CrossMark

Studies in Second Language Acquisition, 2015, **37**, 457–486.
doi:10.1017/S0272263114000370

MEASURING UNIVERSITY-LEVEL L2 LEARNERS' IMPLICIT AND EXPLICIT LINGUISTIC KNOWLEDGE

Runhan Zhang
Central University of Finance and Economics

Although many theoretical issues revolving around implicit and explicit knowledge in second language (L2) acquisition hinge on the ability to measure these two types of knowledge, few empirical studies have attempted to do so. However, R. Ellis (2005) did develop a battery of tests intended to provide relatively separate measures. This study aims to validate the use of Ellis's test battery in an English as a foreign language context and to investigate the extent of Chinese first-year university students' implicit and explicit L2 knowledge. Test scores loaded on two factors, as in R. Ellis (2005), thus demonstrating construct validity for the tests as measures of implicit and explicit knowledge in a population of Chinese university-level learners of English in a Chinese (as opposed to English as a second language) context. These learners were found to have higher scores on measures of explicit knowledge than on those of implicit knowledge because of the instruction they had received and their English learning environment.

Implicit linguistic knowledge is knowledge of language that is intuitive and procedural. It often underlies communicative competence. *Explicit linguistic knowledge*, on the other hand, is knowledge about language. Implicit and explicit second language (L2) knowledge are two central constructs in the field of SLA. The distinction and relationship between them are of particular interest and importance for SLA research and

Correspondence concerning this article should be addressed to Runhan Zhang, School of Foreign Studies, Central University of Finance and Economics, Haidian District, Beijing, P.R. China. E-mail: zhangrunhan_nora@hotmail.com

© Cambridge University Press 2014

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

language pedagogy (N. Ellis, 1994, 2008) because the ultimate, most highly prized goal of language learning is spontaneous, unreflecting language use, which relies primarily on implicit L2 knowledge (Sharwood Smith, 1981), although L2 performance usually involves a combination of implicit and explicit knowledge (R. Ellis, 2009a). The lack of validated instruments for the measurement of these two types of knowledge has seemed to prevent empirical examination of them and of their relationship. One difficulty in measuring the two types of knowledge is, as proposed by Bowles (2011) and mentioned previously, that learners' "interlanguage contains both types of knowledge" (p. 248), making it tricky in practice to operationalize them separately. The current study explores the validity of the tests designed by R. Ellis (2005) to provide separate measures of implicit and explicit knowledge with a population of university-level learners of English as a foreign language (EFL) at the upper-intermediate stage of proficiency who have learned English explicitly in an EFL context and have had a large amount of formal instruction (i.e., who had begun to learn English before entering postsecondary education). All their teachers were using the grammar-translation method.[1]

In addition to the benefit of validating the measures of implicit and explicit knowledge, measuring the extent to which first-year Chinese university students possess each of these two types of knowledge has implications for both SLA theory and language pedagogy. As mentioned previously, spontaneous, unreflecting language use is the final goal of language learning. However, in reality, it is assumed that most L2 learners, especially those in an EFL context like China's, will have much more explicit L2 knowledge than implicit L2 knowledge. If this is found to be so in this study, it will show the need for attention to be directed to the creation and implementation of teaching methods that can equip L2 learners with more implicit knowledge.

CONCEPTUAL ISSUES

Implicit and Explicit Linguistic Knowledge

In neurolinguistic, psycholinguistic, and SLA literature, there have been a number of terms used to refer to implicit and explicit knowledge: *unanalyzed* and *analyzed knowledge* (Bialystok, 1978), *unconscious* and *conscious knowledge* (Schmidt, 1990), *acquired* and *learned knowledge* (Krashen, 1981), *procedural* and *declarative knowledge* (DeKeyser, 1998; Paradis, 1994), *tacit* and *explicit knowledge* (Reber, 1989), and *implicit linguistic competence* and *explicit metalinguistic knowledge* (Roehr, 2008). In this article, the terms *implicit* and *explicit* knowledge are used to avoid confusion.

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

Implicit linguistic knowledge refers to knowledge of language; it is "an intuitive feeling for what is correct and acceptable" (Bialystok, 1981, cited in Sharwood Smith, 1981, p. 159). Implicit knowledge is also the knowledge underlying communicative ability and is used in spontaneous comprehension and production. R. Ellis (1994) distinguishes two kinds of implicit linguistic knowledge: *formulaic* and *rule-based knowledge*. As the names imply, formulaic knowledge is knowledge of internalized expressions like "how are you" and "nice to meet you," which are also described as unanalyzed chunks, whereas rule-based knowledge is knowledge of internalized rules, for example, the rules governing the use of the third person and the plural marker *-s*.

In contrast to implicit knowledge, explicit knowledge refers to knowledge about language, which can be broken down into *analyzed knowledge* and *metalanguage* (R. Ellis, 1997, p. 110). Analyzed knowledge refers to knowledge regarding L2 items and structures of which the learner has metalinguistic awareness (R. Ellis, 2008b). It is assumed in the current study that analyzed knowledge and the "metalinguistic knowledge" of Roehr (2008) refer to the same phenomenon; metalanguage is language used to describe language (Richards, Platt, & Weber, 1985), which can entail technical (e.g., "the most typical semantic role of a subject is agentive," Quirk & Greenbaum, cited in R. Ellis, 2008a, p. 144) or semitechnical terminology (e.g., "the subject typically tells us who does an action," R. Ellis, 2008a, p. 145). According to R. Ellis (2008b), metalanguage "must be learnt through instruction or observation" (p. 144). R. Ellis (2004) provides the following general definition of explicit knowledge:

> Explicit L2 knowledge is the declarative and often anomalous knowledge of the phonological, lexical, grammatical, pragmatic, and sociocritical features of an L2 together with the metalanguage for labeling this knowledge. It is held consciously and is learnable and verbalizable. It is typically accessed through controlled processing when L2 learners experience some kind of linguistic difficulty in the use of the L2. Learners vary in the breadth and depth of their L2 explicit knowledge. (pp. 244–245)

In a word, implicit knowledge is intuitive and beyond people's awareness, whereas explicit knowledge is declarative and is accessible to consciousness; implicit knowledge is at the core of automated language processing and needs online planning, whereas explicit knowledge needs controlled effort and thus is typically used in tasks that allow for careful planning and monitoring; explicit knowledge can be reported, whereas implicit knowledge is nonverbalizable. Hulstijn (2002) argues that, in fluent speakers, L2 knowledge is mostly implicit. Accordingly, there seems to be consensus that the development of implicit knowledge is the ultimate goal of L2 acquisition (Doughty, 2003).

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

The Relationship between Implicit and Explicit L2 Knowledge: Interface Positions

Central to the debate regarding implicit and explicit L2 knowledge is the question of whether there is an interface between the two—that is, whether there is the possibility of one knowledge type becoming the other. Three positions—namely, the noninterface, the strong interface, and the weak interface positions—provide different answers to this question.

The Noninterface Position. The noninterface position claims that there is an absolute distinction between implicit and explicit knowledge and thus that there is no possibility of one knowledge type converting directly into the other. Many researchers hold this position (e.g., Hulstijn, 2002, 2005; Krashen, 1981, 1982, 1985; Paradis, 2009; Schwartz, 1993). Krashen (1981) espoused this position by making a distinction between acquisition and learning. Explicit knowledge, according to Krashen, is only useful and available inasmuch as it allows L2 learners to construct a monitor of performance that can check output to ensure that it is correct (Krashen, 1981). Zobl (1995) held the same view and further stated that explicit knowledge plays no role in acquisition (i.e., the development of implicit knowledge), which "comes about simply through carrying out more successful computations on intake data . . . by deducing unknowns from available representations in the course of processing an input string" (Zobl, 1995, p. 5). Thus, it is believed that implicit knowledge is not influenced by explicit knowledge.

The Strong Interface Position. In contrast to the noninterface position, the strong interface position states that, although there is a distinction between implicit and explicit knowledge, there can also occur conversion of learned knowledge (explicit) to acquired knowledge (implicit). More precisely, those holding this position claim that L2 knowledge is first gained in explicit form and then transformed into implicit form through communicative practice (DeKeyser, 1998); they also acknowledge "the possibility of the transfer of implicit knowledge to explicit knowledge through the process of conscious reflection on and analysis of output generated by means of implicit knowledge" (Milasi & Pishghadam, 2007, p. 19). In Bialystok's (1978) model of L2 learning, knowledge is divided into three kinds: explicit knowledge, implicit knowledge, and other knowledge. The model shows that linguistic information first leads to explicit knowledge and then becomes implicit knowledge only after continuous practice and use. In the meantime, when a L2 learner needs to report an explicit rule, implicit knowledge can be drawn on to formulate explicit knowledge in an ad hoc way. For example, a learner learns

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

the rules of the word class of determiners in English (explicit knowledge) and then comes to be able to use determiners very well after a certain amount of practice (implicit knowledge). In the meantime, if he wants to elaborate a description of determiners, he may draw on his implicit knowledge to formulate explicit knowledge and say, "Determiners refer to words that are used in the premodification of a noun phrase and that typically precede any adjectives that premodify the headword." DeKeyser (2003) articulated the strong interface position as follows:

> Even though implicitly acquired knowledge tends to remain implicit, and explicitly acquired knowledge tends to remain explicit, explicitly learned knowledge can become implicit in the sense that learners can lose awareness of its structure over time, and learners can become aware of the structure of implicit knowledge when attempting to access it, for example for applying it to a new context or for conveying it verbally to somebody else. (p. 315)

The Weak Interface Position. The third position concerning the interface hypothesis is an in-between position referred to as the weak interface position, supported by N. Ellis (1994, 2005, 2008) and R. Ellis (1993, 1994, 2005, 2009a). The proponents of this position have suggested that it is possible for explicit knowledge to convert into implicit knowledge, but they posit conditions on when and how this can happen. There are three separate perspectives within this position.

The first perspective draws on Pienemann's learnability/teachability hypothesis, which claims that "learners can benefit from classroom instruction only when they are psycholinguistically ready for it" (Kumaravadivelu, 2006, p. 77). That is, the learnability of a structure constrains the effectiveness of teaching. Pienemann (1989) claimed that "instruction can only promote language acquisition if the interlanguage is close to the point when the structure to be taught is acquired in the natural setting" (p. 60) and therefore that "teaching is ineffectual, since L2 acquisition can only be promoted when the learner is ready to acquire the given items in the natural context" (p. 61). The learnability/teachability hypothesis concerns the relationship between implicit and explicit knowledge and acknowledges the possibility of explicit knowledge becoming implicit but only when "the learner is developmentally ready to acquire the linguistic form" (R. Ellis, 2009a, p. 21).

The second perspective claims that the two kinds of knowledge are closely related and that learning can indirectly facilitate acquisition because learned knowledge can help to make the hypothesis-testing process more efficient and can help learners acquire rules more rapidly (R. Ellis, 1997, 2001). N. Ellis (1994) similarly points out that teaching "declarative rules can have 'top-down' influences on perception" in that

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

"it can make features perceptually more salient and thus enable the learners to 'notice the gap' between the observed input and their output based on their existing interlanguage system" (p. 16), a process that is crucial for subsequent learning (Schmidt, 1994). Thus, these learners will acquire the features in question through this process. Researchers adopting this perspective believe that implicit and explicit learning processes work together in L2 acquisition, but they also acknowledge that conversion will occur only when a learner is developmentally ready—and only for rules that are developmentally constrained, meaning that they are acquired in a definite sequence from simple to complex (R. Ellis, 1994).

The third perspective thus expands on the second perspective, emphasizing that learners' output derives from explicit knowledge, which can act as auto-input that triggers learners' implicit learning mechanisms (Schmidt & Frota, 1986; Sharwood Smith, 1981).

A number of empirical studies have examined the relationship between L2 learners' implicit and explicit knowledge (Green & Hecht, 1992; Hu, 2002; Macrory & Stone, 2000). However, despite the theoretical debate outlined previously, none of these studies has investigated whether the two types of knowledge can convert into each other or to what degree the two knowledge systems interact (R. Ellis, 2005). The interface debate calls for clear empirical evidence, but the lack of valid instruments for measuring implicit and explicit knowledge has prevented an empirical test of the interface positions (R. Ellis, 2005).

REVIEW OF STUDIES MEASURING IMPLICIT AND EXPLICIT KNOWLEDGE

As Elder and Ellis (2009) note, there is a dearth of studies that examine the relationship between implicit and explicit knowledge. One such study is Han and Ellis (1998). Han and Ellis's empirical study explored ways of measuring the two types of knowledge. Based on their study, R. Ellis (2005) carried out a psychometric study aiming to establish operational definitions of the two types of knowledge in English and also to provide relatively independent measures of them. The results of his study showed that the tests employed (i.e., the elicited imitation test [EIT], oral narrative test [ONT], timed grammaticality judgment test [TGJT], untimed grammaticality judgment test [UTGJT], and metalinguistic knowledge test [MKT]) provided relatively separate measures of implicit and explicit knowledge. Later, Bowles (2011) conducted a study to validate R. Ellis's proposed measures of the two types of knowledge with a different language (Spanish) and a different learner population. The findings of her study provided further empirical support for the construct validity of the test battery in R. Ellis (2005). In their studies, Gutiérrez (2012) and Erçetin and Alptekin (2013) used three of R. Ellis's tests to measure implicit and

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

explicit knowledge of L2 Spanish and English learners, respectively. Their findings also provided a certain degree of support for the construct validity of these tests as measures of implicit and explicit knowledge. The following section focuses on these studies.

Han and Ellis (1998) explored ways of measuring the implicit and explicit L2 knowledge of 48 adult English learners from various countries studying English in the United States. Only one grammatical structure—verb complements—was investigated. Several tests were used to elicit implicit and explicit knowledge: (a) the TGJT, (b) the oral production test (OPT), (c) the delayed grammaticality judgment test (delayed GJT), and (d) metalingual comments. The first two tests were intended to elicit implicit knowledge, and the second two, explicit knowledge. The TGJT was administered twice (TGJT 1 and TGJT 2). All together, scores for five tests were obtained: (a) the TGJT 1, (b) the TGJT 2, (c) the OPT, (d) the delayed GJT, and (e) metalingual comments.

Principal component analysis (PCA) revealed a distinct two-factor solution. The delayed GJT and metalingual comments loaded on one factor, and the OPT and TGJT 1 loaded on the other factor. The results suggested that the tests yielded relatively separate measures of implicit and explicit knowledge, providing some support for distinguishing implicit and explicit L2 knowledge. However, the TGJT 2 was found to load on both factors, on the basis of which Han and Ellis (1998) suggested that the TGJT 2 tapped both types of knowledge. They speculated that "familiarity with the test content and procedure allowed the subjects to access a greater mix of strategies, drawing on both types of knowledge" (p. 17).

Building on Han and Ellis's (1998) exploratory study, R. Ellis (2005) conducted a study aiming to establish the construct validity and reliability of a battery of tests designed to elicit implicit and explicit knowledge of English. The participants were 20 native speakers (NSs) and 91 learners of English with various proficiency levels. The NSs were predicted to outscore the L2 learners on all tests, especially those intended to elicit implicit knowledge. Following R. Ellis (2004) and previous studies, R. Ellis (2005) proposed seven criteria for operationalizing implicit and explicit knowledge: degree of awareness, time available, focus of attention, systematicity, certainty, metalanguage, and learnability. A battery of tests measured learners' implicit and explicit L2 knowledge of 17 English grammatical structures (e.g., verb complements and *yes*/*no* questions), chosen on the basis of three considerations: universal problematicity among learners, developmental properties (a mix of early- and late-acquired structures), and level of instruction (where the structure is typically introduced—beginner, lower/upper intermediate, or advanced level). The tests used were the (a) EIT, (b) ONT, (c) TGJT, (d) UGJT, and (e) MKT. Table 1 shows the design features of the tests. The EIT, ONT, and TGJT were designed to measure students' implicit knowledge, and the UGJT and MKT to measure their explicit knowledge.

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

Table 1. Design Features of the Tests in the Test Battery in R. Ellis (2005)

Criterion	EIT	ONT	TGJT	UGJT	MKT
Degree of Awareness	Feel	Feel	Feel	Rule	Rule
Time available	Pressured	Pressured	Pressured	Unpressured	Unpressured
Focus of attention	Meaning	Meaning	Form	Form	Form
Utility of knowledge of metalanguage	No	No	No	Yes	Yes

Note. This table is adapted from R. Ellis (2009b, p. 47). EIT = elicited imitation test; ONT = oral narrative test; TGJT = timed grammaticality judgment test; UGJT = untimed grammaticality judgment test; MKT = metalinguistic knowledge test.

It was predicted that the five tests would provide relatively separate measures of implicit and explicit knowledge according to the four criteria previously described. The EIT and ONT were predicted to measure implicit knowledge in that the test takers would rely mainly on "feel" (intuition) and would perform under time pressure to prevent them from accessing their metalanguage. In contrast, the MKT was predicted to measure explicit knowledge because it was unpressured and would require attention to form by the test taker. Both of the GJTs also required attention mainly to form, but the TGJT was predicted to primarily measure implicit knowledge, and the UGJT explicit knowledge. The TGJT was pressured and encouraged in learners the use of "feel"; there was little chance to access metalinguistic knowledge in this test. On the contrary, the UGJT was not pressured and encouraged test takers to use "rules," making it more likely that their responses would involve metalinguistic knowledge. R. Ellis (2005) further argued on theoretical grounds that, in an UGJT, ungrammatical sentences provide an especially clear measure of explicit knowledge because "sentences that learners judged to be ungrammatical or that they are not sure about often invoked attempts to make use of declarative knowledge" (R. Ellis, 1991, p. 178).

The results of R. Ellis (2005) could be summarized as follows, and these results provide evidence for the reliability and validity of the five tests used in his study (see also the summary of the results in R. Ellis, 2009b, pp. 49–58):

1. *Reliability.* Four of the tests had alphas greater than .80, and interrater reliability of the ONT was .85.
2. *Degree of awareness.* As predicted, "rule" correlated better with the UGJT ungrammatical items and the MKT (r = .32, p < .05 and .37, p < .05, respectively), which were most likely to elicit explicit knowledge.
3. *Time available.* As predicted, two factors were extracted, and pressured and unpressured tests loaded on different factors (R. Ellis, 2005, 2009b).[2]
4. *Focus of attention.* It was hypothesized that the two tests designed to direct learners' attention to meaning (the EIT and the ONT) would load heavily on the same factor, but that the TGJT, which required a focus on form, would load less heavily on this factor (R. Ellis, 2009b). The result echoed this hypothesis to some extent. However, R. Ellis (2009b) also admits that "this hypothesis cannot be properly tested in this study as the focus and time-pressure variables were confounded in the design of the tests" (p. 55).
5. *Utility of metalinguistic knowledge.* Scores on the MKT correlated more strongly with scores on the UGJT and the UGJT ungrammatical items (i.e., UGJT UG) than with those on the EIT and ONT, indicating that tests of explicit knowledge encouraged greater use of metalinguistic knowledge than tests of implicit knowledge.

Using the same instruments and similar target structures, Bowles (2011) later validated R. Ellis's (2005) test battery in a different language,

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

Spanish, and with a different learner population. The 30 participants were evenly divided into three groups: native Spanish speakers who had immigrated to the United States as adults, L2 intermediate Spanish learners raised in monolingual English-speaking households, and heritage learners of Spanish raised in bilingual English-Spanish households who had used Spanish throughout childhood. The findings provide further empirical support for the construct validity of the test battery with a different language and different learner population and also support its content validity by distinguishing between "L2 learners, who should have higher explicit knowledge, and the HL [heritage language] learners, who should have minimal explicit knowledge" (Bowles, 2011, p. 262). However, it was found that heritage learners scored 81.2% on the UGJT (one of the explicit knowledge measures), which was much higher than the L2 learners' score on this measure (M = 66.9%). So it seems that Bowles's predictions were not entirely borne out. This may need further empirical examination with a larger population size.

Gutiérrez (2012) also administered tests of implicit and explicit knowledge to learners of L2 Spanish. Unlike Bowles (2011), he employed two measures for explicit knowledge (the UGJT and MKT) and one for implicit knowledge (the TGJT) in his study; all structures used had been taught explicitly to participants in class. Principal component analysis and confirmatory factor analysis (CFA) showed that the grammatical items of the GJTs loaded on one factor and the ungrammatical items of the GJTs and the MKT loaded on another factor; thus, the grammatical/ungrammatical model worked for Gutiérrez's participants. Therefore, the grammatical sentences in the timed and untimed GJTs were considered to be measures of implicit knowledge, and the ungrammatical sentences in both GJTs and in the MKT were considered to be measures of explicit knowledge in this study. It should be admitted that Gutiérrez's study distinguished between the grammatical and ungrammatical items in both GJTs as separate measures of implicit and explicit knowledge, which provided more evidence that the ungrammatical items of the UGJT were a better measure of explicit knowledge than the grammatical ones. However, this model still needs further examination because the EIT and the ONT, which are thought to be good measures of implicit knowledge, were not included in this study.

Likewise, Erçetin and Alptekin (2013) employed three of R. Ellis's tests in their study to measure the implicit and explicit knowledge of L2 English of 51 Turkish university-level students. Differing from Gutiérrez (2012), they used one measure for implicit knowledge (the UGJT) and two measures for explicit knowledge (the TGJT and the EIT). Their study aimed to explore and examine three correlations: (a) the relationships between L2 implicit/explicit knowledge and L2 working memory (WM) capacity, (b) the relationships between L2 reading comprehension and L2 WM capacity, and (c) the relationships between L2 reading

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

comprehension and L2 explicit/implicit knowledge sources. Therefore, besides the three measures of implicit and explicit knowledge, two other measures of L2 WM capacity and reading comprehension were also used. The results of the PCA with varimax rotation showed that two factors were extracted from the five test variables. The TGJT and EIT loaded on the same factor, whereas the UGJT and the L2 reading comprehension test (the comprehension section of the Nelson-Denny Reading Test) loaded on a different factor, with the reading span task loading on both factors. To a certain degree, the different loadings of the three implicit and explicit measures provided further empirical evidence for their relative separateness. However, the focus of Erçetin and Alptekin's study was not to validate the three measures. In other words, the two-factor solution was obtained according to five test variables rather than just the three measures of implicit and explicit knowledge. As a result, the loading pattern of these three measures of L2 implicit and explicit knowledge still remains unclear, but the loading pattern of all these five test variables provides us with a broader view in terms of using more tests to measure implicit and explicit knowledge in the future.

THE CURRENT STUDY

Although previous studies have provided partial support for the tests of implicit and explicit knowledge developed by R. Ellis (2005), it has also been argued that these tests do not provide pure measures of implicit and explicit knowledge. Rather, some tests are considered more likely to measure explicit L2 knowledge and others to measure implicit L2 knowledge. The present study tests four of the measures (the EIT, TGJT, UGJT, and MKT) in a new context, that of majors in English at a Chinese university, focusing on the issues of reliability and construct validity.

Hence, the current study was designed to validate that the battery of tests reported in R. Ellis (2005) indeed provide relatively separate measures of implicit and explicit L2 knowledge and also to investigate to what extent Chinese first-year university students have developed these two types of knowledge with regard to English. Whereas Ellis administered a battery of tests to both NSs and L2 learners of English who had been living in an English-speaking country (New Zealand) for up to 2 years, the current study applied four of the tests to a large pool of university-level L2 learners of English with no experience living in an English-speaking country. The L2 learners in the current study were thus expected to perform worse on the EIT than both the L2 learners and the NSs in R. Ellis's (2005) study but not worse on the MKT, because it relies on grammatical terminology and rules that both NSs and L2 learners most likely do not access frequently.

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

Two specific research questions guided the current research:

1. Can the battery of tests provide relatively separate measures of implicit and explicit knowledge in a population of native Chinese-speaking English L2 learners?
2. What is the extent of Chinese first-year university students' implicit and explicit knowledge of English?

Participants

The study was conducted at a key university in Beijing.[3] In China, admission to a particular university is dependent on student scores on the National Matriculation Examination. The lowest entrance mark for key universities is around 40 points higher than that for nonkey universities. The English department in the host university usually has four classes in each year, with 24 students in each class on average. All four classes of English majors in Year 1 (N = 100) participated in this study. They had begun attending the university around one month prior to the study and came from different parts of China, which meant their exposure to English varied. Students from small towns or the countryside had had no opportunities at all to communicate in English with native English speakers. Some of their English teachers had not even been very good at speaking or listening in English. These students' goals of learning English were to pass the exams. In contrast, students from big cities such as Beijing, Shanghai, and Guangzhou had better English resources to resort to, and some of them had had more opportunities to communicate with native English speakers. Their teachers had been more qualified in terms of their English ability. However, generally speaking, all of the participants had received form-focused instruction in class because of China's exam-oriented educational system, in which English grammar is valued and communicative ability is belittled. The majority of the participants were female (n = 93). The participants were told the tests were to be conducted for the purpose of research and were assured that the information collected would not affect their course grades.

Target Structures

Four of the tests from R. Ellis (2005) were used in the current study to measure the participants' implicit and explicit grammatical knowledge (i.e., the EIT, TGJT, UGJT, and MKT). The target structures were the same 17 grammatical structures used in R. Ellis's (2005) study. This choice of

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

the grammatical structures proposed by R. Ellis (2005) was based on several considerations. First, an attempt was made to select grammatical structures that are universally problematic for L2 learners. The second consideration was the developmental properties of the structure; structures were selected whose properties made them representative of both early and late acquisition of L2 grammar. This was based on the results of morpheme studies suggesting a universal order of acquisition for L2 learners (see R. Ellis, 2008b). Third, the structures selected are typically introduced across various levels of instruction (elementary, lower intermediate, intermediate, and advanced). Morphological structures with fewer alternations (e.g., indefinite articles) are usually introduced prior to those with more alternations (e.g., modal verbs), and the teaching of morphological structures on the whole is more likely to happen before that of syntactic structures (e.g., embedded questions). Although this aspect was considered in light of the order of introduction of information in English textbooks used in New Zealand, it was also true for the Chinese context, as reflected in the syllabi for elementary, junior high, and senior high schools and for university English majors. For example, indefinite articles (which have fewer alternations, only *a* and *an*) are introduced before the regular past tense (which has more alternations, including *+ed*, *+d*, and *–y+ied*) in elementary school. Relative clauses (a more complex syntactic feature) are introduced much later (i.e., in senior high school or university) than the regular past tense (i.e., in elementary or junior high school). The last consideration was grammatical type; that is, the structures included both target morphological and syntactic features. Appendix A lists the grammatical structures that occur in the syllabi for Chinese students with exemplars of the structures. One example is given in (1):

(1) *Structure:* Verb complement
Exemplar: Liao says he wants to buy a new car.
Acquisition: Early
Pedagogical introduction: Junior high school
Type: Syntactic

Description of the Tests

The Elicited Oral Imitation Test (EIT). The EIT was used to elicit implicit L2 knowledge in this study because it has been shown to be the "best" measure of implicit knowledge (R. Ellis, 2009b, p. 59). This is because it is reconstructive in nature and attempts to direct participants' attention to meaning first. This test consists of a set of belief statements involving both grammatical and ungrammatical sentences containing

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

the target structures. There were 68 statements in the original version of this test. This number was subsequently reduced to 34 (one grammatical and one ungrammatical sentence per structure; see Appendix B) to shorten the time it took to administer the test. The sentences were presented aurally to the test takers using a computer. Participants were first required to indicate on paper whether they agreed with, disagreed with, or were not sure about the truth of each statement. This was intended to focus the learners' attention on meaning first. The test takers were then asked to repeat the sentences orally in correct English. Examples of belief statements for third-person *-s* are given in (2) and (3):

(2) *A good teacher makes lessons interesting and cares about students.* (Grammatical)
(3) *Everyone loves comic books and read them.* (Ungrammatical)

The students' responses were audio recorded and then analyzed by identifying obligatory occasions for the use of the target structures. Failure on the part of the test taker to imitate a sentence at all or to reproduce it in a form that did not create an obligatory context for the target structure of the sentence was coded as "avoidance." Each imitated sentence was allocated a score of either 1 (the target structure is correctly supplied) or 0 (the target structure is either avoided or attempted but incorrectly supplied). Scores were expressed in terms of percentage correct. The sentences in this test were based on R. Ellis's (2005) study but were changed and retaped to make the test suitable for Chinese students. An example of the structure *comparatives* is shown in (4):

(4) Original statement: *New Zealand is* greener *and* more beautiful *than other countries.*
New statement: *Chinese people are* nicer *and* more polite *than other people.*

New Zealand may be unfamiliar to Chinese students, and they may not know whether New Zealand is greener and more beautiful than other countries. Accordingly, if they had read the original statement, it would have been very hard for them to indicate whether they agreed with, disagreed with, or were not sure about the truth of this statement.

The Timed Grammaticality Judgment Test (TGJT). The TGJT used in the current study is a computer-delivered test consisting of 68 sentences (see Appendix C), evenly divided between grammatical and ungrammatical sentences. Thus, four sentences were judged for each of the 17 grammatical structures, with two grammatical and two ungrammatical exemplars per target structure. Examples of sentences for modals are presented in (5):

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

(5) *I must to brush my teeth now.* (Ungrammatical)
I can to speak French very well. (Ungrammatical)
I must finish my homework tonight. (Grammatical)
I can cook Chinese food very well. (Grammatical)

The sentences, which were different from those in the imitation test, were presented in written form on the computer screen. The test takers were required to indicate whether each sentence was grammatical or ungrammatical by pressing response buttons within a time limit. In R. Ellis (2005), the time limit for each sentence was established by timing NSs' performance in a pilot study, calculating an average response time for each sentence, and then adding an additional 20% for each sentence to allow for the slower processing speed of L2 learners. So, the time allowed to judge the individual sentences ranged from 1.8 to 6.24 s in R. Ellis's study. The same range was used in the current study. Items were scored dichotomously as correct or incorrect; items not responded to were scored as incorrect. Accuracy scores for the full test and for the grammatical and ungrammatical items were calculated as percentages.

The Untimed Grammaticality Judgment Test (UGJT). This paper-based test was the same as the TGJT in that participants read sentences but was different in that the sentences were administered via pencil and paper rather than via computer. The test takers were required to (a) indicate whether each sentence was grammatical or ungrammatical, (b) indicate the degree of certainty of their judgment by a certainty score from 0% to 100% (as proposed by Sorace, 1996), and (c) self-report whether they used "rule" or "feel" for each sentence. This test provided three separate measures: an overall percentage judgment accuracy score based on the participants' dichotomous responses and separate scores for the grammatical/ungrammatical items, a percentage certainty score, and a percentage score for the participants' reported use of "rule" in judging each item.

It should be noted that R. Ellis's (2009b) study employed a computer-delivered version of this test, which prevented the participants from revising their answers after they had finished judging each sentence. In contrast, the paper version allowed the test takers to revise their answers at any time during the test. In addition, the computerized version of the test could record the time taken to judge each sentence, whereas the paper version could not. However, the different ways of administering the test are unlikely to have greatly affected the type of knowledge that the participants accessed because, with no time limit, the students could still access their explicit knowledge.[4]

The Metalinguistic Knowledge Test (MKT). The MKT used in the current study was an adaptation of an earlier metalanguage test devised by

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

Alderson, Steel, and Clapham (1997). It consisted of an untimed, computerized multiple-choice test in two parts. Part 1 presented the test takers with 17 ungrammatical sentences (see Appendix D) corresponding to the 17 structures (see Appendix A) and required them to select the rule that best explained each error out of four choices provided, as in the example in (6):

(6) Hiroshi wants visiting the United States this year.
   a. "Visiting" should be written in the base form.
   b. The verb following "want" must be an infinitive.
   c. We cannot have two verbs together in a sentence.
   d. It should be "visit" because the event is in the future.

As seen here, grammar terms like *infinitive*, *verbs*, and *imperative* appeared in the choices.

Part 2 had two sections (see Appendix D). In Section 1, the students were asked to find the corresponding item for each grammatical feature from a given passage. For example, *the* corresponded to the feature *definite article*. Section 2 contained four sentences; participants needed to find the item corresponding to the grammatical feature in each sentence, as in (7):

(7) *Poor little Joe stood out in the snow.* (Subject)

RESULTS

Reliability and Construct Validity of the Tests

A series of statistical analyses were carried out to investigate the reliability and construct validity of the four tests. Table 2 shows the reliability coefficient for each of the four tests and also for the grammatical and ungrammatical items of both GJTs. According to DeVellis (1991), Cronbach's alpha values ranging from .65 to .70 indicate acceptable reliability, whereas values from .70 to .80 indicate fairly good reliability and thus very good consistency. As seen in the table, all alphas for all tests had high reliability for this learner population (> .70).

Table 3 shows the mean and standard deviations of scores on the four tests. It can be seen that the participants scored highest on the UGJT (M = .86), followed by the MKT (M = .56), and then the TGJT (M = .51). They achieved the lowest scores on the EIT (M = .44).

Table 4 shows the correlations among the four tests. All four measures were significantly intercorrelated, varying between Pearson r values of .29 to .52. According to Cohen (1988) and Cohen and Cohen (2003), a correlation coefficient ranging from .10 to .29 is thought to represent a weak or small correlation, one from .30 to .49 a moderate correlation,

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

Table 2. Reliability Values of the Tests (N = 100)

Test/Submeasures	Reliability (alpha)
EIT	.76
TGJT	.70
UGJT	.72
MKT	.76
TGJT G	.83
TGJT UG	.63
UGJT G	.70
UGJT UG	.72

Note. EIT = elicited imitation test; TGJT = timed grammaticality judgment test; TGJT G = timed grammaticality judgment test grammatical items; TGJT UG = timed grammaticality judgment test ungrammatical items; UGJT = untimed grammaticality judgment test; UGJT G = untimed grammaticality judgment test grammatical items; UGJT UG = untimed grammaticality judgment test ungrammatical items; MKT = metalinguistic knowledge test.

and one of .50 or larger a strong or large correlation. Hence, among all the correlations, the EIT and MKT correlated most strongly with the UGJT (r = .52 and .51, respectively). The TGJT correlated most strongly with the EIT (r = .36).

Two Pearson product-moment correlations were then conducted between the participants' self-reported use of "rule" and "feel" (respectively) and both the grammatical and ungrammatical items. It can be seen from Table 5 that "rule" had a stronger and positive relationship with the ungrammatical items than the grammatical ones (r = .43, p < .01), whereas "feel" had a stronger, but negative, relationship with the ungrammatical items. Such results further support the claim that the ungrammatical items constitute a better measure of explicit knowledge.

Table 3. Descriptive Statistics of the Tests (N = 100)

Tests	Items	*M*	*SD*
EIT	34	.44	.12
TGJT	68	.51	.10
UGJT	68	.86	.08
MKT	40	.56	.12
TGJT G	34	.69	.18
TGJT UG	34	.33	.12
UGJT G	34	.85	.10
UGJT UG	34	.87	.10

Note. EIT = elicited imitation test; TGJT = timed grammaticality judgment test = TGJT G: timed grammaticality judgment test grammatical items; TGJT UG = timed grammaticality judgment test ungrammatical items; UGJT = untimed grammaticality judgment test; UGJT G = untimed grammaticality judgment test grammatical items; UGJT UG = untimed grammaticality judgment test ungrammatical items; MKT = metalinguistic knowledge test.

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

Table 4. Correlations among the Four Tests (N = 100)

	EIT	TGJT	UGJT	MKT
EIT	—	.36**	.52**	.41**
TGJT		—	.34**	.29**
UGJT			—	.51**

Note. EIT = elicited imitation test; TGJT = timed grammaticality judgment test; UGJT = untimed grammaticality judgment test; MKT = metalinguistic knowledge test.
** $p < .01$.

According to Roberts (1999), CFA is often conducted when researchers have "an understanding of the constructs that underlie the data" (p. 3). Confirmatory factor analysis is theory or hypothesis driven rather than data driven. Confirmatory factor analysis allows researchers to test the hypothesis by using their "knowledge of the theory, empirical research, or both" to postulate "the relationship pattern a priori" and then test "the hypothesis statistically" (Suhr, 2006, p. 200).

Two CFAs were run using AMOS (version 19.0), one testing the two-factor implicit/explicit model reported in R. Ellis (2005)[5] and in Ellis and Loewen (2007), and the other testing the grammatical/ungrammatical model proposed in Gutiérrez (2012). In R. Ellis's model, the maximum likelihood method was employed to estimate the model fitting of this sample (N = 100). Figures 1 and 2 and Table 6, respectively, report some important model fit indices. Chi-squared (χ^2) is the most common method of evaluating goodness of fit. A low χ^2 value, indicating nonsignificance, would point to a good fit, whereas a significant χ^2 value ($p < .05$) indicates the model is statistically unlikely to occur. This model had a nonsignificant χ^2 value, indicating acceptability. It should be further noted that a normed fit index (NFI) value of greater than .90 indicates a good model fit, and one greater than .95 indicates a superior fit. Also, root mean square error of approximation (RMSEA) has been regarded as one of the most informative criteria for evaluating model fit because it takes into account the error of approximation in the population (Byrne, 2001). It has been advised that values less than .05 indicate a good fit for the model, values as high as .08 are considered an acceptable

Table 5. Correlations between the Use of Rule and the Test Measures

	UGJT G	UGJT UG
Rule	.24*	.43**
Feel	–.33*	–.34*

Note. UGJT G = untimed grammaticality judgment test grammatical items; UGJT UG = untimed grammaticality judgment test ungrammatical items.
* $p < .05$. ** $p < .01$.

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

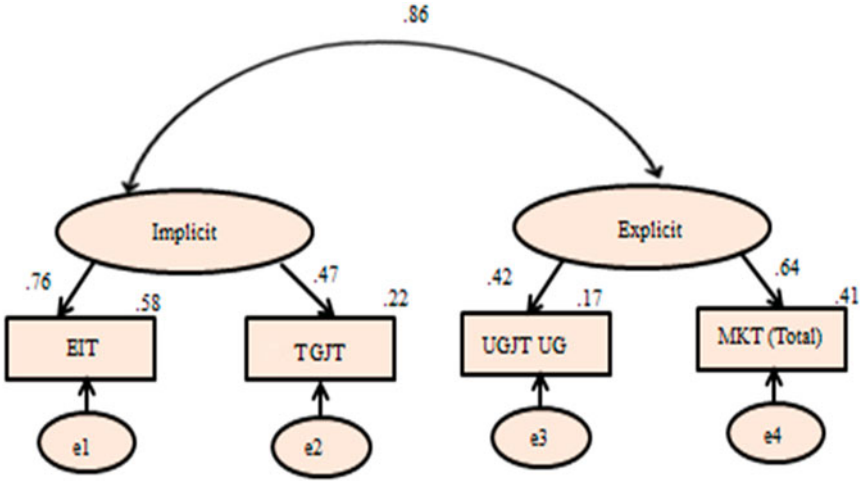


Figure 1. CFA for first-year students' implicit versus explicit knowledge, testing R. Ellis's model. EIT: elicited imitation test; TGJT: timed grammaticality judgment test; UGJT UG: untimed grammaticality judgment test ungrammatical items; MKT: metalinguistic knowledge test.

model fit, values ranging from .08 to .10 are considered a mediocre model fit, and those greater than .10 indicate a poor model fit (Byrne, 2001). Table 6 shows that the NFI was greater than .95 and the RMSEA less than .05, indicating that this model fits the present data.

The results of the CFA thus further support the construct validity of the testing instruments. In R. Ellis's model, the EIT and TGJT were thought more likely to elicit implicit knowledge, whereas the ungrammatical items in the UGJT and MKT were considered the best measures of explicit knowledge. The results of the CFA indicated that R. Ellis's two-factor model, which distinguishes between measures intended to tap implicit and explicit knowledge, was replicated for the present sample. For the grammatical/ungrammatical model in Gutiérrez (2012), χ^2 and RMSEA values indicated it to be acceptable; however, the value for NFI was lower than .90, indicating that this model did not work well for the sample in the current study.

To investigate the construct validity of the test battery in a more fine-grained way, it is helpful to compare the performance of the L2 English learners in the present study to those of the English NSs and L2 English learners in R. Ellis's study. For clarity, *nonnative speakers* (NNSs) is used here to refer to the L2 learner group in R. Ellis (2005), whereas *L2 learners* is used to refer to the L2 learner group in the current study. Table 7 shows the comparison of mean scores and standard deviations on the four tests in R. Ellis's (2005) study and in the present study.[6]

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

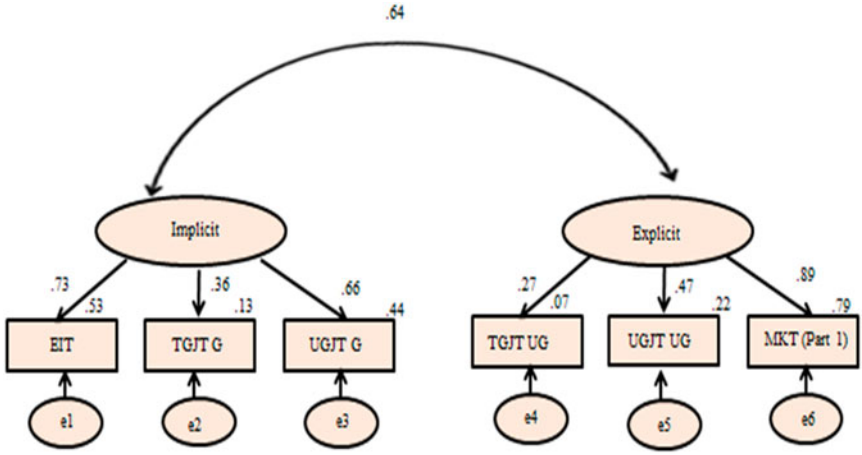


Figure 2. CFA for first year students' implicit versus explicit knowledge, testing Gutiérrez's model. EIT: elicited imitation test; TGJT: timed grammaticality judgment test; TGJT G: timed grammaticality judgment test grammatical items; TGJT UG: timed grammaticality judgment test ungrammatical items; UGJT G: untimed grammaticality judgment test ungrammatical items; UGJT UG: untimed grammaticality judgment test ungrammatical items; MKT: metalinguistic knowledge test.

The NSs in R. Ellis (2005) performed near the ceiling on the EIT and UGJT (M = .94 and .96)—that is, much better than the NNSs (in Ellis's study; M = .51 and .82) or the L2 learners (in the present study; M = .54 and .86). They also performed much better on the TGJT (M = .80) than the NNSs (M = .54) or the L2 learners (M = .51). However, their performance on the MKT (M = .57), a test that relies on grammatical knowledge and rules, was quite close to those of the NNSs (M = .53) and the L2 learners (M = .56). Effect sizes were then calculated to investigate whether time pressure significantly affected accuracy for all three groups on the TGJT and UGJT. The largest effect size (d = 3.87), found for the L2 learners, is much larger than those for the NSs (d = 2.22) and the NNSs (d = 2.43), thus demonstrating that time was a much greater factor for the L2 learners in the current study than for either the NSs or the NNSs in R. Ellis (2005).

Table 6. Summary of Model Fit

Model	χ^2	NFI	RMSEA	*df*
Implicit/explicit	.96 (p = .326)	.98	.00	1
Grammatical/ungrammatical	1.34 (p = .220)	.89	.06	8

Note. NFI = normed fit index; RMSEA = root mean square error of approximation.

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

Table 7. Comparison of the Mean Scores and *SD*s

Test	R. Ellis (2005)				Present study	
	NSs		NNSs		L2 learners	
	M	*SD*	*M*	*SD*	*M*	*SD*
EIT	.94	.04	.51	.17	.44	.12
TGJT	.80	.10	.54	.12	.51	.10
UGJT	.96	.02	.82	.11	.86	.08
MKT	.57	.07	.53	.21	.56	.12

Note. EIT = elicited imitation test; TGJT = timed grammaticality judgment test; UGJT = untimed grammaticality judgment test; MKT = metalinguistic knowledge test.

First-Year Students' Implicit and Explicit Knowledge

As shown previously, the EIT and TGJT were more likely to elicit implicit knowledge, and the UGJT ungrammatical items and the MKT were more likely to measure explicit knowledge. Therefore, four paired-samples *t* tests were conducted to see whether there were statistically significant differences between the participants' scores on the measures of implicit knowledge and on those of explicit knowledge. The results are presented in Table 8. Four statistically significant differences were found between the participants' scores on the measures of implicit knowledge and on those of explicit knowledge. Overall, all participants' scores on the measures of explicit knowledge (i.e., the UGJT UG and MKT) were higher than on those of implicit knowledge (i.e., the EIT and TGJT).

DISCUSSION

The current study was primarily concerned with investigating to what extent Chinese first-year university students have developed implicit and explicit L2 grammatical knowledge. To investigate this issue, it is crucial to examine whether the tests in R. Ellis (2005) can also provide relatively separate measures of these two types of knowledge for this pool of participants. The first research question addressed the issue of validity, asking whether scores on the four tests would load on two separate factors—one representing implicit knowledge and the other representing explicit knowledge. If the tests in R. Ellis (2005) are replicable and tap different sources of knowledge, then the tests adapted for this study in an EFL context in which L2 learners learned English mainly in the classroom should load on separate factors. The results provide evidence that these tests tapped relatively separate pools of knowledge.

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

Table 8. Paired-Samples *t* tests of First-Year Students' Scores on Measures of Implicit and Explicit Knowledge (df = 99)

	UGJT UG			MKT		
	SEM	*t*	Sig.(2-tailed)	*SEM*	*t*	Sig.(2-tailed)
EIT	0.01	–31.99	.000	0.01	–26.38	.000
TGJT	0.01	–9.46	.000	0.01	–3.63	.000

Note. EIT = elicited imitation test; TGJT = timed grammaticality judgment test; UGJT UG = untimed grammaticality judgment test ungrammatical items; MKT = metalinguistic knowledge test; *SEM* = standard error mean.

The comparison of the performances of the NSs and NNSs in R. Ellis (2005) and the L2 learners in the present study should provide more fine-grained evidence for the construct validity of the test battery. If the tests really measure relatively separate pools of linguistic knowledge, then the L2 learners, who learned English in a classroom setting and had no experience living in an English-speaking country, should score lower on tests of implicit knowledge than on tests of explicit knowledge. In contrast, the NNSs in R. Ellis (2005), who also received several years of formal instruction but had around 2 years' experience living in an English-speaking country, should score higher for implicit knowledge than the L2 learners in the current study but much lower than the NSs in R. Ellis (2005). However, the NSs should not perform near the ceiling on the MKT because it "relies on grammatical terminology and rules that NSs most likely do not access frequently" (Bowles, 2011, p. 253).

The results shown in Table 7 provide such evidence. The L2 learners in the present study scored lowest on the tests of implicit knowledge but outperformed the NNSs on measures of explicit knowledge—the UGJT (.86),[7] followed by the MKT (.56). The NSs' mean score on the MKT was almost the same as that of the L2 learners in the present study and the NNSs in R. Ellis's study, meaning their performance on this measure was far less than the ceiling, as expected. These findings point to the content validity of the battery of tests because the measures appear to distinguish between the L2 learners, who had lower implicit knowledge and higher explicit knowledge, and the NNSs, who had higher implicit knowledge and lower explicit knowledge.

A comparison of the participants' scores on the UGJT and TGJT provides additional support for the validity of the implicit-explicit model proposed by R. Ellis (2005). The NSs, NNSs, and L2 learners all scored much lower on the TGJT than they did on the UGJT.[8] The results indicate that time pressure affected accuracy for all three groups to a significant extent, as the large effect sizes reported indicate. The result in which the largest effect size was found among the L2 learners in the

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

present study most likely occurred because they relied more heavily than the other two groups on their explicit knowledge to complete the UGJT. The results, therefore, provide support for the construct validity of the battery of tests in R. Ellis (2005) among a different pool of L2 learners in a different context. The results of the current study may bring the field of SLA research one step closer to establishing reliable and valid measures of implicit and explicit knowledge. The results could also direct the attention of researchers to the continued relevance of long-standing issues such as "the relationship between type of instruction and the nature of resulting linguistic knowledge" (Bowles, 2011, p. 262).

The results for the first research question also provide a basis on which to answer the second research question, which addressed the extent of Chinese first-year university students' implicit and explicit knowledge of English. The results of the paired-samples *t* tests for the participants' scores on the measures of implicit and explicit knowledge, shown in Table 8, indicate that their implicit knowledge and their explicit knowledge of English were significantly different. They obtained much higher scores on both explicit knowledge measures than on the two implicit knowledge measures (as shown in Table 8). In other words, the first-year English majors in the present study seemed to have much greater explicit knowledge of English than implicit knowledge, as anticipated. The English instruction this group of learners had received prior to entering university was mainly form focused and exam directed. They had no experience living in an English-speaking country and had little, if any, experience communicating in English with English NSs. Thus, they were likely to have primarily developed explicit knowledge. This fact indicates that EFL countries like China may not provide the necessary conditions (e.g., a large number of native English-speaking teachers or English-speaking environments outside the classroom, etc.) for English learners to develop their implicit knowledge of English. Thus, the questions that arise include, "How can we create more communicative conditions for learners to foster their implicit knowledge?" and even, "Can we really create communicative conditions for learners to foster their implicit knowledge in EFL countries like China?" Regarding the first question, although traditional methods such as forms-focused instruction and task-based group discussion in and after class are worth employing, more fine-grained methods are still needed. In terms of the second concern, it may be worth taking Spada's (2013) suggestion into account. She argues that a potentially more realistic goal for L2 teachers in settings like China is to employ activities that are more likely to lead to the proceduralization of explicit knowledge, because proceduralized knowledge is qualitatively different from but functionally equivalent to implicit knowledge (DeKeyser, 2003).

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

IMPLICATIONS AND LIMITATIONS

The current study was conducted to validate the measures of implicit and explicit knowledge with a population of university-level learners of English in a foreign language context and also to measure to what extent these students possess each of these two types of knowledge. Although the results suggest that the two aims were largely realized, there are a number of limitations of the study.

Due to the fact that reliability is a property of how a particular sample responded to the testing instruments, it cannot be assumed that the instruments themselves are reliable, although the reliability of the tests for the sample investigated in the current study was established.

The ONT, used in R. Ellis's (2005) study to elicit implicit knowledge, was not included in the current study. With hindsight, the inclusion of the ONT may have provided a better understanding of the students' knowledge pattern because this test does not create any situation for the occurrence of simple imitation and rote repetition, which may possibly occur in the EIT.

In addition, unlike R. Ellis's (2005) study, in which the UGJT was delivered by computer, a paper version of this test was used in the current study. Although the two versions were exactly the same in content, it was quite possible that learners in the current study were able to revise their answers. Therefore, they might have been more accurate in judging the items than the learners in R. Ellis (2005).

Finally, there is the possibility that the tests employed may not be equally difficult and thus that comparability among them may be questioned. It cannot be denied that the nature and form of the EIT make it the most difficult test for the participants to complete, which may be a contributing factor to their lower score on this test as compared to the other three tests. Some other uncontrolled variables may have affected their scores on this test as well. For example, it is quite possible that learners who have sufficient knowledge of phonology and sufficient memory capacity and articulatory skills may perform well on the EIT. Similarly, some strategies, such as guessing, may have been used when the learners did the other three tests. Employing additional and more fine-grained tests that are predicted to elicit implicit and explicit knowledge would be a way to solve this problem. Accordingly, it is hoped that future studies will investigate further ways of operationalizing implicit and explicit knowledge by taking other factors—such as the speed of information processing, WM capacity (as in Erçetin & Alptekin, 2013), pronunciation, listening comprehension, and receptive vocabulary—into consideration.

Received 18 June 2013
Accepted 5 March 2014
Final Version Received 18 May 2014

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

NOTES

1. Most of the participants did not have good listening or speaking abilities, because their teachers paid more attention to grammar and reading.

2. A principal component factor analysis was carried out in R. Ellis's (2005) study. Then two confirmatory factor analyses were carried out in response to Isemonger's (2007) query in R. Ellis and Loewen (2007). The results were included in R. Ellis (2009b). That is why I put two year numbers in brackets.

3. Key and nonkey universities are distinguished according to whether the university is a member of Project 211, a constructive project of nearly 100 universities and disciplines conducted by the Chinese government in the 21st century. The project aims to cultivate high-level talented individuals for national economic and social development strategies.

4. The participants may have analyzed subsequent items and revised previous items; however, the type of knowledge drawn from in this test would have still been explicit knowledge.

5. However, R. Ellis's (2005) model was not tested fully because the ONT was not included in the current study.

6. It should be noted that it would have been better to conduct one-way ANOVAs to examine whether these three groups had statistically significant differences on all four tests. However, there was no way to access the individual scores of the two groups in R. Ellis (2005); therefore, it was not possible to conduct one-way ANOVAs.

7. An anonymous *SSLA* reviewer commented that the higher score of the L2 learners in the present study on the UGJT compared to that of the NNSs in R. Ellis (2005) may be a reflection of the fact that learners in my study were able to revise their answers. They may be more accurate in judging the items than the learners in R. Ellis (2005). This does indeed seem to be a possibility.

8. See Note 6.

REFERENCES

Alderson, J. C., Steel, D., & Clapham, C. (1997). Metalinguistic knowledge, language aptitude and language proficiency. *Language Teaching Research*, *1*, 93–121.

Bialystok, E. (1978). A theoretical model of second language learning. *Language Learning*, *28*, 69–83.

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

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

Cohen, J., & Cohen, P. (2003). *Applied multiple regression/correlation analysis for the behavioural sciences*. Mahwah, NJ: Erlbaum.

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

DeKeyser, R. (2003). Implicit and explicit learning. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 313–348). Malden, MA: Blackwell.

DeVellis, R. F. (1991). *Scale development: Theory and applications*. Newbury Park, NJ: Sage.

Doughty, C. (2003). Instructed SLA: Constraints, compensation, and enhancement. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 256–310). Malden, MA: Blackwell.

Elder, C., & Ellis, R. (2009). Implicit and explicit knowledge of an L2 and language proficiency. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 167–193). Bristol, UK: Multilingual Matters.

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

Ellis, N. (Ed.). (1994). *Implicit and explicit learning of languages*. London, UK: Academic Press.
Ellis, N. (2005). At the interface: Dynamic interactions of explicit and implicit knowledge. *Studies in Second Language Acquisition*, *27*, 305–352.
Ellis, N. (2008). Implicit and explicit knowledge about language. In J. Cenoz & N. Hornberger (Eds.), *Encyclopedia of language and education* (pp. 119–131). New York, NY: Springer.
Ellis, R. (1991). Grammaticality judgments and second language acquisition. *Studies in Second Language Acquisition*, *13*, 161–186.
Ellis, R. (1993). Second language acquisition and the structural syllabus. *TESOL Quarterly*, *27*, 91–113.
Ellis, R. (1994). *The study of second language acquisition*. Oxford, UK: Oxford University Press.
Ellis, R. (1997). *Second language acquisition*. Oxford, UK: Oxford University Press.
Ellis, R. (2001). Investigating form-focused instruction. *Language Learning*, *51*(Suppl. 1), 1–46.
Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, *54*, 227–275.
Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, *27*, 141–172.
Ellis, R. (2008a). Explicit knowledge and second language learning and pedagogy. In J. Cenoz & N. Hornberger (Eds.), *Encyclopedia of language and education* (pp. 143–153). New York, NY: Springer.
Ellis, R. (2008b). *The study of second language acquisition* (2nd ed.). Oxford, UK: Oxford University Press.
Ellis, R. (2009a). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3–25). Bristol, UK: Multilingual Matters.
Ellis, R. (2009b). Measuring implicit and explicit knowledge of a second language. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 31–64). Bristol, UK: Multilingual Matters.
Ellis, R., & Loewen, S. (2007). Confirming the operational definitions of explicit and implicit knowledge in Ellis (2005): Responding to Isemonger. *Studies in Second Language Acquisition*, *29*, 119–126.
Erçetin, G., & Alptekin, C. (2013). The explicit/implicit knowledge distinction and working memory: Implications for second-language reading comprehension. *Applied Psycholinguistics*, *34*, 727–753.
Green, P., & Hecht, K. (1992). Implicit and explicit grammar: An empirical study. *Applied Linguistics*, *13*, 168–184.
Gutiérrez, X. (2012). Implicit knowledge, explicit knowledge, and achievement in second language (L2) Spanish. *Canadian Journal of Applied Linguistics*, *15*, 20–41.
Han, Y., & Ellis, R. (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research*, *2*, 1–23.
Hu, G. (2002). Psychological constraints on the utility of metalinguistic knowledge in second language production. *Studies in Second Language Acquisition*, *24*, 347–386.
Hulstijn, J. H. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research*, *18*, 193–223.
Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second language learning. *Studies in Second Language Acquisition*, *27*, 129–140.
Isemonger, I. (2007). Operational definitions of explicit and implicit knowledge: Response to Ellis (2005) and some recommendations for future research in this area. *Studies in Second Language Acquisition*, *29*, 101–118.
Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford, UK: Pergamon.
Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford, UK: Pergamon.
Krashen, S. (1985). *The input hypothesis issues and implication*. London, UK: Longman.

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

Kumaravadivelu, B. (2006). *Understanding language teaching: From method to postmethod*. Mahwah, NJ: Erlbaum.
Macrory, G., & Stone, V. (2000). Pupil progress in the acquisition of the perfect tense in French: The relationship between knowledge and use. *Language Teaching Research*, *4*, 55–82.
Milasi, A. F., & Pishghadam, R. (2007). The interplay between explicit and implicit knowledge of English native speakers and ESL learners. *Iranian Journal of Applied Linguistics*, *10*(1), 15–43.
Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications of bilingualism and SLA. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 393–419). San Diego, CA: Academic Press.
Paradis, M. (2009). *Neurolinguistic aspects of bilingualism*. Amsterdam, the Netherlands: Benjamins.
Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics*, *10*, 52–79.
Reber, A. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*, 219–235.
Richards, J. C., Platt, J., & Weber, H. (1985). *Longman dictionary of applied linguistics*. London, UK: Longman.
Roberts, J. K. (1999, January). *Basic concepts of confirmatory factor analysis*. Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, TX. Retrieved from http://www.eric.ed.gov/PDFS/ED427091.pdf
Roehr, K. (2008). Metalinguistic knowledge and language ability in university-level L2 learners. *Applied Linguistics*, *29*, 173–199.
Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129–158.
Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, *11*, 11–26.
Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: A case-study of an adult learner of Portuguese. In R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237–326). Rowley, MA: Newbury House.
Schwartz, B. D. (1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition*, *15*, 147–163.
Sharwood Smith, M. (1981). Consciousness-raising and the second language learner. *Applied Linguistics*, *2*, 159–169.
Sorace, A. (1996). The use of acceptability judgments in second language acquisition research. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 375–409). San Diego, CA: Academic Press.
Spada, N. (2013). SLA research and L2 pedagogy: Misapplications and questions of relevance. *Language Teaching*. Advance online publication. doi:10.1017/S026144481200050X
Suhr, D. D. (2006). Exploratory or confirmatory factor analysis? In K. J. LeBouton (Ed.), *Proceedings of the thirty-first annual SAS Users Group International Conference* (pp. 200–231). Cary, NC: SAS Institute.
Zobl, H. (1995). Converging evidence for the "acquisition–learning" distinction. *Applied Linguistics*, *16*, 35–56.

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

APPENDIX A

REPRESENTATIVE GRAMMATICAL STRUCTURES

Structure	Exemplars	Acquisition	Pedagogical introduction	Type
Verb complements	Liao says he wants *to buy* a new car.	Early	Junior high school	Syntactic
Regular past tense	Martin *completed* his assignment yesterday.	Intermediate	Primary/junior high schools	Morphological
Question tags	We will leave tomorrow, *won't we?*	Late	Senior high school/university	Syntactic
Yes/*no* questions	Did Keiko *complete* her homework?	Intermediate	Primary/junior high schools	Morphological
Modal verbs	I *must brush* my teeth now.	Early	Various levels	Morphological
Unreal conditions	If he had been richer, she *would have* married him.	Late	Senior high school/university	Syntactic
Since and *for*	He *has been living* in New Zealand *for* three years.	Intermediate	Junior high school	Syntactic
Indefinite articles	They had *a* very good time at the party	Late	Primary school	Morphological

APPENDIX B

A REPRESENTATIVE SAMPLE OF SENTENCES IN THE EIT (G = GRAMMATICAL; UG = UNGRAMMATICAL)

Comparatives

*Chinese people are **nicer** and **more polite** than other people.* (G)
*It is **more harder** to learn math than to learn chemistry.* (UG)

Verb Complements

*Most college students **want to get** high marks in their examinations.* (G)
*Young people usually **want leaving** their families as soon as possible.* (UG)

Adverb Placement

*People play Ping Pong **well** and football **badly** in China.* (G)
*Young people play **often** internet games and drink **a lot**.* (UG)

Dative Alternation

*People should **report the police stolen money**.* (UG)
*People should **report a car accident to the police**.* (G)

APPENDIX C

A REPRESENTATIVE SAMPLE OF SENTENCES IN THE TIMED AND UNTIMED GRAMMATICALITY TESTS (G = GRAMMATICAL; UG = UNGRAMMATICAL)

Since (G): *I haven't seen him for a long time.*
Comparative (G): *I think that he is nicer and more intelligent than all the other students.*
Dative (G): *The teacher explained the problem to the students.*
Verb complement (UG): *Liao says he wants buying a car next week.*
Past *-ed* (UG): *Martin completed his assignment and print it out.*
Tag (UG): *We will leave tomorrow, isn't it?*

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press

Adverb (G): *He plays soccer very well.*
Auxiliary *do* (UG): *Did Keiko completed her homework?*
Modal (UG): *I must to brush my teeth now.*
Conditional (UG): *If he had been richer, she will marry him.*
Since (UG): *He has been living in New Zealand since three years.*
Reported (G): *Pam wanted to know what I had told John.*
Article (UG): *They had the very good time at the party.*
Passive (UG): *Between 1990 and 2000 the population of New Zealand was increased.*
Possessive (UG): *Liao is still living in his rich uncle house.*
Plural (UG): *Martin sold a few old coins and stamp to a shop.*
Since (UG): *I have been studying English since a long time.*

APPENDIX D

A REPRESENTATIVE SAMPLE OF SENTENCES IN THE METALINGUISTIC KNOWLEDGE TEST

Metalinguistic Knowledge Test, Part 1

(1) You must to wash your hands before eating.
   a. "Must to" is the wrong form of the imperative.
   b. Change to "must have to wash" to express obligation.
   c. Modal verbs should never be followed by a preposition.
   d. After "must" use the base form of the verb not the infinitive.

Metalinguistic Knowledge Test, Part 2

(1) Read the passage below. Find ONE example in the passage for each of the grammatical features listed in the table. Write the examples in the table in the spaces provided. The first one is done for you. Note: it may be possible to choose the same example to illustrate more than one grammatical feature.

Passage:
The materials are delivered to the factory by a supplier, who usually has no technical knowledge, but who happens to have the right contacts. We would normally expect the materials to arrive within three days, but this time it has taken longer.

Grammatical feature	Example
Definite article	*The*

https://doi.org/10.1017/S0272263114000370 Published online by Cambridge University Press