manager and postdoctoral researcher in applied linguistics in the interdisciplinary team ITEC at KU Leuven (University of Leuven), Kulak Kortrijk campus, and at the strategic research institute imec in Belgium. His research in computer-assisted language learning takes a particular interest in skill acquisition, feedback, individual differences, and digital gaming. He is associate editor at *ReCALL*.

E-mail: Frederik.Cornillie@kuleuven.be

Wim Van Den Noortgate is a professor of statistics at the Faculty of Psychology & Educational Sciences and ITEC at KU Leuven, Kulak Kortrijk campus, and imec. His main research interests and publications are in the field of multilevel analysis, meta-analysis, and item response theory.

E-mail: wim.vandennoortgate@kuleuven.be

Kris Van den Branden is a professor of linguistics and teacher educator at the Faculty of Arts of KU Leuven. At the same university, he is the academic promoter of the Centre for Language and Education. He is the co-editor of the journal *ITL International Journal of Applied Linguistics*, and one of the volume series editors of *Task-Based Language Teaching: Issues, Research, and Practice*.

E-mail: kris.vandenbranden@kuleuven.be

Piet Desmet is a full professor of French and applied linguistics and computer-assisted language learning at KU Leuven, Kulak Kortrijk campus, and imec, Belgium. He coordinates the research team imec-ITEC-KU Leuven, focusing on domain-specific educational technology with a main interest in language learning and technology. He leads a range of research projects in this field devoted to the integration of human language technologies into CALL and to the effectiveness of adaptive and personalized learning environments.

E-mail: piet.desmet@kuleuven.be

REFERENCES

Anderson, J. R. (1992). Automaticity and the ACT* theory. *American Journal of Psychology*, *105*(2), 165–180. doi:10.2307/1423026

Anttila, A., Adams, M., & Speriosu, M. (2010). The role of prosody in the English dative alternation. *Language and Cognitive Processes*, *25*(7–9), 946–981. doi:10.1080/01690960903525481

Bresnan, J., & Ford, M. (2010). Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language*, *86*(1), 168–213.

Carroll, S. E., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition*, *15*(3), 357–386. doi:10. 1017/S0272263100012158

Cornillie, F., Van den Branden, K., & Desmet, P. (2015). From language play to linguistic form and back again: Lessons from an experimental study for the design of task-based language practice supported by games. In J. Colpaert, A. Aerts, M. Oberhofer, & M. Gutiérez-Colón Plana (Eds.), *Proceedings of XVIIth international CALL conference: Task design and CALL. 6–8 July 2015. Tarragona, Spain* (pp. 214–222). Antwerp, Belgium: Universiteit Antwerpen.

De Jong, N., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, *61*(2), 533–568. doi:10.1111/j.1467-9922.2010.00620.x

deHaan, J., Reed, W. M., & Kuwada, K. (2010). The effect of interactivity with a music video game on second language vocabulary recall. *Language Learning & Technology*, *14*(2), 74–94. Retrieved from http://llt.msu.edu/vol14num2/dehaanreedkuwada.pdf

DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition*, *19*(2), 195–221. doi:10.1017/s0272263197002040

DeKeyser, R. M. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 42–63). Cambridge, UK: Cambridge University Press.

DeKeyser, R. M. (2007a). Conclusion: The future of practice. In R. M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 287–304). New York, NY: Cambridge University Press.

DeKeyser, R. M. (2007b). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 97–114). New York, NY: Routledge.

Ellis, R. (2009). Implicit and explicit learning, knowledge, and instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing, and teaching* (pp. 3–29). Bristol, UK: Multilingual Matters.

Gropen, J., Pinker, S., Hollander, M., Goldberg, R., & Wilson, R. (1989). The learnability and acquisition of the dative alternation in English. *Language*, *65*(2), 203–257. doi:10.2307/415332

Gutiérrez, X. (2013). The construct validity of grammaticality judgment tests as measures of implicit and explicit knowledge. *Studies in Second Language Acquisition*, *35*(3), 423–449. doi: 10.1017/S0272263113 000041

Heift, T., & Schulze, M. (2015). Tutorial computer-assisted language learning. *Language Teaching*, *48*(3), 471–490. doi:10.1017/S0261444815000245

Hulstijn, J. H. (2007). Psycholinguistic perspectives on language and its acquisition. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching* (Vol. II, pp. 783–796). Norwell, MA: Springer.

Hulstijn, J. H., & de Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA Review*, *11*, 97–112. Retrieved from http://www.aila.info/download/publications/review/AILA11.pdf

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning*, *57*(1), 1–33. doi:10.1111/j.1467-9922.2007.00397.x

Leech, G., & Svartvik, J. (1994). *A communicative grammar of English*. London, UK: Longman.

Lim, H., & Godfroid, A. (2015). Automatization in second language sentence processing: A partial, conceptual replication of Hulstijn, Van Gelderen, and Schoonen's 2009 study. *Applied Psycholinguistics*, *36*(5), 1247–1282. doi:10.1017/S0142716414000137

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, *19*(01), 37–66. doi:10.1017/S0272263197001034

Maindonald, J., & Braun, W. J. (2010). *Data analysis and graphics using R: An example-based approach* (3rd ed.). Cambridge, UK: Cambridge University Press.

Paulston, C. B., & Bruder, M. N. (1976). *Teaching English as a second language: Techniques and procedures*. Cambridge, MA: Winthrop.

Pendergrast, M. (1997). *For God, country, and Coca-Cola: The unauthorized history of the Great American soft drink and the company that makes it*. New York, NY: Touchstone.

Ranta, L., & Lyster, R. (2007). A cognitive approach to improving immersion students' oral language abilities: The Awareness–Practice–Feedback sequence. In R. M. DeKeyser (Ed.), *Practice in a Second Language: Perspectives from applied linguistics and cognitive psychology* (pp. 141–160). New York, NY: Cambridge University Press.

Robinson, P. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition*, *19*(2), 223–247. doi:10.1017/s0272263197002052

Rodgers, D. M. (2011). The automatization of verbal morphology in instructed second language acquisition. *IRAL - International Review of Applied Linguistics in Language Teaching*, *49*(4), 295–319. doi:10.1515/iral.2011.016

Schulze, M. (2003). Grammatical errors and feedback: Some theoretical insights. *CALICO Journal*, *20*(3), 437–450. Retrieved from https://calico.org/html/article_287.pdf

Segalowitz, N., & Hulstijn, J. H. (2005). Automaticity in bilingualism and second language learning. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371–388). Oxford, UK: Oxford University Press.

Serafini, E. J. (2013). *Cognitive and psychosocial factors in the long-term development of implicit and explicit second language knowledge in adult learners of Spanish at increasing proficiency*. Georgetown University, Washington, DC.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, *60*(2), 263–308. doi:10.1111/j.1467-9922.2010.00562.x

Tops, G. A. J., Dekeyser, X., Devriendt, B., & Geukens, S. (2001). Dutch speakers. In M. Swan & B. Smith (Eds.), *Learner English: A teacher's guide to interference and other problems* (2nd ed., pp. 1–20). Cambridge, UK: Cambridge University Press.

Wong, W., & VanPatten, B. (2003). The evidence is IN: Drills are OUT. *Foreign Language Annals*, *36*(3), 403–423. doi:10.1111/j.1944-9720.2003.tb02123.x