 doi: 10.3969/j.issn.1674-8921.2014.12.004

Development of Implicit and Explicit Knowledge of Grammatical Structures in Chinese EFL Learners

MAN LI
University of Maryland, USA

XIAOQING QIN
Central China Normal University, China

Although the theoretical distinction between implicit and explicit knowledge of a second language has long been recognized, minimal documented description exists of the development of the two types of knowledge in L2 learners due to the difficulty in finding separate measures of the two types. To address this gap, this study first investigated the construct validity of the four tests developed by R. Ellis (2005) intended to distinguish between the two knowledge types: an oral imitation test, a timed grammaticality judgment test (GJT), an untimed GJT and a metalinguistic knowledge test. The four tests, all examining 17 English grammatical structures, were administered to 88 native speakers of Chinese studying English as their major at the first three year levels in a Chinese university. The results from a confirmatory factor analysis (CFA) confirmed the two-factor solution for the implicit/explicit model proposed in R. Ellis (2005). Repeated measures ANOVA was then used to reveal developmental trends of the two knowledge types in the EFL learners across the three year levels. The results showed that learners at higher year level (or higher proficiency level) generally had higher level of both implicit and explicit knowledge of grammatical structures, and overall, the Chinese EFL learners possessed higher level of explicit knowledge than implicit knowledge. With regard to individual grammatical structures, the results of this study confirmed R. Ellis' (2006b) finding that the level of difficulty of grammatical structures varied according to whether it was implicit or explicit knowledge of the structures involved.

Keywords: implicit knowledge, explicit knowledge, grammatical structure, L2 development, Chinese learners of English

INTRODUCTION

A theoretical distinction between implicit and explicit learning of a second language was first proposed by Krashen (1981) and Bialystok (1978). This distinction at the *process* level between the two types of learning entailed the distinction at the *product* level between implicit and explicit language knowledge that an L2 learner may possess. These distinctions as well as the interfaces between them (both at the *process* level and at the *product* level) have attracted considerable debate and research in the fields of both cognitive psychology and second language acquisition (SLA) over the last three decades (e.g., DeKeyser, 1995, 1998, 2003; N. C. Ellis, 1994, 2005, 2007; R. Ellis, 1994b, 2004, 2005; Hulstijn, 2005; Krashen, 1981, 1994; Paradis, 1994; Tokowicz & MacWhinney, 2005). As one major goal of SLA research is to describe the development of L2 linguistic knowledge (R. Ellis, 1994a, 2005; R. Ellis, Loewen, & Elder, 2009), the distinction and relationship between implicit and explicit knowledge are of particular interest in SLA.

Though the existence and distinction of implicit and explicit knowledge are now commonly held (e.g., DeKeyser, 2003; N. C. Ellis, 1994, 2007; R. Ellis, 1994b, 2004, 2005; R. Ellis *et al.*, 2009; Krashen, 1981, 1994; Paradis, 1994), the relationship between the two types of knowledge, which has been discussed in terms of the interface between them (N. C. Ellis, 1994), remains a central topic of continuous debate in SLA. Specifically, the issue mainly concerns the role of explicit knowledge in the development of implicit knowledge, and the question asked is whether explicit knowledge of a second language can become implicit (DeKeyser, 2003; N. C. Ellis, 1994, 2005; R. Ellis, 1997, 2005; Krashen, 1981, 1994; Paradis, 1994).

Regarding this question, researchers have identified three distinct positions—the non-interface, the strong interface and the weak interface. Krashen firmly holds to the non-interface position arguing for the separateness of implicit ("acquired") and explicit ("learned") knowledge. He rejects the possibility of transfer of knowledge from explicit to implicit. This

Correspondence should be addressed to Man Li, Second Language Acquisition, 3215 Jiménez Hall, University of Maryland, College Park, MD 20742, USA. Email: manli@umd.edu

position has been forcibly argued by Paradis (1994), who finds neurological evidence for the separate representation of the two types of knowledge. Acknowledging the possibility of interaction between the two knowledge systems, Paradis disputes that there is transformation between the two types of knowledge. The strong interface position, as argued by DeKeyser (1998, 2003), on the contrary, holds that explicit knowledge can be turned into implicit through practice. In his later work, he further clarifies that the strong interface hypothesis does not imply that the more there is of implicit knowledge, the less there is of explicit knowledge in the sense of metamorphosis (DeKeyser, 2009). On the other hand, DeKeyser (2009) posits that the presence of explicit knowledge plays a "causal" role in the development of implicit knowledge (p. 127).

The weak interface position taken by R. Ellis (1993) claims that explicit knowledge can convert into implicit knowledge only when the learner is developmentally ready. R. Ellis (2006a) later proposes that explicit knowledge can facilitate the processes responsible for implicit learning, such as noticing the gap. Acknowledging that implicit and explicit knowledge are dissociative because they reside in different parts of the brain and involve different types of representations, N. C. Ellis (2005) argues for an indirect role of explicit knowledge as a facilitator of the development of implicit knowledge when he claims that a dynamic interface between explicit and implicit knowledge happens "transiently during conscious processing, but influence upon implicit cognition endures thereafter" (p. 305).

The problem regarding the above debate is that discussions on these positions have stagnated at the level of theoretical modeling and are in need of subjecting to empirical inquiry (R. Ellis, 2005). Developing valid and reliable tests that measure implicit and explicit knowledge separately is essential to confirming whether explicit knowledge of a second language can become implicit (DeKeyser, 2003). R. Ellis (2005) makes an important attempt in SLA in this regard. In his study (R. Ellis, 2005), he established operational definitions of the two constructs, based on which five tests were developed to measure implicit and explicit knowledge of seventeen grammatical structures. In a follow-up study, R. Ellis (2006b) administered four of these five tests to a larger population of L2 learners to investigate learning difficulty of specific grammatical structures in relation to the two knowledge types and the differential contribution of the two types of knowledge to L2 proficiency. Clearly, the construct validity of the tests developed in R. Ellis (2005) is in need of further validation. Given the availability of relatively separate measures of the two knowledge types, one further step along this line of research is to investigate the development of the two types of knowledge in L2 learners, which has been recognized as one major goal of SLA research.

The present study investigates the development of implicit and explicit knowledge of grammatical structures in Chinese learners of English in an EFL (English as a Foreign Language) context. The instruments used in this study were originally developed in R. Ellis (2005) and also used in his later study in 2006. This study set out to investigate the construct validity of the tests with a group of Chinese EFL learners. After the construct validity of the tests were confirmed, this study further explored the developmental trends of implicit and explicit knowledge of grammatical structures in the L2 learners at both the macro level (i.e., the overall development of implicit and explicit knowledge of grammatical structures) and the micro level (i.e., the development of the two types of knowledge of individual grammatical structures). Before delving into the details of this study, we will first review the methodological issues in the SLA literature of operationalizing the constructs of implicit and explicit knowledge and then discuss R. Ellis (2005) and R. Ellis (2006b) in more detail.

Operationalization of implicit and explicit knowledge

When developing separate measures of implicit and explicit knowledge of a second language, it is critical to operationalize the distinction between the two types of knowledge. Dating back to the first dissociation between implicit and explicit systems, Krashen (1981) distinguishes between language *acquisition*, which is "a subconscious process that results in linguistic knowledge that is subconsciously stored in the brain (tacit knowledge)" (Krashen, 1994, p. 45) and language *learning*, which is "a conscious process that results in knowledge about language" (Krashen, 1994, p. 45). He equates the *acquired*

and *learned* system with implicit and explicit knowledge (Krashen, 1982). As can be seen from his definition, the presence or absence of consciousness is what distinguishes the two types of knowledge. R. Ellis (2005) identifies seven criteria by which implicit and explicit linguistic knowledge can be distinguished—degree of awareness, time available, focus of attention, systematicity, certainty, metalinguistic knowledge and learnability (p. 152). Implicit knowledge is thus defined here as nonverbalizable procedural knowledge of language which is unconscious, unanalyzed and automatized, and is therefore easily retrievable for spontaneous language performance. Explicit knowledge, however, tends to be learned and is conscious, declarative, verbalizable, and accessible only via controlled processing when L2 learners experience linguistic difficulty in the use of L2. Among the seven criteria, R. Ellis emphasizes "conscious awareness" as the defining feature that sets the two types of knowledge apart.

Though it is now generally accepted that the criterion of *consciousness* is the intrinsic characteristic of the distinction between implicit and explicit knowledge (N. C. Ellis, 1994, 2005, 2007; R. Ellis, 2004, 2005, 2008; Hulstijn, 2005; Krashen, 1981, 1982, 1994; Paradis, 1994), consciousness is difficult, if not impossible, to subject to scientific investigation (Searle, 2000), and a number of alternatives have been developed to operationalize the distinction, such as articulability (Butler, 2002; Green & Hecht, 1992; Hu, 2002; Sorace, 1985), and automaticity (Bialystok, 1978, 1979; R. Ellis, 1984; Jiang, 2004, 2007).

In terms of the operationalization of L2 explicit knowledge, R. Ellis (2004) provides a critical review and argues that explicit knowledge should be investigated primarily as analyzed knowledge and secondarily metalanguage. The two methods that have been most often used in measuring L2 explicit knowledge are grammaticality judgment tasks (GJT) and tests of metalinguistic knowledge (R. Ellis, 2004). Using some versions of the GJT, most studies (Alderson, Clapham, & Steel, 1997; Bialystok, 1979; Green & Hecht, 1992; Han & Ellis, 1998; Hu, 2002; Masny, 1987; Sorace, 1985) asked participants to (a) identify the error in an ungrammatical sentence, (b) correct the error, and (c) verbalize the rule that has been violated, except that Green and Hecht (1992) omitted (a) and Masny (1987) omitted (c). Learners were considered to have explicit knowledge if they could state the rules in addition to correcting the errors. Though articulability as a criterion of the distinction between implicit and explicit knowledge is much easier to operationalize, and tasks that ask learners to articulate grammatical rules elicit explicit knowledge, many researchers caution that such verbal report tasks can only provide a rather conservative picture of explicit knowledge that a L2 learner may have (e.g., Bialystok, 1979; R. Ellis, 2004; Hu, 2002). L2 learners may have explicit knowledge of a rule but cannot verbalize it simply because they do not have the necessary language to articulate it (Hu, 2002). A receptive test of metalinguistic knowledge instead has been recognized as a better option. Bialystok (1979) and R. Ellis (2005) asked learners to indicate from a list of rules which rule was violated after they had judged and identified the error. A number of studies also tested learners' ability to identify named grammatical features (e.g., Alderson *et al.*, 1997; R. Ellis, 2005).

When talking about the measurement of implicit language knowledge, Hulstijn and De Graaff (1994) note that implicit knowledge is such a theoretical construct that it cannot be directly accessed by language tasks. The behavioral correlate of implicit knowledge has been generally viewed as automaticity in language use (Hulstijn & De Graaff, 1994; Schmidt, 1994). In SLA research, the instruments most frequently used include free language production (R. Ellis, 2002, 2005; Han & Ellis, 1998), oral elicited imitation (R. Ellis, 2005; Erlam, 2006), and Timed grammaticality judgment test (R. Ellis, 2005, 2006b; Han & Ellis, 1998).

Free language production has often been used to measure language gains when researchers investigate the effectiveness of L2 instruction (see reviews by R. Ellis, 2002; Norris & Ortega, 2000). It is believed that when L2 learners are engaged in unplanned, meaning-focused speech production, they rely on implicit competence rather than explicit knowledge (R. Ellis, 1994b, 2005; Schmidt, 1994). This approach of measuring implicit knowledge, however, is not free of criticism. One major criticism is that free language production tasks do not preclude the involvement of explicit knowledge. As R. Ellis (2002)

points out, "free-production tasks make it difficult but not impossible for learners to perform on the basis of explicit knowledge" (p. 234). The problem is that learners can retrieve their explicit knowledge to monitor their language production even under time pressure (Jiang, 2004). Another shortcoming of free production tasks is the difficulty in targeting the specific structures that a study aims to elicit, which further makes it difficult to rate (Erlam, 2006).

The oral elicited imitation test is another kind of test that has been used to measure implicit knowledge (e.g., R. Ellis, 2005; Erlam, 2006). In R. Ellis (2005) and Erlam (2006), participants were first asked to indicate their beliefs about a statement they heard, and then to repeat the statement in correct English. Of note is that half of the statements were grammatical and the other half ungrammatical. The assumptions behind the design of the test were firstly that requiring participants to respond to the meaning of a sentence reduced the likelihood of focusing on linguistic form and thus reduced the likelihood of their access to explicit knowledge, and secondly that their spontaneous correction of the ungrammatical sentences would be a powerful indication of their internalized language knowledge (Erlam, 2006). This approach is superior to the free language production to the extent that it is easier for researchers to target specific structures they want to elicit, and is therefore easier to rate. However, the validity of the oral elicited imitation test cannot be assured unless its "reconstructive" nature is guaranteed. If the design of the task allows for simple rote imitation, the elicited response may not be an accurate reflection of the participants' internalized implicit linguistic knowledge. Another problem similar to free language production is that participants can still use their explicit knowledge to monitor their output when they are asked to repeat the sentence.

Compared with free language production and oral elicited imitation, which are productive tests by nature, a GJT only involves receptive language use, which can be considered as a promising method in measuring implicit knowledge. A large body of literature has emerged on the validity and reliability of GJTs as measures of L2 knowledge (e.g., R. Ellis, 2004; Gass, 1994; Han, 2000; Loewen, 2009; Mandell, 1999; Murphy, 1997). The key construct validity issue concerns what a GJT measures—explicit knowledge, implicit knowledge or a mix of both. It seems that what type of knowledge a GJT measures depends on certain test conditions—presence or absence of time pressure, presentation modality of test materials (auditory or written) and task stimulus (grammatical or ungrammatical). The untimed GJT with written presentation of ungrammatical sentences tends to encourage learners to access explicit L2 knowledge whereas the timed written GJT (grammatical items) appears more conducive to measuring implicit L2 knowledge (e.g., R. Ellis, 2005; Han & Ellis, 1998; Loewen, 2009). Auditory GJT (e.g., Murphy, 1997; Whong-Barr & Schwartz, 2002) has been considered to be conducive to measuring implicit knowledge.

One issue with auditory GJT is its difficulty in interpreting the test results because it requires both grammaticality competence and phonological decoding ability to successfully complete the task. As Jiang (2004) notes, for EFL learners who learn English as a foreign language, their poor performance on an auditory GJT can be attributed more to their poor phonological decoding ability than to their insufficient grammatical competence. Johnson (1992) provided evidence that when the same test items were presented to the same participants in a written format, significant improvement were made. The phonological decoding factor can indeed be eliminated when test stimuli are in writing; however, it creates the problem of potential involvement of explicit knowledge (Jiang, 2004). When a sentence is visually presented, test takers can go back and forth to check the grammar. This is why a time constraint is extremely important in such a task. However, even a timed GJT task cannot completely eliminate the potential involvement of explicit knowledge. To summarize, both an auditory GJT and a timed written GJT have their limitations as measures of implicit L2 knowledge, either inviting an interwoven phonological decoding factor, or allowing for the involvement of explicit knowledge.

The above review of the methodological issues in measuring implicit and explicit L2 knowledge does not reveal a rosy picture for finding pure and exhaustive measures of the two types of knowledge. DeKeyser (2003) commented that "at this point researchers have to content themselves with eliciting knowledge

under conditions that are more or less conducive to the retrieval of implicit and explicit knowledge" (p. 319). Though the behavioral measures are unlikely to provide pure and exhaustive measures of implicit and explicit knowledge, researchers did find that these tests are more likely to separately measure them. The construct validity of these tests as separate measures of implicit and explicit knowledge is in need of empirical validation.

R. Ellis (2005, 2006b)

In order to address the construct validity of the tests used to measure implicit and explicit knowledge in SLA studies, R. Ellis (2005) proposed a set of seven criteria to distinguish implicit and explicit knowledge. Based on these criteria, he further developed a battery of five tests—an oral imitation test, an oral narrative test, a timed GJT, an untimed GJT, and a metalinguistic knowledge test, with the first three hypothesized to measure implicit knowledge and the latter two explicit knowledge. All the five tests examined 17 English grammatical structures. The tests were administered to 20 native speakers of English and 91 learners of English at a university in New Zealand. R. Ellis (2005) then used an exploratory factor analysis (EFA) to analyze the scores for the five tests and drew a two-factor solution with the scores for the first three tests loading on the first factor and the scores for ungrammatical items in untimed GJT and metalinguistic knowledge test loading on the second factor, which were interpreted as implicit and explicit knowledge, respectively. Isemonger (2007) criticized R. Ellis (2005) for using an EFA on the grounds that the design of the study had a prior model and a verificational agenda, and suggested that a confirmatory factor analysis (CFA) be used. R. Ellis and Loewen (2007) then used CFA and confirmed the construct solution of the implicit/explicit model predicted by the design of the study.

In a follow-up study, R. Ellis (2006b) administered four of the five tests to a larger population of L2 learners (N = 220) in order to investigate how learning difficulty of the grammatical structures is related to the two knowledge types and how the two knowledge types are related to L2 proficiency. In order to determine whether some grammatical structures were easy in terms of explicit knowledge but difficult in terms of implicit knowledge and vice versa, R. Ellis calculated the difference score between explicit and implicit knowledge for each grammatical structure. Those difference scores, however, were not tested for statistical significance, so it is not clear whether the differences were meaningful or due to random error. This study also attempted to explore the relationship between the two types of knowledge and L2 proficiency, which was operationalized as the four language skills—listening, speaking, reading and writing—as measured by the International English Language Testing System (IELTS). Generally, low to medium positive correlations were found between implicit/explicit knowledge of specific grammatical structures and the IELTS scores (total, listening, speaking, reading and writing), indicating that the development of L2 proficiency entails the development of implicit and explicit knowledge of grammatical structures. However, the development of implicit and explicit knowledge in L2 learners was not the focus of the study.

THE STUDY

The goal of the present study was twofold. First, it was set up to investigate the construct validity of four of the tests originally developed in R. Ellis (2005) on a different population, i.e., Chinese learners of English in an EFL context. Second, it attempted to reveal the developmental trends of implicit and explicit knowledge of grammatical structures in the L2 learners at both the macro level, i.e., the overall development of implicit and explicit knowledge of grammatical structures, and the micro level, i.e., the development of the two types of knowledge of specific grammatical structures. In accordance with the purposes of the study, the following research questions (RQs) were formulated:

1. Do the four tests—the oral imitation test, the timed GJT, the untimed GJT and the metalinguistic test—measure two distinct constructs (i.e., implicit and explicit knowledge)?
2. What are the developmental trends of *overall* implicit and explicit knowledge of grammatical structures in Chinese learners of English?
3. What are the developmental trends of implicit and explicit knowledge of *specific* grammatical structures in Chinese learners of English?

Participants

Participants in this study were native speakers of Chinese studying English as a foreign language in China. They were 88 undergraduate students majoring in English language and literature at a major university in China, ranging in ages from 18 to 22 years. The participants consisted of 30 students in Year One, 30 in Year Two, and 28 in Year Three of a four-year program, among which 26 were male and 65 were female. All participants had finished at least 6 years of junior and senior high school English education. None of the participants had ever any experience of studying in or living in an English-speaking country.

Instruments

This study used four of the five tests developed by R. Ellis' (2005). The oral narrative test was not used in the study due to its difficulty in targeting specific structures. This study also used the same test items as those by R. Ellis (2005). The 17 grammatical structures and their properties were summarized by R. Ellis (2005) in Table 1. His choice of the grammatical structures was motivated by a number of considerations. First, all the selected grammatical structures were problematic to learners, resulting in identifiable errors. Second, they were representative of a full range of structures acquired at both early and late stages in L2 acquisition. Third, they were representative of a broad range of proficiency levels according to a typical teaching syllabus. Finally, they included both morphological and syntactical structures.

Table 1 Target Grammatical Structures (from R. Ellis, 2005, p. 155)

Structure	Example of learner error	Acquisition	Pedagogic introduction	Type
Verb complements	Liao says he wants *buying* a new car.	Early	Lower intermediate	S
Regular past tense	Martin *complete* his assignment yesterday.	Intermediate	Elementary/lower intermediate	M
Question tags	We will leave tomorrow, *isn't it*?	Late	No clear focus at any level	S
Yes/no questions	Did Keikko *completed* her homework?	Intermediate	Elementary/lower intermediate	M
Modal verbs	I must to *brush* my teeth now.	Early	Various levels	M
Unreal conditionals	If he had been richer, she *will* marry him.	Late	Lower intermediate/intermediate	S
Since and *for*	He *has been living* in New Zealand *since* three years.	Intermediate	Lower intermediate	S
Indefinite article	They had *the* very good time at the party.	Late	Elementary	M
Ergative verbs	Between 1990 and 2000 the population of New Zealand *was* increased.	Late	Various levels	S
Possessive -s	Liao is still living in his rich *uncle* house.	Late	Elementary	M
Plural -s	Martin sold a few old *coin* to a shop.	Early	No clear focus at any level	M
Third person -s	Hiroshi *live* with his friend Koji.	Late	Elementary/lower intermediate	M
Relative clauses	The boat that my father bought *it* has sunk.	Late	Intermediate/advanced	S
Embedded questions	Tom wanted to know what *had I done*	Late	Intermediate	S
Dative alternation	The teacher explained *John the answer*.	Late	No clear focus at any level	S
Comparatives	The building is *more bigger* than your house.	Late	Elementary/intermediate	S
Adverb placement.	She writes *very well* English.	Late	Elementary/lower intermediate	S

Note. S = syntactic; M = morphological.

Oral Imitation Test

This test was comprised of a set of 34 belief statements, involving one grammatical sentence and one ungrammatical sentence for each of the 17 target structures. The sentences were aurally presented to participants one at a time by a Windows Media Player on a Dell laptop. The participants were required to first complete an answer sheet indicating whether they agreed with, disagreed with, or were not sure about the content of each statement and then to repeat the sentences orally in correct English. Their responses were audio recorded. The whole process of participants taking this test was monitored by the first author in order to ensure that each participant indicated their belief about each statement before they repeated the sentence.

The responses were analyzed by identifying obligatory occasions for the use of target structures.

Test takers' failure to imitate a sentence or to reproduce it in a form that lacked an obligatory occasion for the target structure was coded as avoidance. Each imitated sentence was allocated a score of either 1 (the target structure was correctly supplied) or 0 (the target structure was either avoided or attempted but incorrectly supplied). Scores were expressed as percentage correct.

Timed Grammaticality Judgment Test

This was a computer-delivered test consisting of 68 sentences (4 per grammatical structure: 2 grammatical and 2 ungrammatical). The sentences were different from those in the oral imitation test and were presented in written form on a computer screen. Participants were required to indicate whether each sentence was grammatical or ungrammatical by pressing response buttons within a fixed time limit of 6.5 seconds[①].

The participants were asked to judge the grammaticality of each sentence as quickly as they could via their feel of the sentence. Participants were told that both their response time and judgment for each sentence would be recorded by the computer and used as data for the study. They were also told that if their response time for a sentence exceeds the time limit, that question would be left unanswered, resulting in being scored as incorrect. Each item was scored dichotomously as correct/incorrect. A percentage accuracy score was calculated.

Untimed Grammaticality Judgment Test

This was a paper-delivered test with the same items as in the timed GJT. The participants were required to (a) indicate whether each sentence was grammatical or ungrammatical, (b) indicate the degree of certainty of their judgment as 0%, 20%, 40%, 50%, 60%, 80% or 100%, and (c) to self-report whether they used *rule* or *feel* for judging each sentence. Thus, this test provided three separate measures: a percentage judgment accuracy score, a percentage certainty score, and a percentage score based on the participants' reported use of rule in judging each item.

Metalinguistic Knowledge Test

This was an untimed paper-based test in two parts. The first part presented the participants with 17 ungrammatical sentences (one sentence for each target structure) and required them to select the rule that best explained each error out of four choices provided. The second part had two sections. In the first section, the participants were asked to read a short text and then to find examples of 19 specific grammatical features from the text. In the second section, they were asked to identify the named grammatical parts in a set of sentences. A total percentage accuracy score was calculated.

Procedure

Each participant came to a private office and completed the four tests individually in the following order: the oral imitation test, the timed GJT, the untimed GJT and the metalinguistic knowledge test. The directions for each test were first delivered in English either aurally in the oral imitation test or in written form in the other three tests, and then explained by the researcher in Chinese. The battery of four tests was completed in a single session, and the administration time for the complete set of tests ranged approximately from one hour to one and a half hours.

Analysis

Two phases of statistical analysis were conducted. In Phase 1, the purpose was to examine the construct validity of the four tests in response to the first RQ. The internal consistency reliability of the four psychometric tests was calculated using Cronbach's Alpha. Descriptive statistics (means and standard deviations) for the four tests were also calculated. In order to validate the predictions about the type of knowledge each test measured and the argument about the number of factors that could be extracted, two confirmatory factor analyses using maximum likelihood estimation were performed using AMOS 5.0 (Arbuckle, 2004). The first analysis tested a one-factor construct solution. The second analysis tested the two-factor solution that was argued by R. Ellis (2005), i.e., the oral imitation test and the timed GJT would load on an implicit factor, and the untimed GJT (ungrammatical sentences) and the metalinguistic knowledge test would load on an explicit factor. The two models were then compared to find the best fitting model.

The results in Phase 1 suggesting that the two-factor solution for the implicit/explicit model best fit the study's data enabled us proceed to statistical

analysis in Phase 2. To answer RQ 2 concerning the developmental trends of overall implicit and explicit knowledge of grammatical structures, a combined mean score for the oral imitation test and the timed GJT for all 17 grammatical structures was calculated as a measure of implicit knowledge, and a combined mean score for the ungrammatical sentences on the untimed GJT and the metalinguistic knowledge test was calculated as a measure of explicit knowledge. Then, a repeated measures ANOVA with year level (Year 1, Year 2 and Year 3) as a between-subjects factor and type of knowledge (implicit vs. explicit) as a within-subjects factor was conducted in order to examine (a) the effect of year level on the participants' mastery of implicit and explicit knowledge of the grammatical structures, (b) the effect of type of knowledge, and (c) whether there was an interaction[②] between year level and type of knowledge. When significant differences were found with the three year levels, the Tukey's post hoc test was used to locate the significant differences among pairs of the three levels.

In order to answer RQ 3 concerning the developmental trends of implicit and explicit knowledge of each specific structure, a combined mean score for the oral imitation test and the timed GJT for each of the 17 structures was calculated, and a combined mean score for the untimed GJT (ungrammatical) and the metalinguistic knowledge test for each structure was calculated. Then, a repeated measures ANOVA with year level as a between-subjects factor and type of knowledge as a within-subjects factor for each grammatical structures was conducted to explore (a) the effect of year level on participants' mastery of each of the grammatical structures, (b) the effect of type of knowledge for each structure, (c) whether there was an interaction between year level and type of knowledge for each structure. As 17 independent hypotheses were tested on the set of data, the Bonferroni correction method was used to maintain the familywise error rate (Abdi, 2007). Thus, each individual hypothesis was tested at a significance level of 0.003 (i.e., 0.05/17).

RESULTS

Table 2 presents means and standard deviations of the percentage accuracy scores for each of the four tests completed by the 88 Chinese participants as well as reliability measures for each test. The results demonstrate that the participants achieved the highest mean score on the untimed GJT test (M = 0.91), followed by the timed GJT (M = 0.82), then the metalinguistic knowledge test (M = 0.67), and finally the oral imitation test (M = 0.53). The dispersion of the distribution of the scores for each test was reflected by the standard deviations, ranging from .05 for the untimed GJT to .11 for the metalinguistic knowledge test. The Cronbach's alpha values varied between 0.70 for the metalinguistic knowledge test and 0.57 for the oral imitation test, which will be discussed in detail in the discussion section.

Table 2 Reliability and Descriptive Statistics for the Four Tests

Test	Items	Participants	M	SD	Reliability
Oral Imitation	34	88	0.53	0.10	$\alpha = 0.57$
Timed GJT	68	88	0.82	0.07	$\alpha = 0.62$
Untimed GJT	68	88	0.91	0.05	$\alpha = 0.62$
Metalinguistic knowledge	40	88	0.67	0.11	$\alpha = 0.70$

Construct validity of the four tests

Table 3 presents a summary of model fit statistics for both the one-factor and the two-factor solutions, with the χ^2 statistics (testing the goodness of fit), comparative fit index (CFI) and root mean square error of approximation (RMSEA) presented. The higher the p value associated with χ^2, the closer the fit between the hypothesized model and the perfect fit, and a significant value represents a statistically unlikely event (i.e., a misfitting model) (Byrne, 2001). As can be seen from this table, the one-factor model yielded a χ^2 value of 13.125, with 2 degrees of freedom and a probability of 0.001, thereby suggesting that the fit of the data to this hypothesized model is inadequate. The two-factor solution for the implicit/explicit model had a χ^2 value

of 1.602 with a probability of 0.206, indicating that this model fits the data. Values for the CFI range from zero to 1.00, and a value>0.95 is considered representative of a well-fitting model (2001). Here we can see that the CFI value for the one-factor model was far lower than the threshold of 0.95, thereby indicating that the model was inadequate. The CFI value for the implicit/explicit model, however, was 0.984, thus adding more evidence that the implicit/explicit model represented an adequate fit to the data.

The RMSEA takes into account the error of approximation in the population and this index depends on the degrees of freedom and sample size. It has been advised that values less than 0.05 indicate good fit, values as high as 0.08 represent reasonable errors of approximation in the population, values ranging from 0.08 to 0.10 indicate mediocre fit, and those greater than 0.10 indicate poor fit (Byrne, 2001). The table shows that the implicit/explicit model yielded a RMSEA value of 0.083 while the one-factor model produced a value of 0.254. It is found by Chen *et al*. (2008) from an empirical evaluation that when the degrees of freedom are small and sample size is not large ($n \leqslant 100$), a 0.05 cutoff value of the RMSEA tends to over-reject true population models. Because the sample size was relatively small and the degree of freedom of the implicit/explicit model was only 1, it is arguable that the RMSEA value as low as 0.083 indicated an adequate fit for this model. Thus, all the above indices indicate that the one-factor model does not fit while the implicit/explicit two-factor model adequately fits the data in this study.

Table 3 Summary of the model of fit for the two models

Model	χ^2	df	CFI	RMSEA
One-factor model	13.235** (p = 0.001)	2	0.608	0.254
Implicit/explicit model	1.602 (p = 0.206)	1	0.984	0.083

Note: CFI = comparative fit index; RMSEA = root mean square error of approximation.
** $p<0.00$.

Development of overall implicit and explicit knowledge of grammatical structures

Since the hypothesized implicit/explicit two-factor construct model was confirmed by the confirmatory factor analysis, a combined mean score for the oral imitation test and the timed GJT was calculated as a measure of implicit knowledge, and a combined mean score for the ungrammatical sentences on the untimed GJT and the metalinguistic knowledge test was calculated as a measure of explicit knowledge. The reliability for the measures of implicit and explicit knowledge is presented in Table 4. As can be seen, the Cronbach's alpha value for implicit knowledge was 0.70 and 0.74 for explicit knowledge, indicating that the measures for the two types of knowledge were sufficiently reliable. The means and standard deviations of the percentage accuracy scores in terms of implicit and explicit knowledge across year levels are presented in Table 5.

Table 4 Reliability for Implicit and Explicit Measures

Measure	Items	Participants	Reliability
Implicit	102	88	α = 0.70
Explicit	74	88	α = 0.74

Table 5 Descriptives for Implicit and Explicit Measures

Year	N	Implicit knowledge		Explicit knowledge	
		M	SD	M	SD
1	30	0.64	0.07	0.77	0.09
2	30	0.68	0.06	0.79	0.07
3	28	0.71	0.07	0.83	0.06
Total	88	0.68	0.07	0.80	0.08

The accuracy scores were analyzed in a 3 x 2 repeated measures ANOVA with year level (Year 1, Year 2, and Year 3) as a between-subjects factor and type of knowledge (implicit vs. explicit) as a within-subjects factor. Levene's test for Equality of Variances showed that the variances of population samples were equal, $p = 0.709$ for implicit knowledge, and $p = 0.066$ for explicit knowledge. The ANOVA yielded a significant main effect for year level, $F(2, 85) = 7.621$, $p = 0.001$, $\eta_p^2 = 0.152$[③], and a significant main effect for type of knowledge, $F(1, 85) = 194.328$, $p < 0.001$, $\eta_p^2 = 0.696$, while no significant interaction was found between year level and type of knowledge, $F(2, 85) = 0.338$, $p = 0.714$, $\eta_p^2 = 0.008$. These results suggest that participants at each year level achieved significantly higher scores on explicit knowledge measures than on implicit knowledge measures. In addition, a post hoc analysis using Tukey's test (alpha = 0.05) further revealed that participants in Year 3 achieved significantly higher scores in both implicit and explicit knowledge measures than participants in Year 1, and no significant difference was found between scores achieved by participants in Years 2 and 1 or between Years 2 and 3 in terms of both implicit and explicit knowledge. Figure 1, demonstrating the mean scores for the two type of knowledge at each year level, delineates the developmental trends of implicit and explicit knowledge across year levels.

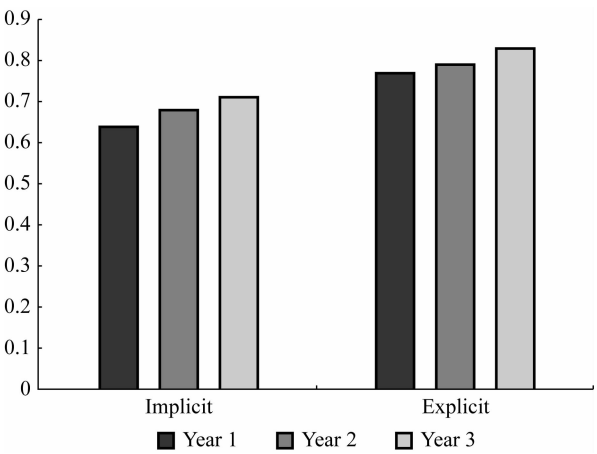

Figure 1 Mean scores for implicit and explicit knowledge across year levels

Development of implicit and explicit knowledge of specific grammatical structures

Development of implicit and explicit knowledge of specific grammatical structures across year levels

A repeated measures ANOVA for each grammatical structure with year level (Year 1, Year 2, and Year 3) as a between-subjects factor and type of knowledge (implicit vs. explicit) as a within-subjects factor was conducted. The equality of variance assumption was not met for each structure. Statisticians note that if group sizes are equal or approximately equal (largest/smallest < 1.5), the F statistics is robust for unequal variance (Steven, 2007). In this study, the group sizes were almost equal with two groups having 30 participants in each and the other group 28 participants, thus the repeated measures ANOVA could proceed accurately.

Results from the 17 repeated measures ANOVAs (alpha = 0.003, i.e., 0.05/17) showed that there was no interaction between year level and type of knowledge for all the structures, meaning that the effect of year level was the same for both implicit and explicit knowledge of each structure. Significant year level effect was found only for the structure verb 3rd person -s ($F(2, 85) = 7.128$, $p = 0.001$, $\eta_p^2 = 0.144$). Post-hoc pairwise comparisons analyses using Tukey's test revealed that participants in Year 3 achieved significantly higher scores than those in Years 1 and 2, and no significant difference was found between Years 1 and 2.

Difference scores between explicit and implicit knowledge of specific grammatical structures

Table 6 presents the means and standard deviations for the 17 grammatical structures in both implicit and explicit knowledge and the differences between explicit and implicit mean scores. The structures are listed in the descending order of the mean difference

between explicit and implicit knowledge. The mean scores for both implicit and explicit knowledge of each structure are presented graphically in Figure 2.

The result from the repeated measures ANOVA for each grammatical structure showed that participants scored significantly higher in explicit measures than in implicit measures on 12 grammatical structures—unreal conditionals, comparative, indefinite article, embedded questions, plural, question tags, verb 3rd person -s, regular past tense -ed, since/for, possessive -s, relative clauses, and verb complement. Conversely, significantly higher score in implicit measures than in explicit measures was found only on one grammatical structure modals. For the other four structures, no significant difference was found in scores between implicit and explicit measures.

Table 6 Means and Standard Deviations of Explicit and Implicit scores for 17 Grammatical Structures and Differences between Explicit and Implicit Mean Scores (n = 88)

Grammatical structures	Explicit		Implicit		Difference
	M	SD	M	SD	
Indefinite article	0.91	0.18	0.63	0.17	0.29**
Comparative	0.87	0.20	0.59	0.20	0.28**
Embedded question	0.93	0.14	0.65	0.16	0.28**
Question tags	0.93	0.14	0.67	0.14	0.26**
Plural -s	0.92	0.18	0.66	0.18	0.25**
Regular past tense -ed	0.96	0.14	0.72	0.16	0.25**
Since/for	0.92	0.14	0.76	0.19	0.16**
Relative clauses	0.95	0.14	0.80	0.17	0.16**
Unreal conditionals	0.73	0.22	0.58	0.15	0.15**
Third person -s	0.84	0.17	0.70	0.18	0.15**
Possessive -s	0.88	0.21	0.79	0.16	0.09**
Verb complement	0.95	0.15	0.86	0.17	0.09**
Yes/no question	0.90	0.17	0.83	0.16	0.07
Ergative verbs	0.83	0.23	0.79	0.15	0.05
Adverb placement	0.65	0.29	0.68	0.15	−0.03
Dative alternation	0.62	0.36	0.67	0.16	−0.05
Modals	0.85	0.19	0.93	0.11	−0.08**

** $p<0.003$

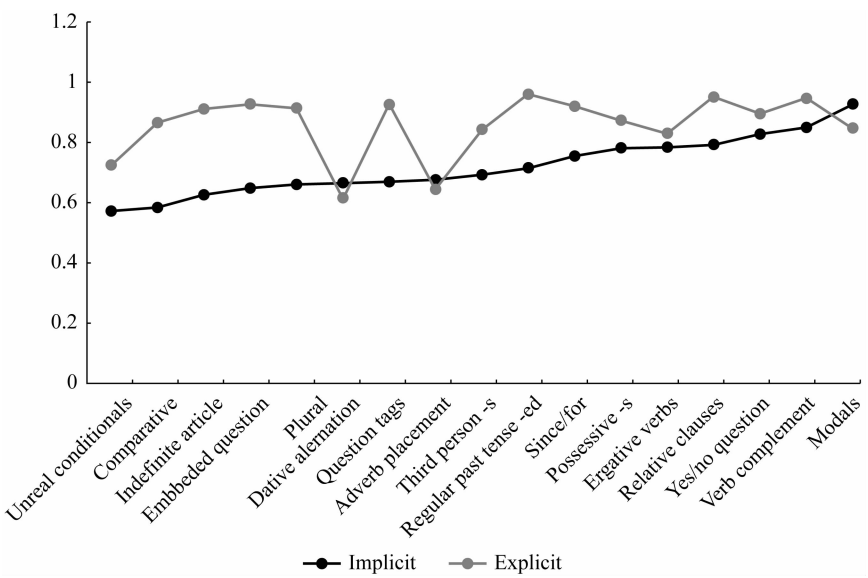

Figure 2 Implicit and explicit mean scores for the seventeen grammatical structures

DISCUSSION

Before discussing the results corresponding to each research question, we would like to first discuss the reliability of the four tests used in this study. This study employed the same four tests in R. Ellis (2005, 2006b)—the oral imitation test, the timed GJT, the untimed GJT, and the metalinguistic knowledge test. Compared with the reliability measures in R. Ellis' studies (see Table 7), this study obtained overall lower Cronbach's alpha values. It should be noted that this study used the identical testing items for the four tests as those in R. Ellis (2005, 2006b), therefore, the item quality of the tests were assured. In addition, the administrative procedures of the tests in this study were very similar to those in R. Ellis' studies with the exception of the timed GJT. The overall lower Cronbach's values for the four tests could probably be explained by the fact that the participants in this study were less various in terms of English proficiency as compared with those in R. Ellis'. As Cronbach's alpha is a function of the population variance of the test scores (Bachman, 2004, p. 163; Brown, 2005, p. 179), the larger the variance of the test scores, the higher the value of Cronbach's alpha. The lower alpha values in this study could be justified by the more uniform population variance of the test scores that can be seen from the much smaller standard deviations.

Table 7 Reliability and Descriptive Statistics for the Four Tests in R. Ellis (2005, 2006b)

Test	R. Ellis (2005, p. 159)				R. Ellis (2006b, p. 449)			
	N	M	SD	Reliability	N	M	SD	Reliability
Oral Imitation	91	.51	.17	.88	228	.50	.19	.88
Timed GJT	91	.54	.12	.81	227	.56	.12	.96
Untimed GJT	91	.82	.11	.83	225	.81	.13	.83
Metalinguistic knowledge	91	.53	.21	.90	228	.55	.16	.79

Comparing the means and standard deviations for the four tests in this study with those in R. Ellis' (2005, 2006b), it can be seen that participants in this study achieved higher mean scores with smaller standard deviations on each test. This indicates that participants in this study, who were English majors at a prominent university of China, probably had higher English proficiency than those in R. Ellis' studies. The smaller standard deviations on the four tests indicates that the English proficiency of participants in this study were more uniform than that in R. Ellis' studies.

One notable difference between this and R. Ellis' study is the substantially higher mean score for the timed GJT. One reason for this difference may be the different time constraints set for the timed GJTs in the two studies. In R. Ellis (2005), a time limit was established for each sentence based on 10 native speakers' average response time for each sentence in a pilot study, and an additional 20% of the time taken for each sentence was added to allow for the slower processing speed of L2 learners, and the time limit for each sentence allowed for L2 learners ranged from 1.8 to 6.24 seconds. In this study, however, a time limit of 6.5 seconds was set for each sentence due to the difficulty of finding a reasonable number of native English speakers in the local area in China to set the baseline of the time limit for each sentence. The relatively less strict time limit in this study was compensated by asking participants to respond as quickly as they could, where they were told that both their response time and accuracy for each sentence would be recorded and used as data for the study. It is possible that the less rigorous time limit in the timed GJT in the study might have allowed time for participants to access some of their explicit knowledge on some sentences, even though they were asked to respond as quickly as possible.

To investigate the first research question, i.e., whether the four tests measure two distinct constructs (i.e., implicit and explicit knowledge), two confirmatory factor analyses were performed. The results suggest that while the one-factor model does not fit the data, the two-factor solution for the implicit/explicit model fits. Specifically, it was confirmed that the oral imitation test and the timed GJT measure the construct of implicit knowledge while the untimed

GJT ungrammatical items and the metalinguistic knowledge test measure the construct of explicit knowledge. The results of this study confirmed the construct validity of the four tests developed by R. Ellis (2005) as relatively separate measures of implicit and explicit knowledge on a different population of participants.

The second research question concerned the developmental trends of overall implicit and explicit knowledge of grammatical structures in Chinese learners of English. The results suggest that there were significant improvement in terms of both implicit and explicit knowledge across year levels. Specifically, participants in Year 3 had significantly higher levels of both implicit and explicit knowledge of grammatical structures than those in Year 1, and no significant difference was found between Years 1 and 2 or between Years 2 and 3. The cross-sectional design of this study enables us to tentatively conclude that with the increase of L2 proficiency, adult L2 learners who learned English as a foreign language in China usually possess a higher level of both implicit and explicit knowledge of grammatical structures.

The study also found that participants achieved significantly higher scores in the explicit knowledge measures than in the implicit knowledge measures. This indicates that EFL learners generally possess a significantly higher level of explicit knowledge of grammatical structures than that of implicit knowledge. One caveat is that the higher explicit knowledge scores might partly be an artifact of the test scoring procedures. The oral imitation test (meaning implicit knowledge) in this study was scored using obligatory occasion analysis whereas the ungrammatical GJTs and the metalinguistic knowledge test (measuring explicit knowledge) involve selective response scoring, which might have contributed to the higher explicit knowledge scores to a slight extent.

Caution should also be taken when we attempt to generalize the findings from this particular population of participants to other populations even in EFL contexts. Of note is that the participants in this study were L2 learners who learned the English language as their major in an undergraduate degree program in China. In other words, one of their career goals was to teach the English language. Given that, it was not surprising that the participants improved their explicit knowledge along with higher year levels. English majors were more likely to be motivated to learn the metalinguistic knowledge of English than L2 learners in other majors who learned English for other purposes rather than teaching the language. Therefore, the findings may not be generalizable to non-English majors in this foreign context; neither should they be generalized to situations where learners acquire English as their second language in an English-speaking country. Further studies are clearly needed to reveal the developmental trend of implicit and explicit knowledge in other populations, such as non-English majors at college level in foreign contexts, EFL students at high school level, and English learners in ESL contexts.

The third research question concerned the developmental trends of implicit and explicit knowledge of specific grammatical structures. The results suggest that implicit and explicit knowledge of only one grammatical structure verb 3rd person -s improved significantly across year levels. Students in Year 3 did significantly better on verb 3rd person -s than those in Year 1 and Year 2. For the other sixteen structures in terms of both implicit and explicit knowledge, no significant improvement was found across year levels. This finding is significant in that it documented the development of these 17 grammatical structures across proficiency levels in adult English learners in the context of China. It is difficult to say why and what exactly contributed to the significant improvement of implicit and explicit knowledge of this particular structure verb 3rd person -s across year levels instead of the others. Further study of a longitudinal rather than cross-sectional design on the development of these grammatical structures in terms of both implicit and explicit knowledge would be more revealing about the process of acquisition.

Comparing the scores for implicit and explicit knowledge on each grammatical structure allows us to see whether there were any grammatical structures that are easy in terms of one type of knowledge (implicit or explicit) but difficult in terms of the other type (R. Ellis, 2006b). This study found that 12 grammatical structures (i.e., unreal conditionals, comparative, indefinite article, embedded questions, plural, question tags, verb 3rd person -s, regular past tense -ed, since/for, possessive -s, relative clauses, and verb complement) were significantly easier in terms of explicit knowledge than they were in terms

of implicit knowledge. This implies that L2 learners in this EFL context may very well know the rules for these structures but have difficulty in using them accurately spontaneously. Conversely, when it comes to the structure of modals, with the testing point being that modals should be followed by the base form of the main verb, it was significantly easier in terms of implicit knowledge. For yes/no question, ergative verbs, dative alternation and adverb placement, no significant difference was found in scores between implicit and explicit measures. These results of this study with the support of inferential statistical analysis further confirmed the previous finding by R. Ellis (2006b) that the level of difficulty of grammatical structures varied according to whether it was implicit or explicit knowledge of the structures involved.

CONCLUSION

With the aim of describing the development of implicit and explicit knowledge of grammatical structures in Chinese learners of English, this study used the four tests developed by R. Ellis (2005) that were intended to distinguish the two types of knowledge. The construct validity of the four tests was first investigated, and results confirmed that the four tests tended to measure two distinct constructs—implicit and explicit knowledge. Based on this result the development of the two types of knowledge of grammatical structures in Chinese adult L2 learners of English was explored. It was found that learners with higher proficiency level had higher level of both implicit and explicit knowledge of grammatical structures, and overall, the learners possessed higher level of explicit knowledge than implicit knowledge. On the development of the seventeen individual grammatical structures, only verb 3rd person -s improved significantly in terms of the two types of knowledge across year levels. In addition, it was found that the majority of the structures were much easier to learn as explicit knowledge than implicit knowledge for adult L2 learners in the foreign context.

In terms of the interface between implicit and explicit knowledge of a second language, this study by the nature of design cannot directly address the question about the role of explicit knowledge on the development of implicit knowledge. However, this study is contributive in two aspects. First, it provided strong support for having available relatively separate measures of the two types of knowledge. It is only when sensitive measures of the two knowledge types are well established that we can empirically address the theoretical positions on the interface between them. Second, this study provided a description of the development of both types of knowledge of a wide variety of grammatical structures in a group of adult EFL learners. The finding that both types of knowledge increased with year level provides empirical evidence for the argument that explicit knowledge does not transform into implicit knowledge in the sense of metamorphosis (DeKeyser, 2009; Paradis, 1994) because explicit knowledge remains available or even increases after implicit knowledge is acquired.

This study also has limitations. The validity of the findings depends heavily on the four tests used to measure the two types of knowledge. As has been discussed earlier, while the four tests were not pure measures of implicit or explicit knowledge, they are relatively separate measures of the two knowledge types. A timed GJT with more rigorous time limit and a self-paced word by word reading task (Jiang, 2004) could be more sensitive for measuring implicit knowledge. Future work on developing more sensitive measures of the two knowledge types, especially of implicit knowledge, is clearly needed.

ACKNOWLEDGEMENTS

This study was supported by the National Social Science Foundation of China (10BYY035) and National Research Centre for FL Education Foundation (ZGWYJYJJ2010A30). This study was presented at the 2010 American Association of Applied Linguistics conference. We are grateful to Rod Ellis and two other anonymous *Language Learning* reviewers for helpful comments on an earlier version. We are also grateful to Michelle Dove for proofreading this manuscript. Any remaining errors and omissions are of our own.

NOTES

1 This procedure was different from R. Ellis (2005), who established a time limit for each sentence on the basis of 10 native speakers' average response time for each sentence in a pilot study, to which an additional 20% of the time taken for each sentence was added to allow for the slower processing speed of L2 learners. In his study, the time limit for each sentence ranged from 1.8 to 6.24 seconds.

2 An interaction means that "the effect one independent variable has on the dependent variable is not the same for all levels of the other independent variable" (Steven, 2007, p. 125).

3 η_p^2 is a measure of effect size. Cohen (1977) characterizes $\eta_p^2 = 0.01$ as corresponding to a small effect size, $\eta_p^2 = 0.06$ to a medium effect size, and $\eta_p^2 = 0.14$ to a large effect size (see Steven, 2007, p. 118).

REFERENCES

Abdi, H. (2007). Bonferroni and Šidák corrections for multiple comparisons. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 103-107). Thousand Oaks: Sage.

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

Arbuckle, J. (2004). *AMOS 5.0*. Chicago: Small Waters Corporation.

Bachman, L. F. (2004). *Statistical analysis for language assessment*. Cambridge: Cambridge University Press.

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

Bialystok, E. (1979). Explicit and implicit judgements of L2 grammaticality. *Language Learning*, *29*, 81-103.

Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment* (New ed.). New York: McGraw-Hill College.

Butler, Y. (2002). Second language learners' theories on the use of English articles. *Studies in Second Language Acquisition*, *24*, 451-480.

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

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in Structural Equation Models. *Sociological Methods & Research*, *36*(*4*), 462-494.

DeKeyser, R. M. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition*, *17*, 379-410.

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

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. Doughty & M. Long (Eds.), *Handbook of second language learning* (pp. 313-348). Oxford: Blackwell.

DeKeyser, R. M. (2009). Cognitive-pscychological processes in second language learning. In M. H. Long & C. Doughty (Eds.), *Handbook of second language teaching* (pp. 119-138). Oxford: Blackwell.

Ellis, N. C. (1994). Implicit and explicit processes in language acquisition: An introduction. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 1-32). San Diego: Academic Press.

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

Ellis, N. C. (2007). Implicit and explicit knowledge about language. In J. Cenoz & N. H. Hornberger (Eds.), *Encyclopedia of language and education, second edition, volume 6: Knowledge about language* (pp. 119-132). Berlin: Springer.

Ellis, R. (1984). *Classroom second language development*. Oxford: Pergamon.

Ellis, R. (1993). Second language acquisition and the structural syllabus. *TESOL Quarterly*, *27*, 91-113.

Ellis, R. (1994a). *The study of second language acquisition*. Oxford: Oxford University Press.

Ellis, R. (1994b). A theory of instructed second language acquisition. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 79-114). San Diego: Academic Press.

Ellis, R. (1997). *SLA research and language teaching*. Oxford: Oxford University Press.

Ellis, R. (2002). Does Form-Focused Instruction affect the acquisition of implicit knowledge? *Studies in Second Language Acquisition*, *24*(*2*), 223-236.

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, *54*(*2*), 227-275.

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

Ellis, R. (2006a). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly*, *40*(*1*), 83-107.

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

Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford: Oxford University Press.

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*(*1*), 119-126.

Ellis, R., Loewen, S., & Elder, C. (2009). *Implicit and explicit knowledge in second language learning, testing and teaching*. Bristol: Multilingual Matters.

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

Gass, S. (1994). The reliability of second-language grammaticality judgments. In E. Tarone, S. Gass & A. Cohen (Eds.), *Research methodology in second language acquisition* (pp. 303-322). Hillsdale: Erlbaum.

Green, P. S., & Hecht, K. (1992). Implicit and explicit grammar: An empirical study. *Applied Linguistics*, *13* (*2*), 168-184.

Han, Y. (2000). Grammaticality judgment tests: How reliable and valid are they? *Applied Language Learning*, *11*(*1*), 177-204.

Han, Y., & Ellis, R. (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research*, *2*(*1*), 1-23.

Hu, G. (2002). Psychological constraints on the utility of metalinguistic knowledge in second langauge production. *Studies in Second Language Acquisition*, *24* (*3*), 347-386.

Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning: Introduction. *Studies in Second Language Acquisition*, *27*(*2*), 129-140.

Hulstijn, J. H., & De Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA Review*, *11*, 97-113.

Isemonger, I. M. (2007). Operational definitions of explicit and implicit knowledge: Response to R. Ellis (2005) and Some Recommendations for Future Research in This Area. *Studies in Second Language Acquisition*, *29*(*1*), 101-118.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, *25*, 603-634.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning*, *57*(*1*), 1-33.

Johnson, J. S. (1992). Critical period effects in second language acquisition: The effect of written versus auditory materials on the assessment of grammatical competence. *Language Learning*, *42*, 217-248.

Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon.

Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.

Krashen, S. (1994). The input hypothesis and its rivals. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 45-77). San Diego: Academic Press.

Loewen, S. (2009). Grammaticality Judgment Tests and the mesuresment of implicit and explicit L2 knowledge. 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. 94-112). Bristol: Multilingual Matters.

Mandell, P. B. (1999). On the reliability of grammaticality judgement tests in second language acquisition research. *Second Language Research*, *15*(*1*), 73-99.

Masny, D. (1987). The role of language and cognition in second language metalinguistic awareness. In J. Lantolf & A. Labarca (Eds.), *Research in second language learning: Focus on the classroom* (pp. 59-73). Norwood: Ablex.

Murphy, V. (1997). The effect of modality on a grammaticality judgement task. *Second Language Research*, *13*(*1*), 34-65.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417-528.

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and SLA. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 393-419). San Diego: Academic Press.

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, *11*, 11-26.

Searle, J. R. (2000). Consciousness. *Annual Review of Neuroscience*, *23*, 557-578.

Sorace, A. (1985). Metalinguistic knowledge and language use in acquisition-poor environments. *Applied Linguistics*, *6*(*3*), 239-254.

Stevens, J. P. (2007). *Intermediate statistics: A modern approach*. New York: Lawrence Erlbaum Associates.

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition*, *27*(*2*), 173-204.

Whong-Barr, M., & Schwartz, B. D. (2002). Morphological and syntactic transfer in child L2 acquisition of the English dative alternation. *Studies in Second Language Acquisition*, *24*, 579-616.