Available online at www.ejal.eu
http://dx.doi.org/10.32601/ejal.775806
Eurasian Journal of Applied Linguistics, 6(2) (2020) 265–284

EJAL
Eurasian Journal of Applied Linguistics

The structural sensitivity of elicited imitation as a measure of implicit grammatical knowledge

Hedayat Sarandi [a] *

[a] *Istanbul Sabahattin Zaim University, Dep. of English Language Teaching, Halkalı, Istanbul, 34303, Turkey*

Received 9 November 2019 | Received in revised form 19 March 2020 | Accepted 5 May 2020

APA Citation:
Sarandi, H. (2020). The structural sensitivity of elicited imitation as a measure of implicit grammatical knowledge. *Eurasian Journal of Applied Linguistics, 6*(2), 265-284.

Doi: 10.32601/ejal.775806

Abstract

One of the major issues in SLA is developing language tests that could produce good measures of implicit knowledge. This study examines the validity of an English Elicited Imitation (EI) test as a measure of L2 implicit grammatical knowledge. Forty freshman university students in Turkey took a set of language tests: an EI test, two storytelling tasks, a picture description task, IELTS listening sample tests and a speaking test. Four English morphemes were chosen as the target structures: third person '-s', plural '-s', simple past '-ed', and comparative '-er'. Results from a principal component analysis showed that all measures were loaded on a single component labelled as implicit knowledge. Significant correlations with varying magnitude were also recorded between learners' EI scores for the target structures and their scores on other time-pressured measures: r = .63, r = .63, r = .65 and r = .43, for third person '-s', simple past '-ed', plural '-s' and comparative '-er', respectively. These findings suggest that the likelihood of EI measuring L2 implicit grammatical knowledge may vary depending on language structures.

© 2020 EJAL & the Authors. Published by *Eurasian Journal of Applied Linguistics (EJAL)*. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (CC BY-NC-ND) (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Elicited imitation; implicit knowledge; learning English grammar; second language acquisition

1. Introduction

Elicited imitation (EI) is a language test that requires participants to listen to some stimulus sentences of varying length and then repeat them as accurately as they can (Underhill, 1987). The success in the test performance is assessed by the ability of the test takers to repeat these sentences or the target linguistic features implanted inside them verbatim. The collected data can later be analyzed either manually or through computer-generated programs (e.g., Graham, Lonsdale, Kennington, Johnson, & McGhee, 2008). Despite its non-communicative design, EI has attracted the attention of scholars with various research interests. It was initially used in child language studies to measure grammatical and lexical acquisition (e.g., Smith, 1970) and

* Corresponding author. Tel.: +90-212-692-8940
E-mail address: hidayet.sarandi@gmail.com
http://dx.doi.org/10.32601/ejal.775806

neuropsychological studies to assess language problems (e.g., Berry 1976), but later it was employed in second language acquisition (SLA) research as well (e.g., Naiman, 1974).

In SLA, the interest in EI as a test of implicit knowledge was triggered by R. Ellis's (2005) claim that time-pressured and meaning-focused language tests such as interviews, oral narration and EI could produce likely measures of learners' intuitive use of their internal grammar. The assertion triggered off several validating studies whose findings were not in the same direction (see Bowles, 2011; Erlam, 2006; Kim & Nam, 2017; Spada, Shiu, & Tomito, 2015; Sarandi, 2015; Suzuki & DeKeyser, 2015; Zhang, 2014). While some supported the claim that EI is a valid method to measure implicit grammatical knowledge (see Bowles, 2011; Erlam, 2006; Kim & Nam, 2017; Sarandi, 2015; Zhang, 2014), others either produced mixed results (Spada et al., 2015), or challenged it altogether (Suzuki & DeKeyser, 2015). The present study aims to further investigate the construct validity of EI as a measure of implicit grammatical knowledge. It aims to examine whether EI has structural sensitivity and as such can produce a better measure of some grammatical structures than others.

1.1. Types of linguistic knowledge

There is a consensus that linguistic knowledge can be of two different types: implicit and explicit. Implicit knowledge is unconscious, systematic, and may not be verbalizable whereas explicit knowledge is conscious, less consistent, and verbalizable, at least through a nontechnical language (R. Ellis, 2005). Implicit knowledge is mainly developed through communicative message-focused interaction whereas explicit knowledge is mainly developed in formal classroom settings (Bialystok, 1978). Different task conditions, likewise, induce learners to draw on different knowledge types. Implicit knowledge is accessed when the focus is on meaning and when there is time pressure to perform the task online. Explicit knowledge, on the other hand, is accessed when the focus is on language forms, and when there is ample time to plan and monitor language use (R. Ellis, 2005; Erlam, 2006).

Of importance is also the nature of interaction between explicit and implicit knowledge. According to weak interface theory, explicit knowledge can facilitate and promote implicit learning of language features (N. Ellis, 2005). This is mainly conducted through noticing. The conscious involvement in the construction of rules, which usually characterize the initial stages of language learning, helps learners notice structures and recognize patterns in the incoming input as well. Once the association is created between language features, noticing is no longer necessary and implicit tallying and consolidation of rules can occur through mere exposure. Language knowledge is subsequently developed through "the memories of utterances in the history of language use and the abstraction of regularities within them" (N. Ellis, 2005, p. 306).

There are several reasons why the measurement of implicit knowledge is of importance in SLA. Firstly, even though language users usually have both implicit and explicit knowledge, the main source of linguistic knowledge is implicit (R. Ellis, 2005; Erlam, 2006; Paradis, 2009). It is mainly the tacit representations of language elements and the ability to use them under real-life conditions that account for linguistic knowledge. Secondly, SLA researchers are interested in discovering whether implicit learning without explicit attention to form is possible. They intend to find out if learners who are flooded with certain linguistic features are capable of developing their implicit knowledge (Ellis & Roever, 2018). Furthermore, developing valid and reliable measures of implicit knowledge can also help researchers resolve the controversy related to interaction between implicit and explicit knowledge (interface theory). There are different theories regarding the possible interaction between explicit and implicit knowledge: non-interface, weak interface and strong interface (see Ellis, 2009 for a review). A valid measure of implicit knowledge allows researchers to probe the nature of interaction between the two knowledge types. It enables them, for example, to examine whether different types of explicit form-focused instruction can affect learners' development of implicit knowledge, and consequently, their spontaneous use of language structures in daily interaction (see, for example, Ellis, Loewen, & Erlam, 2006; Spada, Jessop, Tomita, Suzuki, & Valeo, 2014; Toth & Guijarro-Fuentes, 2013). Thirdly, a valid measure of implicit knowledge can help test designers to find out about the construct validity of their tests and the nature of language knowledge they measure. They can determine whether their tests induce test takers to draw on implicit knowledge, explicit knowledge or both (Ellis & Roever, 2018).

1.3. EI as a test of implicit knowledge

The theoretical underpinning of EI is that the successful perception and production of stimulus sentences presuppose the existence of some degree of linguistic knowledge on the part of test takers. During the perception phase, they need to parse the stimulus into manageable chunks, and map them against their internal grammar and lexical memory to create meaningful representations. A reverse process happens in the production stage when the surface acoustic memory of stimulus fades away and learners attempt to reproduce original statements by passing the stored representations through their filter of existing grammatical knowledge (Eisenstein, Bailey, & Madden, 1982; Potter & Lombardi, 1990).

Another way to explain how EI taps into core linguistic knowledge is by examining the existing relation between the consumption of working memory capacity and the linguistic knowledge storied in long term memory. For learners with limited linguistic knowledge in long term memory, the processing of language is heavily dependent of working memory capacity which fills up quite fast (Baddeley, Gathercole, & Papagno, 1998; Scott, 1994). These learners may not be able hold larger stretches of language in their working memory and reproduce them. As learners become more familiar with the syntax and vocabulary of a language, they manage to pack information into larger

chunks, which occupy less space in their memory than unrelated words. Learners become capable of producing larger stretches of language with more accuracy and efficiency. In this sense, prior language knowledge counterbalances the limitation of working memory and even increases its capacity allowing language users to conduct more effective and elaborate language processing in their working memory (Baddeley, 2015).

Even though EI has long been the focus of studies in child language development, examining the construct validity of EI as a measure of L2 implicit grammatical knowledge is a relatively new line of research. In a series of articles Ellis (2005, 2006) and Erlam (2006) argued that EI, like other time-restricted language tests, can provide a reliable measure of implicit knowledge. According to Erlam (2006), EI design can be manipulated to increase the likelihood of measuring implicit grammatical knowledge. This can be done by: (a) drawing test takers' attention to meaning (e.g., through employing factual statements and asking learners to respond to the truth value of these statements) (b) inclusion of some delay between the listening and repetition of stimulus sentences to avert memory effect (c) using ungrammatical statements along with the grammatical ones. It is assumed that the automatic correction of ungrammatical items can reflect the reconstructive nature of EI; the fact that learners use their own internal grammar to reproduce the sentences and not the memory of what they just heard.

Several studies have examined the construct validity of EI as a measure of L2 implicit grammatical knowledge. In R. Ellis (2005) native and non-native participants completed a battery of five tests in English: an EI test, an oral narrative task, a timed grammaticality judgement test (TGJT), an untimed grammaticality judgement test (UGJT) and a metalinguistic test. Seventeen English language structures were selected as the target structures. An exploratory factor analysis showed that the results obtained from the EI, oral narrative task and TGJT loaded significantly on one factor whereas the results obtained from the metalinguistic test and UGJT, especially the ungrammatical items, loaded heavily on another factor. The two factors were labelled as implicit and explicit knowledge, respectively. Of the three tests that loaded on factor one, the EI turned out to have the heaviest loading implying that the EI may measure implicit knowledge better than the other two tests.

Erlam (2006) required native and non-native participants to respond to 34 EI truth statements targeting 17 language structures in English. A comparison between learners' EI scores and the IELTS test revealed higher correlations for the listening and speaking parts (r = .72, r = .67, respectively) than the reading and written parts of IELTS (r = .51, r = .46, respectively). A moderate correlation was also recorded between learners' EI performance and their scores on an oral narrative task (r = .47). Erlam interpreted the findings as evidence that shows EI and other time-restricted tests measure the same/similar underlying construct, i.e., implicit knowledge.

Further validating evidence for EI as measure of implicit knowledge comes from the studies that showed EI is capable of determining leaners' developmental stages just

the same way as naturally occurring language does. Ellis (2008), for example, examined learners' EI performance on four English structures (since/for, third person '-s' tag questions and possessive -s), which corresponded to four different categories suggested by Processability Theory (Pienemann, 1998). The accuracy and emergence analyses for these structures showed that EI and Processability Theory made identical predictions regarding the order of acquisition of these items. Similarly, Baten and Cornillie (2019) examined the performance of Dutch university students on a German EI task in order to find out whether EI could differentiate between three difficulty levels of German case making (positional, prepositional, functional) as predicted by Processability Theory (Pienemann, 1998). The accuracy and emergence analyses showed a similar order of difficulty especially between the first two stages and the last stage. These findings offer further evidence that EI might be a likely measure of implicit grammatical knowledge.

Bowles (2011) replicated R. Ellis (2005) using the same set of language tests but this time in Spanish. The results of a confirmatory factor analysis once more created a two factor model confirming R. Ellis' (2005) earlier findings. The EI along with the oral narrative task and the TGJT loaded on one factor, and the metalinguistic language test and the UGJT loaded on another factor. Of three implicit knowledge tests, the EI again had the heaviest load on the first factor. Also, significant correlations were recorded between learners' EI scores, and their scores on the oral narrative task and TGJT (r = .78 and r = .64, respectively) implying the three tests may measure the same knowledge type. In another replication study, Zhang (2014) also reported similar findings with Chinese EFL learners. The results of a confirmatory factor analysis showed that learners' EI scores and TGJT loaded on one factor and the ungrammatical items of the UGJT and meta-linguistic test loaded on another factor confirming the earlier suggestion made by R. Ellis (2005) that EI and TGJT may provide a reliable measures of implicit knowledge.

Sarandi (2015) and Spada et al. (2015) drawing on the inadequacy of the number of test items for each structure in R. Ellis (2005) and Erlam (2006) focused on a single language feature, using more test items instead. There were 18 test items targeting English third person '-s' in Sarandi (2015) and 14 test items targeting English passive tenses in Spada et al. (2015). Sarandi (2015) found a significant correlation between Turkish learners' EI scores and their accuracy scores on an oral narrative task (r = .73). Spada et al. (2015), however, reported mixed results. In their study, EFL learners in Taiwan completed five tests: an EI, an error correction task, a picture description task and a timed written and a timed aural GJT. The results of an exploratory factor analysis revealed that the EI and timed aural GJT were loaded on one factor, labelled as implicit knowledge, and the error correction task was loaded on a second factor, labelled as explicit knowledge. However, no significant correlation was recorded between learners' EI scores and their scores on the picture description task.

Suzuki and DeKeyser (2015) compared the performance of L2 Japanese learners on a Japanese EI task and a word monitoring task in order to find out which of the two

measures could offer a better estimate of implicit knowledge. To this objective, learners' performance on the EI and word monitoring tasks were compared with their performance on a metalinguistic test and an SRT (serial reaction time, a test of implicit language learning). The results showed that the EI significantly correlated with the metalinguistic knowledge test, (r = .46), but not with the SRT. The opposite results were reported for the word monitoring task though. While no significant correlation was found between the word monitoring task and the metalinguistic test, a significant correlation was shown between the word monitoring task and the SRT for the learners who stayed in Japan for longer period of time, (r = .43). The researchers concluded that compared to EI, word monitoring tasks might be a better measure of implicit knowledge. They also suggested that EI might be a better measure of automated explicit knowledge than implicit knowledge.

Granena (2016) investigated whether Japanese English language learners' performance on EI is affected by their memory. Learners' EI performance on two language structures (English third person and plural) was compared with their performance on an Operation Span test (a test of working memory) and a visual letter span test (a short-term memory capacity test). An UGJT was also used to select learners who had considerable explicit knowledge of the two structures (85% and above). The results showed no significant relationship between the EI and any of the cognitive tests. The performance of learners on the EI was also considerably poorer on the EI task compared to the UGJT. The researcher claimed that these findings show that memory factors do not moderate learners' performance on EI and that EI is likely a valid measure of implicit knowledge.

Kim and Nam (2017) examined the performance of native and non-native EFL university students in Korea on five language tests: controlled and uncontrolled time-pressured EI tests, written and oral TGJTs and a metalinguistic knowledge test. The results of a principle factor analysis revealed a three-factor solution with controlled and uncontrolled time-pressured EIs loading heavily on the first factor, the written and oral TGJTs on the second factor and the metalinguistic knowledge test on the third factor. The researchers argued that both factor one and two represented implicit knowledge whereas the third factor represented explicit knowledge. They further reported that the EI tests were stronger measures of implicit knowledge compared to the TGJTs.

Finally, Suzuki (2019) applied fMRI (functional magnetic resonance imaging), a technique used to detect brain activities, to examine which parts of the brain, frontal-basal ganglia circuits (related to procedural memory) or hippocampus and medial temporal lobe (related to declarative memory) is activated by different language and cognitive tests. Twenty-five Chinese learning Japanese as their second language completed a set of language tests in the fMRI scanner. The brain images of the learners showed that the EI, GJT, and metalinguistic tests mainly activated parts of the brain related to the declarative memory. The researcher interpreted the findings as evidence that EI is more likely to involve automated explicit knowledge than implicit knowledge.

Overall, of the studies mentioned above the findings of R. Ellis (2005), Ellis (2008), Erlam (2006), Bowles (2011), Sarandi (2015), Zhang (2014) Granena (2016), and Kim and Nam (2017) lend credence to the assumption that EI can provide measures of implicit grammatical knowledge. Suzuki and DeKeyser (2015) and Suzuki (2019), however, challenged this claim, and Spada et al. (2015) reported mixed results. An important point to consider is the methodological variations employed in the design of the EI studies. Even though EI tasks included both grammatical and ungrammatical statements, learners received different instructions on how to repeat these statements. In Ellis (2005), Erlam (2006), Bowles (2011), Zhang (2014), Sarandi (2015), Kim and Nam (2017) learners were required to repeat all statements in correct English. However, Spada et al. (2015), Suzuki and DeKeyser (2015) and Suzuki (2019) demanded learners to correct the statements when/if they realized that they were ungrammatical. The instructions were hence differed on whether or not learners were made aware of the existence of ungrammatical statements.

The use of ungrammatical statements is problematic. If test takers are informed of the existence of ungrammatical items, then they may invoke their explicit knowledge at some stages of task completion. This might be the reason why Spada et al. (2015), Suzuki and DeKeyser (2015) and Suzuki (2019) found no convincing evidence in favor of EI tasks as a measure implicit knowledge. On the other hand, if learners are not informed of the existence of the ungrammatical statements, as R. Ellis (2005), Erlam (2006) Bowles (2011), Zhang (2014), Sarandi (2015) and Kim and Nam (2017) did, then it is hard to rule out the possibility that test takers' incorrect repetition is not affected by their exposure to ungrammatical items. The present study aims to avert this problem by excluding ungrammatical statements and using only grammatical ones.

Another important difference in the design scheme of EI studies is the number of target structures employed. The number of target structures in R. Ellis (2005), Erlam, (2006), Bowles (2011), and Zhang (2014) was 17, and they used two test items per each structure. Suzuki and DeKeyser (2015) used five Japanese target structures, but they reported on the average performance of learners on these structures, and there were no separate reports for each language feature. Sarandi (2015) and Spada et al. (2015) targeted a single language structure, but their findings were not in the same direction. While Sarandi (2015) found a high correlation between learners' performance of the EI and the oral narrative task, Spada et al. (2015) reported no correlations between the EI and the picture description task. These confounding results beg the question as to whether learners' performance on the EI is affected by the features of target structures. Earlier EI studies in child language acquisition show that some language features are repeated with greater accuracy than the others. Connell and Myles-Zizer (1982), for example, found that children repeated nouns and verbs more accurately than verbal morphemes and articles. The present research aims to address this question in SLA. It aims to find out whether EI is more sensitive to measure the implicit knowledge of some language structures than the others. In the light of discussion above, the following research question is formulated: Is

construct validity of EI as a measure of implicit grammatical knowledge dependent on target language structures?

2. Method

2.1. Participants

The participants of the study were 40 freshman English language learners, 35 females and 5 males, coming from three intact classrooms in an ELT department of a private university in Turkey. The age group of the learners varied from 18-24 with the majority being 19 or 20. The proficiency level of the majority of the learners was B2 (according to CEFR). In order to be eligible to pursue their major in their department, they needed to pass, at least, a B1 level in the language proficiency test of the university. Learners who failed to achieve the passing score were required to complete some general English courses at the English preparatory school of the University and take a similar language proficiency test at the end of the program. The native language of all participants was Turkish. The participants' consent and the approval of school authorities were obtained before the onset of the study.

2.2. Target structures

Four inflectional morphemes were chosen as the target structures: simple present third person '-s', plural '-s', comparative '-er' and simple past tense marker '-ed'. Some of the most common types of errors for Turkish language users on these morphemes are as follows: for third person '-s', the omission of '-s' from tense endings, (e.g., he go instead of he goes), for plural '-s', the omission of '-s' from noun endings after cardinal numbers (e.g. two brother instead of two brothers), for comparative 'er', the use of wrong comparative marker (e.g., more fast instead of faster) or double comparative markers (e.g., more easier instead of easier), and for simple past, the omission of '-ed' from verb endings, (e.g., ask instead of asked), or misformation where other inflections are used instead of '-ed' (e.g., working instead of worked) (see Ellis, 2007). The four target structures belong to different stages of language acquisition according to Pienemann's Processability Theory (1998). The simple past belongs to the category procedure, the plural and comparative belong to the phrasal procedure and the third person '-s' belongs to the -s procedure. The selected structures hence vary in terms of the amount of processing difficulties that they exert on language learners.

2.3. Research instruments

2.3.1. Elicited imitation test

The EI consisted of 42 grammatically correct belief or factual statements of which 12 targeted third person '-s', 10 simple past '-ed', 10 comparative '-er' and 10 plural '-s'. The target structures were implanted in the middle of the statements to avoid the memory effect associated with the beginning and end of sentences (Yasuyao, Wataru, & Lorena, 2009). Learners were required to listen to the statements and choose on the

answer sheet if they were true, not true or they were not sure about them and then repeat them as much as they could after a beep sound. The beep sound was inserted 5 seconds after the listening part. Learners had 8-10 seconds to repeat the statements depending on the length of the statements. The average length of the statements varied from 10 to 19 syllables. Before the learners took the test, they completed four test items in order to familiarize themselves with the test procedure. The main test took roughly 15 minutes to complete. Learners' repetitions were recorded on the computers in the laboratory of the university. Ungrammatical statements were not used because of the instructional issues explained earlier. Examples of the target structures items were:

- An experienced doctor works in a private hospital. (e.g., for third person '-s')
- Short people are usually smarter than tall people. (e.g., for comparative '-er')
- Online shopping started more than twenty years ago. (e.g., for simple past '-ed')
- There are three huge airports in New York City. (e.g., for plural '-s')

2.3.2. Storytelling tasks

Two narrative stories were used to elicit learners' oral production of third person '-s' and simple past '-ed'. The first one was a 336-word story about the daily life of a nurse working in a hospital in London. Learners were given 5 minutes to prepare for the narration, and then they were asked to retell the story in 3 minutes by looking at a list of important events in the story. The story was originally taken from 'Understanding Ideas', a reading book by Swan (1976), but considerable modifications were made to the text to make it fit the level of learners and serve the purpose of the study. The text was written in first person singular but learners were required to narrate the story in third person '-s'. Learners were only informed of the change in the narration before they started to retell it. The shift in narration was carried out to increase the likelihood of learners' use of their implicit knowledge (see Sarandi, 2015).

The second story was 332-words long, and it was about the life of the narrator's grandfather who worked for the Navy during the Second World War. The first paragraph of the story was taken from an on-line page of the language center at the University of Victoria, but the rest of the paragraphs were written by the researcher. The story contained 31 instances of simple past '-ed' tense. A list of important events was also written at the end of the story. An identical process to the first story was carried out for the second story as well. The narration of both stories was conducted in the laboratory of the university, and learners' voices were recorded on the computers.

2.3.3. Picture description task

The picture description task consisted of 15 pictures on PowerPoint slides. The pictures were employed to elicit learners' production of plural '-s' and comparative '-er'. The pictures were taken from the internet using Google search engine and presented to each learner in a private office. At the top of each slide, there was an adjective. Learners were asked to describe the pictures and then compare the entities on the pictures using the adjective given. For example, one of the pictures showed four

dogs chasing a cat, and at the top of the picture was written 'fast'. Following the instruction the learners were expected to produce sentences like 'there are four dogs and they are following/ chasing a cat' and also 'the dogs are faster than the cat' or 'the cat is faster than the dogs'. The first slide was described by the researcher to familiarize learners with the task procedure. Of the 15 adjectives used in the slides, 12 required the use of comparative '-er', the target structure. The other three required the comparative marker 'more', and were used as distractors. Of the 12 adjectives targeting '-er', 7 were identical with the ones used in the EI and 5 were different. Even though there was no time limit for the task completion, the researcher could speed up the process by asking the learners to move to the next task. Learners' voices were recorded on a computer in the researcher's office. The task took roughly 8-12 minutes for each learner to complete.

2.3.4. IELTS tests

The listening parts of three sample IELT tests of Cambridge IELTS textbooks (2000, 2005) were used to measure learners' listening comprehension performance. Each test consisted of four sections with the total number of 40 test items. The tests were used as evaluation measures for the listening and pronunciation course that learners took during the first semester. Each test took roughly 30-35 minutes to complete. Learners took the three tests with two weeks interval.

2.3.5. Speaking test

The speaking test was part of an oral communication course evaluation procedure. The testing session took place in the laboratory of the university. Each learner was given two questions and asked to answer either one of the questions or both. Learners were given one minute to prepare their answers on a piece of paper and three minutes to reply. The testing session took place in the laboratory of the university and learners' answers were recorded on the computers. Below are the questions used in the speaking test: *1. Describe a job that you find boring or extremely difficult. What features should people have to become successful in that job? 2. Explain some environmental issues in your city. What solutions can you suggest*?

2.4. Data analysis

Learners' performance on the EI task was scored for the accuracy of the target structures. Following Erlam (2006), learners' repetition was granted 1 point when obligatory occasions for the target structures were created and the structures were repeated correctly. They did not receive any credit when they repeated the stimulus incorrectly, when they avoided repetition, and when they corrected their initial incorrect production. Learners' self-correction was assumed to stem from their explicit knowledge and was therefore not given any points (see Ellis, 2006). The researcher and his colleague rated 20 percent of the data. The interrater reliability was .98. The rest of the data were coded by the researcher alone. Learners' performance on each structure was converted into a percentage. The internal reliability was .81 for third

person '-s', .63 for simple past '-ed', .58 for comparative -er, and .75 for plural '-s'. The reliability for the total EI items was 0.90.

Accuracy ratios were used to score learners' performance on the storytelling tasks. Learners' scores for the correct use of third person '-s' and simple past '-ed' in the obligatory occasions were calculated and then divided into the sum of correct and incorrect use of these structures. Like the EI task, learners' self-correction was scored as incorrect. An obligatory occasion analysis was also carried out for the picture description tasks to calculate the accuracy ratio for the use of plural '-s' and comparative '-er'. To estimate interrater reliability, 20% of the data in the storytelling and picture description tasks were coded by the researcher and researcher's colleague. The reliability was .88, .90 .96 and .98, for the third person '-s', simple past '-ed', comparative '-er' and plural '-s', respectively. After the points of disagreement were discussed, the rest of the data were analyzed by the researcher alone. The overall scores of the learners on each structure both in the storytelling and picture description tasks were also converted into a percentage.

As for the sample IELTS listening tests, the average score on the three tests was calculated for each learner and converted into a percentage. The average alpha coefficient for the internal reliability of the tests was .87. For the speaking test, the TOEFL independent speaking rubric was used to score learners' responses (https://www.ets.org/s/toefl/pdf/toefl_speaking_rubrics.pdf). Learners' performance was scored 0-4 and then converted into a percentage for the sake of comparison. The researcher and his colleague scored all speaking data. The interrater reliability was .86.

3. Results

Learners completed the storytelling tasks, the EI, and the picture description task in that order. They completed the listening tests throughout the semester and the speaking test at the end of the semester. Table 1 shows the descriptive statistics for the participants in each test.

As for the EI, the average accuracy scores of the learners on the target structures was notably high (M = 73.31). The highest mean score in the EI was recorded for comparative '-er', where the learners' performance was at the ceiling level (M = 86.67), and the lowest score was recorded for plural '-s' (M = 63.00). When the performance of the learners on different tests was compared, it turned out that they were more successful in the EI than in the picture description and storytelling tasks with regard to the three structures i.e., third person '-s', simple past '-ed' and comparative '-er' (M = 70.15 vs. M = 55.65, M = 75.07 vs. M = 66.58 and M = 86.67 vs. M = 72.85, respectively), while the opposite was true with regard to plural '-s' (M = 63.00 vs. M = 72.60).

Table 1. Descriptive statistics for the EI and the other measures

Test	Mean	SD
EI total	73.31	16.92
EI third person '-s'	70.15	22.62
EI past '-ed'	75.07	14.49
EI comparative '-er'	86.67	14.55
EI plural '-s'	63.00	24.09
Storytelling third person '-s'	55.65	26.85
Storytelling simple past '-ed'	66.58	24.24
Picture description comparative '-er'	72.85	22.81
Picture description plural '-s'	72.60	21.54
Listening	47.87	13.00
Speaking	73.57	13.98

Principal component analysis (PCA) was performed on the data. The KMO value was .85, which indicated that the sampling was adequate. The PCA resulted in a single component that accounted for 57% percent of the total variance. Table 2 provides the loadings of the variables on this component. As shown, the loadings on the component varied from .57 (good), and .63, .66 (very good) to .74, .77, .80, .81, .83, .85 (excellent) according to Comrey and Lee's (1992) criteria.

Table 2. Loadings for the principal component analysis

Measures	Component 1
EI third person '-s'	.83
EI past '-ed'	.85
EI comparative '-er'	.74
EI plural '-s'	.81
Storytelling third person '-s'	.77
Storytelling simple past '-ed'	.63
Picture description comparative '-er'	.57
Picture description plural '-s'	.80
Speaking	.83
Listening	.66

Table 3 shows the results of Pearson correlations recorded between learners' EI scores on the four structures, and their scores on the storytelling and picture description tasks. To avoid the outliers' effects on correlations, the data related to one participant was removed from the analyses. As is shown, the magnitude of correlations was large for the three structures (r = .63, r =. 63, r = .65 for the third person '-s', simple past '-ed' and plural '-s', respectively). The recorded correlation was, however, somewhat lower for comparative '-er' (r = .43).

Table 3. Correlations between scores of learners on the EI, storytelling and picture description tasks

Measure	Storytelling third person '-s'	Storytelling past '-ed'	Picture description comparative '-er'	Picture description plural '-s'
EI third person '-s'	.632**			
EI past '-ed'		.634**		
EI comparative '-er'			.433**	
EI plural '-s'				.657**

**p < 0.006

Pearson correlations were also carried out between the learners' total EI scores on the four target structures, and their listening and speaking skills. Table 4 shows that the EI correlated highly with the speaking and listening tests, and that the correlation was higher for the speaking than the listening tests (r = .73 vs. r = .61).

Table 4. Correlations between the scores of learners on the EI and the speaking and listening tests

Measure	EI total	Listening	Speaking
EI total	1		
Listening	.615**	1	
Speaking	.739**	.648**	1

** p<0.001

4. Discussion

As Table 2 shows, all measures used in the study were loaded on a single factor indicating that these measures may tap into a common construct. The fact that these measures were all meaning-focused and time-pressured implied that this construct is likely to be the learners' implicit knowledge. The research question asked if EI construct validity as a measure of implicit knowledge is dependent on target structures. As the data analysis results showed, the magnitude of correlations for the third person '-s' plural '-s' and simple past '-ed' was considerably higher than the one recorded for comparative '-er'. This implies that EI might be a better measure of implicit knowledge of some language structures than others.

As Table 1 shows, for the comparative structure, the learners performed more poorly on the picture description task than they did on the EI (M = 72.85 vs. M = 86.67). This is also evident in the data analysis of seven comparative adjectives which were identical in the picture description task and the EI. Table 5 shows the number of obligatory occasions created for each adjective in both tests, and the number of times that the adjectives were used correctly in the EI and picture description task. As the table shows, when the learners displayed discrepancies in terms of producing the comparative adjectives in the two tests, they were more likely to get it right in the EI than in the picture description task. These findings indicate that the picture description task resembled more real-life language use than the EI.

Table 5. Identical adjectives and the percentage of their correct and incorrect uses in the EI and picture description task (PDT)

Adjectives shared between EI and PDT	adjectives used in both tests (max 40)	correct use in both tests	correct use in EI only	correct use in the PDT only	incorrect use in both tests
Smarter	31	22	9	0	0
Nicer	34	19	10	3	2
Quicker	37	25	10	1	1
Easier	32	27	4	0	1
Funnier	36	21	9	3	3
Cleaner	37	21	8	3	5
Stronger	35	22	9	2	2

But what could account for the superior performance of learners on the EI compared to the picture description task for the comparative structure? The answer may lie in the nature of the interaction between explicit and implicit knowledge and how this is mediated by the features of language structures. The interaction between implicit and explicit knowledge is less straightforward for comparative '-er' than for the other structures used in the study. Compared to comparative '-er', third person '-s', plural '-s and somewhat '-ed' have more conceptual clarity (a single morpheme is used to establish form and meaning connection, see Ellis, 2006) and high regularity (the rules are highly generative). The correct application of comparative '-er', however, requires learners to make a choice between two morphemes, '-er' and 'more', a cognitively demanding task especially when adjectives contain two syllables. Each adjective needs to go through a system with a binary outcome (e.g., funnier or more funny, happier or more happy, fatter or more fat, etc.). Such conscious rule application needs to be initially carried out for each adjective separately so that the follow-up exposure to L2 data can feed and consolidate the implicit knowledge.

Following the above-mentioned argument, learners may vary with regard to their implicit knowledge of this structure. Some learners may have already established the correct associations between language features in their implicit memory. For these learners, the initial rule-based generation of language structures has successfully led to the follow-up instance-based language use (Skehan, 1998). Another group of learners may use comparative rules in more generative yet inaccurate ways. They may, for example, widen the scope of the comparative marker 'more' to instances where it does not apply (e.g., more easy), or they may use double comparative markers (e.g., more easier). In such cases, the association between language features is established but not in a correct way. For a third group of learners, the correct association between language features might be established for some comparative adjectives (e.g., taller and more interesting) but not for others (e.g., sadder or fatter). These learners may need to revert to the explicit rule-based application of comparatives that they have stored in their declarative memory, once the occasions for the use of novel comparative adjectives arises in communication.

Arguably, it is the third group of learners that benefit more from the current EI design features than purely productive tasks such as picture description tasks. This is so since EI stimuli can act as models to create initial connections or strengthen the existing ones between language features. Learners do not need to decide which type of comparative makers, 'er' or 'more', should be used to produce the new adjective as it is already done for them by the stimulus sentences. The subsequent production, therefore, goes through the filter of newly activated rules. For the other three structures, however, because of the generative nature of rules, the number of learners in the third group is either non-existent or highly limited making the effect of the exposure to the EI stimuli a less decisive factor. The EI, therefore, turns out to offer more likely measures of implicit knowledge for third person '-s', plural '-s', simple past '-ed' but not as much for comparative '-er'.

It is also worth noting that as Table 4 shows the correlation recorded between learners' overall performance on the EI and speaking tests (r = .73) was higher than any correlations recorded between learners' accuracy scores on the EI, and other oral production tests for the four target structures (r = .63, r = .63, r = .65 and r = .43). These findings suggested that EI might be more sensitive to measure aural/oral language proficiency than implicit grammatical knowledge of specific structures. The reason may lie in the fact that the processes involved in the successful performance on EI demand a broader linguistic knowledge than simply morpho-syntactic one. Yan, Maeda, Lv, and Ginther (2016) also found that EI is more responsive when it is used to measure general language proficiency than knowledge of specific linguistic features, such as grammatical, phonological or lexical.

The present study investigated the construct validity of EI as a measure of implicit knowledge. It employed a different design from the previous studies in that it targeted four English structures using a range of test items for each structure. This made it possible to examine whether EI is more likely a measure of implicit knowledge of some structures than others. Furthermore, to take care of the instructional issues related to ungrammatical items, only grammatical statements were employed. The findings with regard to three structures (i.e., simple past, third person '-s' and plural making) provided further evidence in favor of EI as a measure of implicit knowledge. These findings are in line with those of R. Ellis (2005), Erlam (2006), Bowles (2011), Zhang (2014), Sarandi (2015) and Kim & Nam (2017) where it was also found that the EI tasks correlated highly with other time-restricted tests and/or they loaded on a same factor. The findings, however, contradicted with the ones reported by Suzuki and DeKeyser (2015) and Spada et al. (2015). They found that the EI scores significantly correlated with measures of explicit knowledge and not with other measures of implicit knowledge (an oral narration task in Spada et al., 2015 and a SRT task in Suzuki and DeKeyser, 2015).

As mentioned earlier, an important point with regard to the studies whose finding either did not find conclusive evidence in favor of EI as a measure of implicit knowledge (Spada et al., 2015) or challenged it altogether (Suzuki & DeKeyser, 2015; Suzuki, 2019) is that they informed learners of the existence of ungrammatical

statements and required them to correct these statements. Therefore, it is likely that the type of instruction employed in these studies may have raised the test takers' attention to language form and render the test from becoming a purely implicit knowledge measure into a test of automated explicit knowledge as suggested Suzuki and DeKeyser (2015). The learners' post-EI questionnaires revealed that this might be the case. In Spada et al. (2015), for example, almost 40% of learners claimed that they somehow paid attention to language form when they completed the EI task. In Suzuki (2019), 70% of participants claimed that they attended to language errors when they were repeating the sentences. Future studies can examine the correlation between EI tasks that contain only grammatical statements with other tests of implicit learning such as SRT or word monitoring tasks. They could also examine which parts of the brain are activated when learners are required to simply repeat grammatical statements.

As far as learners' comparative scores are concerned, however, the findings impose some limitations on the range of language structures for which EI can produce a good measure of implicit knowledge. It suggests that EI is structurally sensitive and its potential for measuring implicit knowledge may not apply to all language features to a similar extent. In other words, EI might offer more fine-grained estimates of implicit knowledge for some language structures than others. The present study examined four of these structures. Future studies need to examine other language structures with adequate number of test items to ascertain the language structures for which EI is a good estimate of implicit knowledge.

5. Conclusions

The validity of any tests is based on its ability to measure what it intends to measure (McDade, Simpson, & Lamb, 1982). While the results of the principal component analysis indicated that EI, in general, may offer a valid measure of learners' L2 implicit knowledge, the results of correlation analyses highlighted the structural sensitivity of this test. The discrepancy between the magnitudes of the correlations recorded for the four target structures suggested that the likelihood of EI offering valid and reliable measures of implicit grammatical knowledge may vary according to language structures. The EI in the current study was more sensitive to tap into learners' implicit knowledge of third person '-s', simple past '-ed' and plural '-s', but it was less likely to do so with regard to comparative '-er'.

Like all studies, the present study also suffers from some limitations. First, the majority of the participants had considerable prior knowledge with regard to the target structures. The average score of the learners on all four structures was above 50% in all measures used in the study. Interestingly, the opposite was reported in Spada et al. (2015), where the average score of the learners for passive structures was only 15% on the EI and 33% on the storytelling task. It is advisable that future EI validity studies employ a more heterogeneous group of participants with a wider level of prior knowledge on target structures so that learners' poor or strong performance

does not inflate the correlation. Second, even though the descriptive statistics showed that the picture description task was more demanding than the EI, the possibility still existed that the use of adjectives at the top of each slide prompted some learners to invoke and apply their explicit knowledge. A post-test questionnaire could have provided some information on whether learners' attention was drawn to language forms at any stages of task completion. Third, no explicit knowledge measures, such as untimed grammaticality judgement tests, or meta-linguistic tests were used in the study. The application of these measures could have revealed whether they converge on factors different from the implicit measures. Fourth, it is recommended that future studies recruit a larger population to produce more generalizable findings and ensure that test results are not affected by individual factors. Further studies are also required to examine how methodological variations (e.g. the use of ungrammatical statements, the different instruction related to these statements, the different amount of time delay between listening and repeating stimuli sentences, etc.) can affect learners' scores on EI and change its likelihood to tap into implicit knowledge.

To conclude, the findings of the present study suggest that EI can produce a likely measure of implicit grammatical knowledge, but its potential to do that is mediated by the features of target structures. The structural sensitivity of EI has important implications for SLA researchers. They are advised to use EI to measure implicit grammatical knowledge on the language structures for which validating studies have been carried out and promising outcomes have been achieved. They should, however, refrain from making bold claims regarding the changes in learners' implicit grammatical knowledge of language structures for which validating data have not yet been obtained.

Acknowledgements

The earlier version of this paper was presented at the SLRF 2018 in Montreal. I would like to express my gratitude to Prof. Emma Marsden for her helpful comments on the initial draft of this paper. I would also like to thank two anonymous reviewers for their excellent feedback. All remaining errors are my own responsibility.

The Conflict of Interest Statement

In line with the statement of Committee on Publication Ethics (COPE), we hereby declare that we had no conflicting interests regarding any parties of this study.

References

Baddeley, A. (2015). Working memory in second language learning. In Z. Wen, M. Borges & A. McNeill (Eds.), *Working memory in second language acquisition and processing*. Bristol: Multilingual Matters.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158-173.

Baten, K., & Cornillie, F. (2019). Elicited imitation as a window into developmental stages. *Journal of the European Second Language Association, 3*(1), 23–34.

Berry, P. B. (1976). Elicited imitation of language: Some ESNS population characteristics. *Language and Speech, 19*(4), 350-362.

Bialystok, E. (1978). A theoretical model of second language learning. *Language Learning, 28*(1), 69-84.

Bowles, M. A. (2011). Measuring implicit and explicit linguistic knowledge: What can heritage language learners contribute? *Studies in Second Language Acquisition*, *33*(2), 247-271.

Cambridge IELTS 2 (2000). *Examination papers from the university of Cambridge local examination syndicate*. Cambridge: Cambridge University Press.

Cambridge IELTS 4 (2005). *Examination papers from the university of Cambridge local examination syndicate*. Cambridge: Cambridge University Press.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Erlbaum.

Connell, P. J., & Myles-Zitzer, C. (1982). An analysis of elicited imitation as a language evaluation procedure. *Journal of Speech and Hearing Disorders, 47*(4)*,* 390-396.

Eisenstein, M., Bailey, N., & Madden, C. (1982). It take two: Contrasting tasks and contrasting structures. *TESOL Quarterly 16*(3), 381-93.

Ellis, N. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, *27*(2), 305-352.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, *27*(2), 141-172.

Ellis, R. (2006). Modeling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge. *Applied Linguistics, 27*(3), 431-463.

Ellis, R. (2007). The differential effects of corrective feedback on two grammatical structures. In. A. Mackey (Ed.), *Conversational interaction in SLA: A collection of empirical studies*, (pp. 339-360). New York, NY: Oxford University Press.

Ellis, R. (2008). Investigating grammatical difficulty in second language learning: Implications for second language acquisition research and language testing. *International Journal of Applied Linguistics, 18*(1), 4-22.

Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. In Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J., & Reinders, H. (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3-25). Bristol: Multilingual Matters.

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition, 28(2),* 339-368.

Ellis, R., & Roever, C. (2018). The measurement of implicit and explicit knowledge. *The Language Learning Journal,* 1-16.

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics*, *27*(3)*,* 464-491.

Graham, C. R., Lonsdale, D., Kennington, C., Johnson, A., & McGhee, J. (2008). Elicited imitation as an oral proficiency measure with ASR scoring. *Proceedings of the 6th International Conference on Language Resources and Evaluation* (pp.1604-1610). Paris: European Language Resources Association.

Granena, G. (2016). Elicited imitation as a measure of implicit L2 knowledge: The role of working memory and short-term memory capacity. In G. Granena, D. Jackson & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 185-204). Amsterdam: John Benjamins.

Kim, J. E., & Nam, H. (2017). Measures of implicit knowledge revisited. *Studies in Second Language Acquisition*, *39*(3), 431-457.

McDade, H., Simpson, M., & Lamb, D. (1982). The use of elicited imitation as a measure of expressive grammar: A question of validity. *Journal of Speech and Hearing Disorders, 47*(1), 19-24.

Naiman, N. (1974). The use of elicited imitation in second language acquisition research. *Working Papers on Bilingualism, 2*, 1-37.

Paradis, M. (2009) *Declarative and procedural determinants of second languages.* Amsterdam: John Benjamins Publishing Company.

Pienemann, M. (1998). Language processing and second language development: Processability theory. Amsterdam: Benjamins.

Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, *29*(6), 633-654.

Sarandi, H. (2015). Reexamining elicited imitation as a measure of implicit grammatical knowledge and beyond…?. *Language Testing, 32*(4), 485-501.

Scott, M. L. (1994). Auditory memory and perception in younger and older adult second language learners. *Studies in Second Language Acquisition*, *16*(3), 263-281.

Skehan, P. (1998). A cognitive approach to language learning. Oxford: Oxford University Press.

Smith, C. S. (1970). An experimental approach to children's linguistic competence. In J.R. Hayes (Ed.), *Cognition and the development of language* (pp.109-133). New York, NY: John Wiley.

Spada, N., Jessop, L., Tomita, Y., Suzuki, W., & Valeo, A. (2014). Isolated and integrated form focused instruction: Effects on different types of L2 knowledge. *Language Teaching Research, 18*(4), 453-473.

Spada, N., Shiu, J. L. J., & Tomita, Y. (2015). Validating an elicited imitation task as a measure of implicit knowledge: Comparisons with other validation studies. *Language Learning*, *65*(3), 723-751.

Suzuki, Y. (2019, December). The neural foundations of explicit and implicit knowledge: An fMRI study. Paper presented at the meeting of Neurocognitive Foundations of Foreign Language Learning: ESRC UK-Japan Project, University College London.

Suzuki, Y., & DeKeyser, R. (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge. *Language Learning*, *65*(4), 860-895.

Swan, M. (1976). *Understanding ideas: Advanced reading skills*. Cambridge: Cambridge University Press.

Toth, P. D., & Guijarro-Fuentes, P. (2013). The impact of instruction on second-language implicit knowledge: Evidence against encapsulation. *Applied Psycholinguistics, 34*(6), 1163-1193.

Underhill, N. (1987). *Testing spoken language: A handbook of oral testing techniques.* Cambridge: Cambridge University Press.

Yan, X., Maeda, Y., Lv, J., & Ginther, A. (2016). Elicited imitation as a measure of second language proficiency: A narrative review and meta-analysis. *Language Testing*, *33*(4), 497-528.

Yasuyao, T., Wataru, S., & Lorena, J. (2009). Elicited imitation : Toward valid procedures to measure implicit second language grammatical knowledge. *TESOL Quarterly*, *43*(2), 345-350.

Zhang, R. (2014). Measuring university-level L2 learners' implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition*, *37*(3), 457-486.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the Journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (CC BY-NC-ND) (http://creativecommons.org/licenses/by-nc-nd/4.0/).