Measuring implicit and explicit linguistic knowledge: What can heritage language learners contribute?

Melissa A. Bowles
University of Illinois at Urbana-Champaign


Address for correspondence
Melissa Bowles
Department of Spanish, Italian & Portuguese
University of Illinois at Urbana-Champaign
4080 Foreign Languages Building, MC-176
707 South Mathews Avenue
Urbana, IL 61801
USA

Phone: (217) 244-3944
Fax: (217) 244-8430
E-mail: bowlesm@illinois.edu
Abstract

Two major goals of second language acquisition (SLA) research are to define and describe learners’ second language (L2) linguistic knowledge and to explain how this knowledge develops over time (R. Ellis, 2004, 2005). One difficulty in measuring L2 learners’ implicit and explicit language knowledge is that their interlanguage contains both types of knowledge, making measurement tricky. The purpose of this study was to validate that the battery of tests reported in Ellis (2005) indeed provide relatively separate measures of implicit and explicit language knowledge. Whereas Ellis (2005) tested only L2 learners (of English), this study tested both L2 and HL learners (of Spanish). Results showed that battery of test scores loaded on a two-factor model, as in Ellis (2005). Furthermore, as a group the L2 learners scored highest on the metalinguistic knowledge test, whereas the HL learners as a group scored lowest on the metalinguistic knowledge test and highest on the oral narration test. These results provide construct validity for the battery of tests reported in Ellis, as tested on a population of HL learners who have little explicit knowledge by virtue of the environment in which they acquired Spanish.
Measuring implicit and explicit linguistic knowledge: What can heritage language learners contribute?

Two major goals of second language acquisition (SLA) research are to define and describe learners’ second language (L2) linguistic knowledge and to explain how this knowledge develops over time (R. Ellis, 2004, 2005). Many theoretical issues, including the debate over how linguistic knowledge changes as proficiency increases (the interface and non-interface position), hinge on the ability to measure learners’ explicit and implicit language knowledge. Yet to date, there has been little research on how learners’ explicit and implicit interlanguage knowledge can be measured (R. Ellis et al., 2009). One difficulty in measuring L2 learners’ implicit and explicit language knowledge is that their interlanguage contains both types of knowledge, making measurement tricky. This study explores the validity of the tests designed by Ellis (2005) to provide separate measures of implicit and explicit knowledge, with a population of heritage learners of Spanish (Valdés, 2001), who have very little explicit knowledge by virtue of the fact that they learned their heritage language naturalistically and have had little formal instruction in it.

In addition to validating the measures of explicit and implicit knowledge, by having a sample of L2 learners and a sample of HL learners complete the same tests of linguistic knowledge, this study can help to provide a more complete picture of the ways in which HL learners may be similar to, or different than, L2 learners. This is not a trivial question, since it has implications both for linguistic theory and for language pedagogy.
Positions on the Relationship Between Implicit and Explicit Language Knowledge

In SLA, there are three main theoretical positions taken with regard to the relationship between implicit and explicit language knowledge. Each is described below. According to the non-interface position, implicit and explicit L2 knowledge are acquired differently (Hulstijn, 2002; Krashen, 1981), are thought to be located in or accessed by different areas of the brain (Paradis, 1994), and are accessed by different processes (N. Ellis, 1993). The hallmark of the non-interface position is that the two knowledge types are completely separate, to such an extent that explicit knowledge can never become implicit. This is the position espoused by Krashen (1981), and his learning-acquisition hypothesis. According to the non-interface position, explicit language knowledge will always remain explicit, despite years of exposure, practice, and proficiency in the language.

The strong interface position is in polar opposition to the non-interface position. Specifically, this position holds that explicit knowledge can be derived from implicit knowledge, and even more importantly that explicit knowledge can become implicit knowledge with enough practice. According to the strong interface position, learners can first learn a rule as declarative (explicit) knowledge and then through repeated use and practice can convert it into an implicit representation. The interface position was first formally advanced by Sharwood Smith (1981) and has subsequently been promoted by DeKeyser (1998).

The weak interface position similarly posits that explicit knowledge can become implicit; however, it is not as categorical as the strong interface position in its formulation. This position stipulates some constraints on the circumstances under which explicit knowledge can become implicit. One version of the weak interface position, espoused by R. Ellis (1993), proposes that the conversion of explicit to implicit knowledge can occur through practice, but only if and when
the learner is developmentally ready to acquire the linguistic form in question. A second version, espoused by N. Ellis (1994), maintains that explicit knowledge contributes indirectly to the creation of implicit knowledge because “declarative rules can have ‘top-down’ influences on perception” (p. 16). The third and final version of the weak interface position, espoused by Schmidt and Frota (1986) and Sharwood Smith (1981), propose that learners use their explicit linguistic knowledge to produce (presumably planned) output, which in turn becomes input for their implicit system.

These positions have been debated in the SLA literature for decades. However, to date there is no conclusive evidence favoring one position over the others. One main reason for this is that although studies have sought to examine the relationship between explicit and implicit linguistic knowledge for many years (Hulstijn & Hulstijn, 1984; Seliger, 1979; Sorace, 1985), most have not been concerned with establishing a valid and reliable set of measurement instruments to do so (R. Ellis, 2004, 2005). It was for this reason that R. Ellis (2004, 2005) set out to examine the constructs of implicit and explicit linguistic knowledge and to develop a rigorous set of instruments designed to measure them.

Features Proposed to Distinguish Between Types of Linguistic Knowledge

In his recent research, R. Ellis (2004; 2005) has considered features that could be used to distinguish between the constructs of implicit and explicit linguistic knowledge. His review of theoretical and empirical work in SLA revealed seven distinguishing features – awareness, type of knowledge, systematicity, accessibility, use of knowledge, self-report, and learnability. Each is described briefly below.
Awareness

Children learning their L1 initially demonstrate intuitive knowledge about what is possible and impossible in their language. Only once they reach a certain age and developmental level (at or beyond age 5, or school age) do children start to demonstrate some ability to analyze their language, or have metalinguistic knowledge and conscious awareness of why structures are grammatical or ungrammatical (Karmiloff-Smith, 1979). In a recent study with English-speaking L2 learners of Spanish and German, the amount of language-learning experience in formal (classroom) settings was found to be the strongest predictor of metalinguistic knowledge (Roehr & Ganem-Gutierrez, 2009).

Type of Knowledge

In psychology, there is a long-standing distinction between declarative and procedural knowledge (Anderson, 1983). Declarative knowledge is factual in nature and accessed slowly, whereas procedural knowledge is highly automated and accessed quickly. Declarative and explicit knowledge are parallel, as are procedural and implicit knowledge.

Systematicity and Certainty of L2 Knowledge

In cognitive psychology, Reber and his colleagues (Reber, Walkenfield, & Hernstadt, 1991) have demonstrated evidence to support the idea that implicit knowledge is less variable, or more consistent, than explicit knowledge. Along similar lines, SLA researchers have claimed that learners’ interlanguages (their implicit knowledge) are highly systematic (Tarone, 1988). Explicit knowledge, on the other hand, tends to be more anomalous, and much less consistent (Sorace, 1985). Taken together, these pieces of evidence suggest that implicit knowledge is more structured than explicit knowledge and that learners are more confident in their judgments when those are based on implicit, rather than explicit, knowledge.
Type of Processing and Accessibility of Knowledge

Implicit knowledge involves automatic processing whereas explicit knowledge entails controlled processing. However, the type of processing cannot be directly measured; rather, type of processing is generally inferred from the speed with which the knowledge is accessed. Some SLA researchers (e.g., Hulstijn, 2002) have indicated that practice will only “speed up the execution of algorithmic rules to some extent” (p. 211) and that there remains a fundamental difference between automated explicit knowledge and implicit knowledge in terms of their accessibility. Others (e.g., DeKeyser, 2003) have indicated that explicit knowledge can be fully automatized and thereby become functionally equivalent to implicit knowledge.

Use of L2 Knowledge

Bialystok (1982) provided evidence that the use of explicit or implicit knowledge varies according to the demands of the tasks learners are asked to perform. Bialystok distinguished task demands in terms of binary dimensions of analysis and control. For example, she provided evidence to show that a written task that requires learners to detect and then correct errors is [+analyzed, -automatic] and taps into explicit knowledge, whereas an aural task requiring participants to follow the same instructions is [+analyzed, +automatic], tapping implicit knowledge.

Self-report

Implicit knowledge cannot be verbalized, whereas explicit knowledge is potentially verbalizable. Although learners are likely unable to verbalize the entire contents of their explicit knowledge, they are able to verbalize at least a subset of it.
Learnability

Some researchers have proposed that there are maturational effects on implicit and explicit knowledge (DeKeyser, 2002). More specifically, Bialystok (1994) claimed that “explicit knowledge can be learned at any age” (p. 566), but there are age-related constraints on L2 learners’ ability to acquire implicit knowledge.

In summary, implicit knowledge can be described as intuitive, procedural, variable in a limited and systematic way, automatic, and available for use in fluent, spontaneous language use. Implicit knowledge cannot be verbalized, and there is debate over the extent to which implicit knowledge can be learned in adulthood. Explicit knowledge, on the other hand, is conscious, declarative, highly variable and only accessible through controlled processing. Therefore, explicit knowledge is largely unavailable in spontaneous language use, although learners can rely on it for monitoring in planned language use. Explicit knowledge can be verbalized, therefore entailing at least some degree of metalinguistic knowledge. Like any other type of declarative knowledge, explicit language knowledge can be learned regardless of age. When linguists and L1 acquisition researchers refer to linguistic knowledge, they typically refer to linguistic competence (i.e., implicit knowledge). But in many empirical SLA studies, learners’ knowledge of targeted structures in a language is often measured using tests that favor the use of explicit knowledge.

Review of Ellis (2005)

Ellis (2005) created a battery of five English language tests designed to tap explicit and implicit knowledge by manipulating the dimensions just mentioned. The tests included (a) an oral imitation test containing both grammatical and ungrammatical sentences, (b) an oral narration test, (c) a timed grammaticality judgment test (GJT), (d) an untimed GJT testing the
same grammatical structures, and (e) a metalinguistic knowledge test. L2 learners of English at a range of proficiency levels ($n = 91$) and a group of English native speakers ($n = 20$) took the battery of tests, and their responses were analyzed using exploratory factor analysis. Results indicated that the scores from the oral imitation test, oral narration test, and timed GJT loaded on one factor, whereas the scores from the metalinguistic knowledge test and scores from the ungrammatical sentences on the untimed GJT loaded on a second factor. Ellis interpreted the two factors as corresponding to implicit and explicit knowledge, respectively.

Ellis’ (2005) use of an exploratory factor analysis (EFA) rather than a confirmatory factor analysis (CFA) was criticized by Isemonger (2007) on the grounds that CFA is recommended in cases where a priori hypotheses are being tested (Byrne, 2001; Kline, 1994; Thompson, 2004; Thompson & Daniel, 1996). Since Ellis’ study was indeed verificational rather than exploratory in nature, and given that he did, in fact intend to test a number of hypotheses, Ellis and Loewen (2007) subsequently re-analyzed the data from the original (2005) study using CFA. The CFA confirmed the study’s original findings, that a two-factor model with the oral imitation test, the oral narrative test, and the timed GJT loading on one factor and the untimed GJT and the metalinguistic knowledge test loading on a second factor was an appropriate fit, with a $\chi^2$ of 1.191, whereas an alternate model proposed by Isemonger (2007), was not a good fit.

The study

The purpose of this study was to validate that the battery of tests reported in R. Ellis (2005) indeed provide relatively separate measures of implicit and explicit language knowledge. Whereas Ellis (2005) used a battery of tests in English and tested them on both native speakers and L2 learners of English, this study includes a battery of tests designed closely following R. Ellis’ guidelines but in Spanish. Additionally, the tests were run on three groups - Spanish native
speakers, L2 learners, and HL learners. If the tests tap relatively separate pools of linguistic knowledge, then the L2 learners (who had learned Spanish in a classroom setting after puberty and had no more than 1 week of study-abroad experience in a Spanish-speaking country) should score higher on tests of explicit knowledge, whereas the HL learners (who had learned Spanish naturally since birth and had less than 2 years of formal classroom instruction in Spanish) should score higher on tests of implicit knowledge. The native speakers, who were monolingually-raised in Spanish and had completed at least secondary school in Spanish, are expected to perform near ceiling on the oral imitation, oral narration, and both timed and untimed GJTs. Only a small amount of variation is expected, since native speakers’ knowledge should be relatively stable. However, they are not expected to perform near ceiling on the metalinguistic knowledge test, since it relies on grammatical terminology and rules that the native speakers most likely do not access frequently. Accordingly, a larger percent of variation (i.e., larger standard deviations in scores) are expected from one native speaker to the next on this test than on the other four.

The specific research questions are as follows:

1. Do scores on the battery of five tests load on two factors, in the way that they did in Ellis (2005)?

2. Does a comparison of HL and L2 learners’ test scores provide construct validity evidence for the tests (i.e., do L2 learners score higher on tests of explicit knowledge, whereas HL learners score lower on those measures because of their naturalistic language acquisition process)?
Participants

A total of 30 participants completed the battery of five tests. Native speakers (n =10) were all monolingually-raised in Spanish-speaking countries and had immigrated to the U.S. as adults. Among the native speakers, there were 5 males and 5 females, ranging in age from 24-36, with an average age of 30.66. They were from a broad range of countries of origin: 1 each from Puerto Rico, Costa Rica, Colombia, Chile, and Mexico, 2 from Peru, and 3 from Spain. They took the battery of tests primarily (1) to ensure that the usage in the test items was normative, and (2) to establish baseline mean reaction times on each sentence used in the timed grammaticality judgment test.

Both the L2 and HL learners were enrolled in intermediate-level Spanish language classes at the university where the study was conducted. However, the language background profiles of the two groups were quite different from each other. The L2 learners (n = 10) all reported that they had been raised in monolingual English-speaking households and had been exposed to Spanish in a classroom setting at or after puberty. Among the L2 learners, there were 6 females and 4 males, ranging in age from 19 to 22, with an average age of 20.2. They reported that they had not had any significant study-abroad experience (in order to qualify for the study, L2 learners could have spent a maximum of 1 week in a Spanish-speaking country). On average, the L2 learners had taken 6 years of Spanish courses (including secondary and post-secondary courses). The HL learners of Spanish (n =10), on the other hand, reported that they had been raised in a bilingual English/Spanish household and that they routinely used Spanish throughout childhood and into the present day with at least one family member. It is important to note that all HL learners were English-dominant, given that they had been schooled in the U.S. and were attending a university where English was the language of instruction. In order to qualify for the
study, HL learners also had to have had two years or less of formal instruction in Spanish. This was done to ensure that the HL learners would have as little explicit knowledge as possible, while still being literate in Spanish. Among the HL learners, there were 7 females and 3 males, ranging in age from 19 to 23, with an average age of 20.6. They had taken an average of 1.3 years of Spanish courses (including secondary and post-secondary courses).

Test Content

Just as in R. Ellis (2005), the tests in this study were designed to provide measures of the learners’ knowledge of 17 different grammatical structures. The particular structures were selected primarily because they have been shown to be problematic for both L2 and HL learners of Spanish. Ample research has shown that both L2 and HL learners of many languages, including Spanish, display morphological variability and have persistent difficulty with nominal and verbal inflection (see Montrul, this issue). In the nominal domain the structures tested were gender agreement (Bruhn de Garavito & White, 2003; Hakansson, 1995; Montrul, Foote, & Perpiñán, 2008; Polinsky, 2008; White, Valenzuela, Kozlowska-Macgregor, & Leung, 2004) and case marking (Montrul & Bowles, 2009; Polinsky, 2006; Song, O’Grady, Cho, & Lee, 1997), both with dative psych verbs and with animate, direct objects. In terms of verbal inflection, several structures involving tense-aspect-mood morphology were tested, including subject-verb agreement, present subjunctive (in both nominal and adverbial clauses) (Montrul, 2007), past (imperfect) subjunctive (Potowski, Jegerski, & Morgan-Short, 2009), perfective vs. imperfective aspect (Montrul, 2002; Silva-Corvalán, 1994), and conditionals. (Complete stimuli for the tests are included in Appendices A-C.)

The selection of problematic structures, including those that are resistant to instruction or that are consistently problematic after years of instruction (e.g., aspect and mood distinctions),
helped to ensure that there was a range of scores on the tests and that learners did not reach ceiling. Both early and late acquired forms were selected, as were structures introduced at a range of instructional levels. For instance, subject-verb agreement with regular present indicative verbs (Examples 1 and 2) is early acquired and taught in beginner-level courses, whereas mood selection (indicative vs. subjunctive) in adverbial clauses (Examples 3 and 4) is late acquired and is taught in intermediate and even advanced-level courses.

(1) Mamá y yo  tomamos  mucho  café.  
Mom and I  drink.1PI.Pres.Ind.  much  coffee  
‘Mom and I drink a lot of coffee.’

(2) *Las  chicas  hablo  inglés.  
The.Fem.Pl  girls  speak.1Sg.Pres.Ind.  English  
‘The girls speak English.’

(3) Cuando  el  chico  llegue  a  casa,  te  llamará.  
When  the.masc.sing  boy  arrive.3Sg.Pres.Subj.  to  house,  you.2Sg.Ci.  
call.3Sg.Future.  
‘When the boy gets home, he will call you,’

(4) *Cuando  Juan  tiene  tiempo,  estudiara  francés.  
When  John  have.3Sg.Pres.Ind.  time,  study.3Sg.Future  French  
‘When Juan has time, he will study French.’

Description of the Tests

*Imitation test*

The oral imitation test consisted of 34 statements, half of which were grammatical sentences containing the target structures (n = 17) and half of which were ungrammatical
sentences containing the target structures. That is, there was one grammatical and one ungrammatical sentence per target structure. The sentences were audio recorded by a female native speaker of Spanish, and the resulting audio files were randomized to prevent ordering effects. The sentences were played one at a time for the participants, who had to first state out loud whether they agreed or disagreed with the propositional content of the statement by saying *Sí* ‘Yes’ or *No*. Immediately after making that judgment, participants were required to repeat the sentence in correct Spanish. Participants’ responses were audio recorded and subsequently analyzed by identifying obligatory occasions for the use of the target structures. Each correctly imitated sentence received a score of 1, whereas each sentence for which the target structure was avoided or attempted but incorrectly supplied received a score of 0. (It should be noted that two of the native speakers produced a grammatically correct restatement of the ungrammatical stimulus *Si tantos jóvenes no habían votado, Barack Obama no habría ganado las últimas elecciones.* ‘If so many young people had not voted, Barack Obama would not have won the recent election.’ without using the target structure. This was counted as correct and awarded 1 point. It was decided that if any learners restated the sentence in that way, they would also receive 1 point. However, learners either did not produce the entire sentence or produced it ungrammatically.) For each participant scores on this test were expressed as the percentage of sentences repeated/restated correctly out of a total maximum of 34.

*Oral narrative test*

The story used in this test was designed to elicit the use of a number of the target structures, including reflexive verbs, subject-verb agreement, present or simple past tense inflectional morphology, gender agreement, and aspect (perfective vs. imperfective). The participants saw the vignettes from “The tricky alarm clock,” which have been used in previous
SLA elicitation tasks (Adams, 2003; Swain & Lapkin, 2002). The vignettes were accompanied by a story written in Spanish to go along with the pictures. Participants read the story and looked at the corresponding pictures twice. Then, the third time, only the pictures remained on the screen and they were asked to retell the story aloud. Their narratives were audio recorded and subsequently transcribed. Again, as with the oral imitation test, an obligatory occasion analysis was carried out to establish the percentage of correct suppliance of each target structure. For each participant scores on this test were calculated by averaging the percentage scores for each structure.

A transcript of one of the HL learners’ oral narrations appears in (5) below:

(5) Un día al amanecer había una chica tumbada en la cama que … que no se quería levantar para la escuela. A las seis en punto en la mañana suena la alarma del despertador pero ella no se quería levantar. Decide dormirse unos minutos más después de que sonó su alarma. A las seis y uno la muchacha se volvió a dormir muy rápido. A las seis y dos en punto de la mañana sale una mano de su despertador y le hizo cosquillas en sus pies. La muchacha se despertó. Se levantó y se fue al baño y se peinaba mientras se lavaba los dientes. Y después se fue a vestir. Rápido estuvo lista para ir a la escuela. Se puso su mochila, justo a tiempo para llegar a la escuela a tiempo.

One day at dawn there was a girl lying in bed who…who didn’t want to get up to go to school. At six o’clock in the morning sharp the alarm from the alarm clock goes off but she didn’t want to get up. She decides to sleep a few minutes more after her alarm went off. At 6:01 the girl went back to sleep very
quickly. At 6:02 in the morning sharp a hand comes out of her alarm clock and it tickled her feet. The girl woke up. She got up and went to the bathroom and brushed her hair while she was brushing her teeth. And then she went to get dressed. Quickly she was ready to go to school. She put on her backpack, just in time to get to school on time.

Timed GJT

The timed GJT consisted of 68 sentences, evenly divided between grammatical and ungrammatical. Each sentence was presented in written form on a computer screen, so there were 4 sentences for each of the 17 grammatical structures. The test was delivered using E-Prime v. 1.2, and sentences were automatically randomized. Participants were required to indicate whether each sentence was grammatical or ungrammatical by pressing one of two keys on the computer keyboard within a fixed time limit. The time limit for each sentence was established on the basis of the native speakers’ average response time for each stimulus. Following Ellis (2005), the average response time for the NS was increased by 20% to allow for the slower processing speed of the learners. (Time limits were rounded to the nearest millisecond.) Specifically, the time allowed for judging the individual sentences ranged from 1742 to 5467 milliseconds. Each item was scored dichotomously as correct or incorrect, with items left unanswered at the end of the time limit scored as incorrect¹. A percentage accuracy score was calculated for each participant.

Untimed GJT

The untimed GJT included the same types of sentences (but different tokens) than the timed GJT. It was also delivered in E-Prime v. 1.2 and sentences were presented in written form
and randomized automatically. Participants were required to indicate whether each sentence was grammatical or ungrammatical, just as they had done in the timed GJT, except that they were instructed to answer at their own pace since the test had no set time limit.

*Metalinguistic knowledge test*

The metalinguistic knowledge test was based on the test designed in Alderson, Clapham, and Steel (1997) and reported in R. Ellis (2005). It consisted of a two-section untimed computerized multiple-choice test. The first section contained 17 ungrammatical sentences, one per target structure, and instructed participants to select the rule that best explained why each sentence was ungrammatical. In the second section of the test, participants read a short text written for beginning-level learners of Spanish (Appendix C) and were instructed to find examples of 21 specific grammatical features from the text (e.g., verb in the imperfect, subject pronoun). A total percentage accuracy score was calculated out of the total number of points possible.

These tests, like those in Ellis (2005), were designed to provide a relatively separate measure of either implicit or explicit knowledge, as described below. Specifically, the oral imitation test and the oral narration test were predicted to measure implicit knowledge because the participants would need to rely mainly on feel or intuition, they would be under pressure to perform in real time, they would be focused primarily on communicating meaning, rather than on grammatical form, and they would have no reason to access their metalanguage. In contrast, the metalinguistic knowledge test was predicted to measure explicit knowledge because it involved a high degree of awareness, had no time pressure, focused attention on grammatical form, and, obviously, required the use of explicit knowledge. Both of the GJT's required participants to focus attention primarily on grammatical form by their very nature, but the two differed in that
the timed GJT was predicted to measure primarily implicit knowledge, whereas the untimed GJT was predicted to measure primarily explicit knowledge. The timed task encouraged the use of feel, and it was time-pressured, allowing participants little access to explicit knowledge. On the other hand, the untimed GJT encouraged a high degree of awareness and was self-paced, having no time pressure. These features of the untimed GJT distinguish it from the timed GJT, since participants could take their time and use explicit knowledge if they so wished.3

As in R. Ellis (2005), the tests were administered in a consistent order, with those designed to tap implicit knowledge administered before those designed to tap explicit knowledge. This way, the explicit knowledge tests would not prime learners. Specifically, the oral imitation test was administered first, followed by the oral narration test, the timed GJT, the untimed GJT, and finally, the metalinguistic knowledge test.

Results

Given the methodological issues raised by Isemonger (2007), a confirmatory factor analysis, rather than an exploratory factor analysis as in Ellis (2005), was used to answer the first research question. Following Nunnally (1978), who recommends having 10 times as many participants as factors, and Kass and Tinsley (1979), who recommend having between 5 and 10 participants per factor, it was considered that the minimum acceptable sample size for this type of analysis to be meaningful would be a total n size of 20. By this metric, the final sample size of 30 participants should be sufficient to find meaningful loadings on the two factors.

Table 1 shows the reliability coefficient for each of the five tests. Cronbach’s alpha coefficients were calculated for all except the oral narration test, for which it was necessary to calculate inter-rater reliability. All coefficients were well above .80, indicating that all of the tests were internally consistent.
Table 2 shows the means and standard deviations of scores on the five measures by native speakers and L2 and HL learners. Native speakers, as predicted, scored near ceiling (92% or higher) with very little variance on all measures except the metalinguistic knowledge test. Their scores on all except the metalinguistic knowledge test were higher than those of both the L2 and HL learners. HL learners outperformed L2 learners on all measures, scoring in the 75-80% range except on the metalinguistic knowledge test, where they scored less than 50% and L2 learners scored at nearly the same levels as NSs.

The mean scores of the five tests were submitted to separate one-way ANOVAs, with group (NS, L2, HL) as the between-subjects factor. There were statistically significant between-group differences on all five tests. Table 3 includes the $F$ statistic, $p$ values, and results of the Scheffé post-hoc analyses that were performed to determine the sources of the between-group differences.

Table 4 shows the correlation matrix for L2 and HL learners’ scores ($n = 20$) on the battery of tests. All correlations were significant at the .01 level, and the tests designed to measure implicit knowledge were highly inter-correlated. The tests designed to tap into explicit knowledge showed a different pattern of results; the metalinguistic knowledge test correlated negatively with the other four measures, and the untimed GJT was far less highly correlated with the other measures.

A confirmatory factor analysis was then run using AMOS version 17.0, with the two factor model reported in Ellis (2005) and Ellis and Loewen (2007) specified a priori. The results of the analysis are presented in Figure 1 below, and the summary statistics for the model fit are
presented in Table 5. As in previous studies, the $\chi^2$ statistic was used to test the goodness of fit. A significant $p$ value indicates that the model is statistically unlikely to occur, whereas a non-significant value, such as the one obtained, indicates an acceptable fit (Byrne, 2001). Goodness of fit was also analyzed according to the metrics used in Ellis and Loewen (2007), where a normed fit index value of greater than .95 indicates a superior fit and that a root mean square error of approximation value of less than .05 indicates a good fit for the model.

Discussion

This study was primarily concerned with providing evidence for the construct validity of the battery of five tests designed in R. Ellis (2005) to provide relatively separate measures of implicit and explicit knowledge. The first research question addressed the issue of validity broadly, asking whether scores on the five tests would load on two separate factors, one representing implicit knowledge and the other representing explicit knowledge. If the tests in Ellis (2005) are replicable and tap different sources of knowledge, then the tests adapted in this study for Spanish and tested on both classroom L2 learners and naturalistic HL learners, should load on separate factors.

As shown in Table 5 and Figure 1, confirmatory factor analysis verified that the scores from the battery of tests fit the two-factor explicit/implicit knowledge model established by R. Ellis (2005). These findings provide evidence that these tests, like those developed by R. Ellis in English, indeed tap relatively separate pools of knowledge.

The second research question addressed the construct validity of R. Ellis (2005)’s test battery in a more fine-grained way. Specifically, if the tests really measure relatively separate pools of linguistic knowledge, then L2 learners who have learned Spanish in a classroom setting should score higher on tests of explicit knowledge and lower on tests of implicit knowledge.
Conversely, HL learners, who learned Spanish naturalistically but have received less than two years of formal instruction in it, should score lower on tests of explicit knowledge and higher on tests of implicit knowledge.

The test scores from this study provide such evidence. L2 learners scored highest overall on the two tests that were designed to measure explicit knowledge – the metalinguistic knowledge test (72.4%), followed by the untimed GJT (66.9%). Their scores on the remaining three tests – those that were designed to measure implicit knowledge – were all under 50% accuracy. The HL learners’ scores, however, show the opposite pattern. Whereas L2 learners scored highest on the metalinguistic knowledge test, HL learners as a group scored the lowest on this test (with an average accuracy of just 57.4%). Similarly, whereas L2 learners scored lowest on the timed GJT, oral narration, and oral imitation tests, HL learners scored high on all three, with the oral narration test in fact being the test on which they scored highest as a group (95.9%). These findings point to the content validity of the battery of tests, since the measures appear to distinguish between the L2 learners, who should have higher explicit knowledge, and the HL learners, who should have minimal explicit knowledge. In a recent study in which she examined the relationship between background/contextual variables and English language knowledge, Philp (2009) found some similar results. The participants who scored highest on her implicit language tests clustered together, and subsequent analysis of their language background questionnaires revealed that they had all begun to acquire English at or before the age of 6 and were regular users of English in their homes and/or community. Although she does not use the nomenclature “heritage learner” in her paper, from her descriptions it appears that the participants in this cluster were a mixture of L1 English speakers and heritage learners of English.
A comparison of the participants’ scores on the untimed vs. timed GJT provides additional evidence in support of the validity of the explicit/implicit model proposed by R. Ellis. Specifically, native speakers, L2 learners, and HL learners all scored significantly lower on the timed GJT than they did on the untimed GJT: NS, \( t(9) = 2.98, p = 0.02, d = 1.19; \) L2, \( t(9) = 4.41, p < 0.01, d = 2.64; \) HL, \( t(9) = 5.02, p < 0.001, d = 1.44 \). This indicates that time pressure significantly affected accuracy for all three groups, and to a nontrivial extent, as the large effect sizes indicate. The largest effect size \( (d = 2.64) \), which was found for the L2 learners, is nearly twice as large as that of the HL learners \( (d = 1.44) \) and more than twice as large as that of the NSs \( (d = 1.19) \), thus demonstrating that time was a much bigger factor for L2 learners than for either HL learners or native speakers of Spanish, most likely because the L2 learners relied more heavily on their explicit knowledge to complete the untimed GJT than did participants in either of the other groups.

This study therefore provides support for the construct validity of the battery of tests in R. Ellis (2005), in a different language and with another population of learners. It brings SLA one step closer to having a series of reliable and valid measures available to administer to learners to measure their implicit and explicit linguistic knowledge. Such measures should allow SLA researchers, in R. Ellis’ words to “investigate issues of central theoretical importance in the study of L2 acquisition” (R. Ellis, 2005, p. 169). Certainly, research on the controversial interface/non-interface positions comes to mind, but such measures could be used in part to address other long-standing issues, such as the role of age effects in language acquisition, or the relationship between type of instruction and the nature of resulting linguistic knowledge. In the field of heritage language acquisition, which has been able to benefit from fewer instructional intervention studies than SLA, it would appear crucial to determine how heritage learners’
linguistic knowledge differs from that of monolingually-raised native speakers, or from that of L2 learners who were first exposed to the language in early adulthood. Then, armed with that knowledge, researchers and teachers will be better equipped to design instruction (and empirical studies on the effects of different types of instruction) to determine in what ways various components, such as explicit information, negative evidence, and structured input affect these bilinguals’ heritage language systems. A few studies have begun to address these issues (Montrul & Bowles, 2008, 2009; Potowski et al., 2009), but clear measures of implicit and explicit knowledge will help to further hone and refine measurement and instruction in this growing field of investigation.
Notes

1. An anonymous SSLA reviewer questioned the rationale for scoring unanswered items on the timed GJT as incorrect. S/he rightly pointed out that just because a learner is unable to respond to an item in the given allotment of time, that does not translate into a complete lack of knowledge of the sentence’s syntax on his/her part. This point is well taken; nevertheless, the scoring system used in this study was identical to that used in R. Ellis’ (2005) study for purposes of comparability.

2. An anonymous SSLA reviewer questioned the construct being measured by the metalinguistic knowledge test. Specifically, s/he commented that the test items appeared not to be measuring explicit knowledge “of the L2, that is, underlying L2 rules, but more knowledge of grammatical terminology per se.” It is true that both sections of the metalinguistic knowledge test rely on grammatical terminology. The first section requires learners to select the rule that best explains why each of 17 sentences is ungrammatical, and necessarily makes reference to such concepts as “verb” and “tense”. The second section of the test requires learners to find examples of 21 specific grammatical features in a short text. In creating this test, I have followed R. Ellis’ description of his metalinguistic knowledge test as closely as possible, again for the purposes of comparability, but since his paper did not include detailed test specifications or the complete test itself, it is possible that the metalinguistic knowledge test used in this study focuses more heavily on grammatical terminology than his did, or than Alderson, Clapham, and Steel’s (1997) test did.

Those qualifications not withstanding, Elder (2009) reports on a validity study that was conducted on the metalinguistic knowledge test used in Ellis (2005). Seven of the nine
validation hypotheses were confirmed, lending tentative support for the test as a valid measure of the desired construct.

3. The instructions for the untimed GJT were slightly different in this study than in R. Ellis’ original research. R. Ellis instructed his learners to judge not only the acceptability of the untimed GJT sentences but also to indicate, for each sentence, how they came to the acceptability judgment (by using a rule or by feel) and also to indicate how certain they were about each judgment. An anonymous SSLA reviewer suggests that this difference may have pushed the learners in Ellis’ study to access their explicit knowledge to a greater extent than they did in the present study, and I must agree that this is a possibility.

4. An anonymous SSLA reviewer suggested that the test scores from just the NSs and L2 learners be run, since Ellis’ (2005) study included just those two types of learners. A confirmatory factor analysis using the same two-factor a priori model was run using only the test scores from the NSs and L2 learners ($n = 20$). Similar to what was found when the HL learners were included, the results indicated that there was a good fit between the data and the model, $\chi^2 = 2.9$, $df = 4$, $p = \text{n.s.}$, NFI = .956, RMSEA = .000.
References


language learning, testing, and teaching (pp. 113-138). Bristol, UK: Multilingual Matters.


APPENDIX A

Sentences in the Oral Imitation Test

Grammatical

Los estudiantes generalmente se levantan muy temprano.
Los padres quieren que sus hijos saquen malas notas.
Los estudiantes bailarán mucho cuando tengan tiempo.
Barack Obama quería que muchos jóvenes votaran en las últimas elecciones.
Los políticos conocen a muchas personas.
La universidad de Illinois es una escuela pequeña.
Los profesores miran mucha televisión.
Penélope Cruz es una actriz española.
Stephen Hawking es un físico importante.
Muchos estudiantes están en clase a las 10:00 de la mañana.
La ceremonia de graduación es en mayo.
A los estudiantes les gusta escribir composiciones.
Si no hubiera nevado, muchos estudiantes no se habrían caído.
Muchos estudiantes tomaron clases de español el año pasado.
Las computadoras portátiles son muy útiles.
El día de San Valentín, muchos novios les regalan rosas a sus novias.
La casa de Bill Gates es más grande que la tuya.

Ungrammatical

*Los estudiantes acuestan tarde.
*Los maestros quieren que sus estudiantes estudien poco.
*Los estudiantes beben mucho alcohol cuando irán de vacaciones.
*Sarah Palin quería que muchos demócratas voten en las últimas elecciones.
*Los estudiantes visitan sus familias los fines de semana.
*Los diccionarios son libros pequeñas.
*Los profesores gana mucho dinero.
*Héctor Chávez es el venezolano presidente.
*Stephen King está un escritor famoso.
*Pocos estudiantes son en clase a las 11:00 de la noche.
*Los exámenes finales están en febrero.
*Los estudiantes no les gusta tomar exámenes.
*Si tantos jóvenes no hubieran votado, Barack Obama no habría ganado las últimas elecciones.
*El año pasado recibía una mala nota en la clase de español.
*Clases de matemáticas son aburridas.
*Muchas personas dan caridades dinero.
*El sueldo del presidente es mayor que la tu.
APPENDIX B

Sentences in Timed and Untimed Grammaticality Judgment Tests

**Reflexive verbs**

**Grammatical**
- Carlos siempre se levanta a las 8:00.
- Ana nunca se levanta antes de las 10:00.
- Me ducho por la mañana.
- Me baño por la noche.

**Ungrammatical**
- *Miguel siempre levanta a las 6:00.
- *Elena nunca levanta antes de las 9:00.
- *Lavo los dientes tres veces al día.
- *Lavo la cara por la mañana.

**Present subjunctive with nominative clauses**

**Grammatical**
- Mamá quiere que Paco estudie mucho.
- Papá pide que los niños hagan la tarea.
- La maestra recomienda que los alumnos estudien mucho.
- Mamá quiere que Susana limpie la casa.

**Ungrammatical**
- *Papá quiere que Elena estudia.
- *Mamá pide que hagamos la tarea.
- *La maestra no quiere que miramos la televisión.
- *Ana quiere que salesmos con ella.

**Present subjunctive with temporal clauses**

**Grammatical**
- Cuando el chico llegue a casa, te llamará.
- Cuando José tenga dinero, comprará una casa.
- Cuando tú y yo podamos, iremos a la playa.
- Cuando mis padres vayan de vacaciones, irán a París.

**Ungrammatical**
- *Cuando la chica llega a la escuela, te llamará.
- *Cuando Juan tiene tiempo, estudiará francés.
- *Cuando Paco puede, irá a la piscina.
- *Cuando mis abuelos van a Nueva York, visitarán muchos museos.

**Imperfect subjunctive**

**Grammatical**
- Ana quería que su hijo mirara menos televisión.
Pablo buscaba un carro que costara poco. 
Mamá quería que Manuel limpiara su cuarto. 
Ángel necesitaba una computadora que funcionara bien.

*Ungrammatical* 
*Luis quería que su papá trabaje menos.* 
*Mi padre buscaba un trabajo que pague bien.* 
*Papá quería que Nuria lave la ropa.* 
*Juan necesitaba una secretaria que era bilingüe.*

**DOM** 
*Grammatical* 
Marcos ve al Dr. Sánchez mañana. 
María visita a su abuela cada semana. 
La compañía paga a los trabajadores cada viernes. 
Juan busca a su hija.

*Ungrammatical* 
*Pedro ve el profesor el lunes.* 
*Teresa visita su tío en diciembre.* 
*Carlos paga los empleados cada semana.* 
*Pepe busca su hermano.*

**Noun/Adj gender agreement** 
*Grammatical* 
Manuel vive en una casa roja. 
Mi mejor amiga tiene pelo negro. 
Los estudiantes viven en un apartamento pequeño. 
El chico lleva una chaqueta blanca.

*Ungrammatical* 
*Ana tiene una falda negro.* 
*La maestra tiene un bolígrafo roja.* 
*Los niños quieren un regalo cara.* 
*La chica tiene una pulsera negro.*

**Subject-verb agreement (regular present-tense verbs)** 
*Grammatical* 
Paula mira la televisión. 
Sus padres hablan chino. 
Mamá y yo tomamos mucho café. 
Los niños sacan buenas notas.

*Ungrammatical* 
*Luisa miras una película.* 
*Las chicas hablo inglés.*
*José y yo toman clases por la noche.
*Los hombres saco dinero del banco.

**Adjective placement (postnominal)**

**Grammatical**
Tengo un carro rojo.
Buscamos ese libro blanco.
María es una chica alta.
Pedro es un hombre guapo.

**Ungrammatical**
*Quiero una blanca camisa.
*Necesitamos esa negra falda.
*José es un delgado chico.
*Mario es un malo estudiante.

**Ser/estar (with professions)**

**Grammatical**
Mi padre es médico.
Sus hijos son abogados.
Voy a ser profesor.
Mi tía es enfermera.

**Ungrammatical**
*Mi tío está ingeniero.
*Sus hermanos están actores.
*Ella está maestra.
*Él está dentista.

**Ser/estar (with locations of objects)**

**Grammatical**
Los libros están en la biblioteca.
Mis llaves están en casa.
Los estudiantes están en clase.
Mi padre está en el hospital.

**Ungrammatical**
*Las revistas son en la biblioteca.
*Mi cartera es en el apartamento.
*El estudiante no es en clase.
*Mi mamá es en la iglesia.

**Ser/estar (with events)**

**Grammatical**
La fiesta es el 28 de febrero.
La clase es a las 8:00.
El concierto es en Chicago.
La reunión es mañana.

**Ungrammatical**
*El baile está a las 9:00.
*El examen está el 14 de marzo.
*La fiesta está en Nueva York.
*Mi cumpleaños está el 2 de enero.

**Gustar**

**Grammatical**
A los estudiantes no les gusta el examen.
A los niños les gusta el chocolate.
A mi hermano le gusta la pizza.
A ella no le gustan las clases de matemáticas.

**Ungrammatical**
*Los chicos no les gusta la prueba.
*Los niños les gusta los caramelos.
*Mi hermana le gusta la televisión.
*Ella no le gustan las cucarachas.

**Past Hypothetical Conditionals**

**Grammatical**
Si Roberto hubiera estudiado, no habría sacado malas notas.
Si mi amigo no hubiera salido, no habría gastado mucho dinero.
Si Susana no hubiera estado enferma, habría ido a la fiesta.
Si mis padres hubieran tenido dinero, habrían comprado una casa.

**Ungrammatical**
*Si tu hermano había sido médico, habría ganado mucho dinero.
*Si mi familia hubiera tenido dinero, habríamos ido de vacaciones.
*Si Manuel hubiera tenido cuidado, no habría tenido un accidente.
*Si habías prestado atención, no habríais suspendido el examen.

**Possessives**

**Grammatical**
Como mi apartamento es pequeño, prefiero el tuyo.
Mi perro es grande pero el tuyo es pequeño.
Mi teléfono es viejo pero el suyo es nuevo.
Mi chaqueta es roja y la suya es negra.

**Ungrammatical**
*Mi casa es grande pero el tuyo es pequeña.
*Mi clase favorita es español pero la tuya es matemáticas.
*Mi mochila es nueva y la su es vieja.
*Mi carro es barato pero la su es caro.

**Aspect**

**Grammatical**
Ayer los niños jugaron hasta muy tarde.
Mi padre estuvo en el hospital tres meses.
Cuando era niño, iba al cine todas las semanas.
Teresa vivió muchos años en México.

**Ungrammatical**
*Anoche los chicos estudiaban hasta las 3:00.
*Mi hermano estaba en Venezuela seis meses.
*Cuando era joven, fui de compras todos los fines de semana.
*Mi amiga vivía diez años en Perú.

**Article use with generics**

**Grammatical**
Los delfines son animales inteligentes.
La carne tiene proteína.
Los jóvenes beben mucho alcohol.
Las computadoras son caras.

**Ungrammatical**
*Perros ladran mucho.
*Niños lloran.
*Elefantes son grandes.
*Pájaros vuelan.

**Double Object Construction**

**Grammatical**
Le doy libros a la biblioteca.
Le doy regalos a una amiga.
Le escribo cartas a mi hermano.
Le enseño una foto a Paco.

**Ungrammatical**
*Le doy Paula libros.
*Le regalo mi mamá flores.
*Le doy mi mamá rosas.
*Le doy mi novia un regalo.
APPENDIX C

Metalinguistic Knowledge Test

Part 1: Each sentence below is ungrammatical in Spanish. Please select the option that explains why each sentence is ungrammatical.

1. Los estudiantes duchan en el gimnasio.
   A. Missing reflexive pronoun
   B. Missing indirect object pronoun
   C. Missing subject pronoun
   D. Missing demonstrative pronoun

2. Paula sugiere que estudias para el examen.
   A. A verb is conjugated in the wrong person
   B. A verb is conjugated in the wrong tense
   C. A verb is conjugated in the wrong aspect
   D. A verb is conjugated in the wrong mood

3. Cuando ellos leen, tomarán muchos apuntes.
   A. A verb is conjugated in the wrong person
   B. A verb is conjugated in the wrong tense
   C. A verb is conjugated in the wrong aspect
   D. A verb is conjugated in the wrong mood

4. Todos esperaban que ella se rompa con ese chico.
   A. A verb is conjugated in the wrong person
   B. A verb is conjugated in the wrong tense
   C. A verb is conjugated in the wrong aspect
   D. A verb is conjugated in the wrong mood

5. El médico señala José.
   A. Missing dative marker
   B. Missing accusative marker
   C. Missing object pronoun
   D. Missing subject pronoun

6. La mujer necesita una bufanda barato.
   A. There is incorrect subject-verb agreement.
   B. There is incorrect gender assignment.
   C. There is incorrect gender agreement.
   D. There is incorrect number agreement.
7. Mi hijo bailamos poco.
   A. A verb is conjugated in the wrong person
   B. A verb is conjugated in the wrong tense
   C. A verb is conjugated in the wrong aspect
   D. A verb is conjugated in the wrong mood

8. Carla es una simpática chica.
   A. There is incorrect gender agreement.
   B. There is incorrect adjective placement.
   C. There is incorrect number agreement.
   D. The verb *ser* is used incorrectly.

9. Ramiro está profesor.
   A. There is incorrect gender agreement.
   B. There is incorrect adjective placement.
   C. The wrong verb is used to indicate a profession.
   D. The wrong verb is used to indicate a temporary job.

10. Roberto es en el supermercado.
    A. There is incorrect gender agreement.
    B. There is incorrect adjective placement.
    C. The wrong verb is used to indicate the location of an object.
    D. The wrong verb is used to indicate the location of a person.

11. La boda está a las 4:00.
    A. There is incorrect gender agreement.
    B. There is incorrect adjective placement.
    C. The wrong verb is used to indicate the location of an object.
    D. The wrong verb is used to indicate the location of an event.

12. Sergio no le gusta tocar el violín.
    A. Missing dative marker
    B. Missing accusative marker
    C. Missing or incorrect indirect object pronoun
    D. Missing or incorrect direct object pronoun

13. Si el chico había estado en clase, habría aprendido mucho.
    A. A verb is conjugated in the wrong person
    B. A verb is conjugated in the wrong tense
C. A verb is conjugated in the wrong aspect
D. A verb is conjugated in the wrong mood

14. Tu profesor es interesante pero el mi es aburrido.
   A. A possessive pronoun is used incorrectly.
   B. A demonstrative adjective is used incorrectly.
   C. A demonstrative pronoun is used incorrectly.
   D. A subject pronoun is used incorrectly.

15. Comida orgánica cuesta mucho dinero.
   A. There is a problem with gender agreement.
   B. There is a problem with gender assignment.
   C. There is a problem with definite article use.
   D. There is a problem with indefinite article use.

   A. Missing dative marker
   B. Missing accusative marker
   C. Missing indirect object pronoun
   D. Missing definite article

17. Susana vivía dieciocho años en Rusia.
   A. A verb is conjugated in the wrong person
   B. A verb is conjugated in the wrong tense
   C. A verb is conjugated in the wrong aspect
   D. A verb is conjugated in the wrong mood
Todo empezó el miércoles 12 de mayo. Tres días antes de San Isidro, la fiesta mayor de Madrid. Una fiesta que dura una semana y media, más o menos, con baile y espectáculos todas las noches. Miles de madrileños están por la calle hasta muy tarde, y hay gente y ruido por todas partes pero especialmente en el centro. Y yo vivo en el centro. Además, a mí, las fiestas populares no me gustan. Por eso, ese año había decidido irme unos días de vacaciones. Ese miércoles 12 de mayo, estaba a punto de irme. Pensaba pasar toda la semana en Menorca. En mayo es una buena época: pocos turistas y, seguramente, bastante buen tiempo. Quería tomar el sol y no hacer nada en absoluto. Sólo descansar. Descansar y leer un par de buenas novelas. Pero no pudo ser. En la agencia de detectives no tenemos mucho trabajo normalmente. Pero, siempre que quiero irme de vacaciones, las cosas se complican. Ese miércoles 12 de mayo, un día antes de irme a Menorca, sonó el teléfono.

1. Finite verb
2. Infinite verb
3. Demonstrative adjective
4. Preposition
5. Adverb
6. Definite article
7. Indefinite article
8. Verb in imperfect
9. Verb in preterit
10. Proper noun
11. Subject pronoun
12. Auxiliary verb
13. Indirect object pronoun
14. Verb in indicative mood
15. Verb in third person
16. Verb in first person
17. Conjunction
18. Quantifier
19. Indefinite
20. Feminine noun
21. Masculine noun
TABLE 1

Reliability coefficients for the five tests

<table>
<thead>
<tr>
<th>Test</th>
<th>Items</th>
<th>Participants</th>
<th>Reliability coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral imitation</td>
<td>34</td>
<td>30</td>
<td>$\alpha$=.96</td>
</tr>
<tr>
<td>Oral narration</td>
<td>Variable</td>
<td>30</td>
<td>$r$=.92</td>
</tr>
<tr>
<td>Timed GJT</td>
<td>68</td>
<td>30</td>
<td>$\alpha$=.96</td>
</tr>
<tr>
<td>Untimed GJT</td>
<td>68</td>
<td>30</td>
<td>$\alpha$=.94</td>
</tr>
<tr>
<td>Metalinguistic</td>
<td>38</td>
<td>30</td>
<td>$\alpha$=.93</td>
</tr>
</tbody>
</table>
TABLE 2

Descriptive statistics for the five tests by group

<table>
<thead>
<tr>
<th>Test</th>
<th>NS</th>
<th>L2</th>
<th>HL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean%</td>
<td>SD</td>
<td>n</td>
</tr>
<tr>
<td>Oral imitation</td>
<td>99.7</td>
<td>.9</td>
<td>10</td>
</tr>
<tr>
<td>Oral narration</td>
<td>100</td>
<td>--</td>
<td>10</td>
</tr>
<tr>
<td>Timed GJT</td>
<td>92.1</td>
<td>3.9</td>
<td>10</td>
</tr>
<tr>
<td>Untimed GJT</td>
<td>96.2</td>
<td>2.8</td>
<td>10</td>
</tr>
<tr>
<td>Metalinguistic knowledge</td>
<td>77.1</td>
<td>15.8</td>
<td>10</td>
</tr>
</tbody>
</table>
TABLE 3

Results of one-way ANOVAs on test scores for the three groups

<table>
<thead>
<tr>
<th>Test</th>
<th>F</th>
<th>Results of Scheffé post-hoc analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timed GJT</td>
<td>56.39**</td>
<td>NS&gt;Hl</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HL&gt;L2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NS&gt;L2</td>
</tr>
<tr>
<td>Untimed GJT</td>
<td>22.68**</td>
<td>NS&gt;Hl</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NS&gt;L2</td>
</tr>
<tr>
<td>Metalinguistic</td>
<td>3.69*</td>
<td>NS&gt;Hl</td>
</tr>
<tr>
<td>knowledge test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral imitation test</td>
<td>65.78**</td>
<td>NS&gt;Hl</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HL&gt;L2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NS&gt;L2</td>
</tr>
<tr>
<td>Oral narration test</td>
<td>47.10**</td>
<td>NS&gt;L2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HL&gt;L2</td>
</tr>
</tbody>
</table>

** p < 0.0001
* p < 0.05
TABLE 4

Correlation matrix for the five tests (L2 and HL learners, n=20)

<table>
<thead>
<tr>
<th>Test</th>
<th>Oral imitation</th>
<th>Oral narration</th>
<th>Timed GJT</th>
<th>Untimed GJT</th>
<th>Metalinguistic knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral imitation</td>
<td>--</td>
<td>.78</td>
<td>.64</td>
<td>.51</td>
<td>-.17</td>
</tr>
<tr>
<td>Oral narration</td>
<td>--</td>
<td>--</td>
<td>.60</td>
<td>.32</td>
<td>-.28</td>
</tr>
<tr>
<td>Timed GJT</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>.21</td>
<td>-.32</td>
</tr>
<tr>
<td>Untimed GJT</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>-.14</td>
</tr>
<tr>
<td>Metalinguistic knowledge</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>
TABLE 5

Summary statistics for the model of fit

<table>
<thead>
<tr>
<th>$\chi^2$</th>
<th>NFI</th>
<th>RMSEA</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.9 (p = n.s.)</td>
<td>.974</td>
<td>.000</td>
<td>4</td>
</tr>
</tbody>
</table>