Elicited Imitation as a Measure of L2 English Learners' Interlanguage Representation of Relative Clauses

Bee Eng Wong
(bee@fbmk.upm.edu.my)
Universiti Putra Malaysia, Malaysia

Pauline Hwa Ling Teo
(teohl@mail.tarc.edu.my)
Tunku Abdul Rahman College, Malaysia

Abstract

This study examines the use of the Elicited Imitation Test (EIT) (Ellis, 2004) to measure second language learners' underlying knowledge of restrictive relative clauses which will reflect their interlanguage representation of this property. Two groups of learners, L1 Malay and L1 Chinese speakers, were involved in the study. To test the suitability of the EIT as a test of such implicit knowledge, the Malay and Chinese speakers were brought through the EIT and a grammaticality judgement test (GJT), an established task often used in second language acquisition studies. The results from both tasks were compared and correlated. The results showed that learners were generally better at judging and imitating grammatical items in both tests and a positive correlation is indeed found between both tests for grammatical items. Scores obtained from both groups of learners were also comparable. However, it was found that learners were less determinate in their judgement and production of ungrammatical items. In general, they were less proficient in their ability to imitate or judge and recast the ungrammatical items correctly. The results indicate that the L1 Malay and L1 Chinese learners of L2 English have interlanguage representations that differ from native speakers' underlying representations of the said property. The results also indicate that although there is no correlation between the two tests in measuring the ungrammatical items, it is proposed that it was actually a factor related to the time allotted for participants to respond to items in the GJT rather than the EIT itself that had resulted in this outcome.

1   Introduction

Second language acquisition (SLA) research has traditionally been interested in finding out *what learners know* (language competence) and *how they know* what they know (language acquisition). However, researchers have also begun to take interest in *how to find out* what learners know (language testing). Equal interest in both language acquisition, in particular, SLA, as well as language testing has begun to develop.

Thus, an area that is pertinent in SLA studies is the methodology applied to measure second language (L2) knowledge. Chomsky puts it simply that what we know may not necessarily be what we do and consequently, the goal of many research studies conducted throughout the years is to determine how the information from the latter can help us learn more about the former (in Schutze, 1996). In order to investigate this phenomenon, the Grammaticality Judgement Test

(GJT) has been the main instrument used to measure SLA (Mandell, 1999). Four reasons to use the GJT are its ability to measure response towards sentence types which seldom appear in spontaneous conversation, its ability to use information which can only be obtained within normal language use, its ability to detect reliably slips, and its ability to reduce the extent of communicative and representational functions of language skill (Cook & Newson, 1996). In recent years, the Elicited Imitation Test (EIT) has come to be used more often although the technique itself is not new. Some researchers, such as Gallimore and Tharp (1981), Hamayan, Saegert and Larudee (1977), and Munnich, Flynn and Martohardjono (1994), claim that EIT may be a good instrument to reflect learners' interlanguage (IL) representation (cited in Erlam, 2006).

Many studies on the acquisition of relative clauses conducted in the generative framework employ the GJT (see Hawkins & Chan, 1997; Wong, 2002). The GJT is said to be reliable and can be used for both first language (L1) and L2 studies (Mandell, 1999). However, over the years, other tests have been introduced and used in language research. One of the tests which have been used in the testing of language proficiency and linguistic competence is the EIT. The EIT has been used by researchers such as Slobin and Welsh (1968, cited in Ferguson & Slobin, 1973), among others, with young or even preliterate and barely literate children as their subjects, to test L1 acquisition.

Studies by Ellis (e.g. 2004) and Erlam (2006), however, have shown that the EIT can be, and is used with other older subjects, specifically adults, to measure language competence. Thus, this study will look at the suitability and appropriateness of the EIT as an instrument to tap the language competence of L2 learners. We are interested to see the suitability of the EIT to study the IL representations of L2 learners.

Studies such as those conducted by Gass (1994), Loew (1996), Davies and Kaplan (1998), and Mandell (1999) have investigated the reliability and validity of the GJT. Gass (1994) conducted a study on the reliability of the GJT on judgments of English relative clauses by Chinese, Korean and Japanese ESL learners and found that the subjects' overall performance was significantly reliable. Loew (1996), who looked at the relationship of the GJT and oral production tasks, also concluded that GJTs reflect behavioural patterns of L2 development after the results from his study showed a significant relationship between the tests. Ellis (2004), in his research on measurement of L2 explicit knowledge, made a distinction between the Timed GJT and the Untimed GJT. According to him, timed-tasks such as the Timed GJT were used to elicit implicit knowledge while explicit knowledge could be elicited using the Untimed GJT. In this study, the Timed GJT is used to reflect learners' IL representations or implicit knowledge.

2   Theoretical background

2.1   Implicit and explicit knowledge

The terms implicit knowledge and explicit knowledge are two of the key constructs in this study. As such, they are elaborated here. Both implicit and explicit knowledge of language have been discussed and debated over the years and many definitions for these two types of knowledge have been proposed to distinguish them, albeit using different terms. Chomsky (1986, 1991, cited in Smith, 2004) claims that there is an (I)nternalized and an (E)xternalized language. According to Chomsky, I-language is a property within the mind. It is concerned with knowledge of the language that a speaker possesses rather than the language a speaker produces. On the other hand, E-language relates to grammar generated from data collected from speakers (see also Cook & Newson, 2007).

While many definitions have been proposed to show the differences between implicit and explicit knowledge, the general view by Chomsky (1986, 1991) and the specific view given by Ellis (2004) will be adopted in this study. Ellis claims that there are seven principles that distinguish implicit from explicit knowledge. He further divides the seven principles into two dimensions, the representation dimension and the processing dimension. The representation dimension includes awareness, the type of knowledge and the systematicity and certainty of L2 knowledge, while the processing dimension covers accessibility of the knowledge, the use of the L2 knowledge, self-

report and learnability. In this study, focus is given to the representation dimension to distinguish the two types of knowledge.

A major difference between implicit and explicit knowledge is the area of awareness where unconscious awareness is involved in implicit knowledge while conscious awareness is involved in explicit knowledge. Jong (2005) asserts that grammatical knowledge is explicit when it involves conscious awareness of the formal properties of the target language. According to Karmiloff-Smith (1979, in Ellis, 2004), demonstration of implicit knowledge can be seen when a learner is able to recognise if a sentence is ungrammatical. This is termed as "epilinguistic" behaviour. In contrast, "metalinguistic" behaviour, or conscious awareness of the reasons a sentence is ungrammatical indicates the existence of explicit knowledge. For instance, if an L2 learner with implicit knowledge of the following sentence and is presented with a sentence like "Sam explained Erin the procedure," s(he) will be aware that that the sentence is ungrammatical and would probably even be able to point out the error. However, the learner may not be conscious of the grammar rule that has been violated. On the other hand, an L2 learner with explicit knowledge of dative alternation will understand and be able to verbalise the reason for the ungrammaticality of the sentence; that is, the verb "explain" cannot be followed by an indirect object (*Erin*) without an infinitive "to" when the direct object (*the procedure*) is after the indirect object.

Meanwhile, Eichenbaum (1997, cited in Ellis, 2004) states that explicit knowledge is declarative in that it involves knowledge of abstract rules and knowledge of fragments and examples. Implicit knowledge, on the contrary, draws upon procedural knowledge which is very much automated. According to Anderson's (1983, 1985) ACT model, with practice, it is actually possible for learners to move from the declarative stage where they have "knowledge that something is the case" to the procedural stage where they "know how to do something" which ultimately leads to automatisation (Mitchell & Myles, 2004, p. 102).

SLA researchers have argued that the IL representation of L2 learners or basically, their implicit knowledge is highly systematic (Tarone, 1988) when compared to their explicit knowledge, which is frequently less systematic, in that it is "imprecise, inaccurate and inconsistent" (in Sorace, 1985). Besides, implicit knowledge is accessible automatically while explicit knowledge on the other hand involves controlled processing.

The tasks involved in measuring implicit and explicit knowledge also differ considerably. Ellis (2004) reports that Bialystok (1979) concluded many years ago that requiring learners to verbalise specific grammar rules is a manifestation of their explicit knowledge. In addition, Ellis (2002) maintains that the best measure of implicit knowledge is spontaneous production tasks.

2.2   *Interlanguage representation*

Another notion that is important to the study is interlanguage representation. Interlanguage refers to the linguistic system of L2 learners at a specific point in time, and to the series of interlocking L2 systems of the development of the language. This L2 system has its own set of rules that are independent of both the L1 and the L2. In other words, the abstract and complex properties that are manifested in the learner's IL grammar are different from native speakers' but yet not determined by their L2 input (White, 1998).

Focus on IL representation has long been given to L2 initial stage as well as during the development of L2. However, although the source of the L2 grammar is important, the importance in the nature of the IL representation cannot be denied. Therefore, the question to be answered is "What constitutes L2 learners' underlying knowledge or interlanguage representation?" According to White (1998, 2003), it is crucial that IL grammar is seen as an entity on its own and which are constrained by the principles and parameters of Universal Grammar (UG). "The learner's system is worthy of a study on its own right, not just as a degenerate form of the target system" (Bley-Vroman, 1983, p. 4). White (1998, 2003) states that L2 learner's IL grammar, though not similar to native speakers' mental representation, may arrive at a grammar that explains for L2 input and the underlying knowledge which they possess, and that knowledge and the use of knowledge do

not always coincide. This is supported by Hawkins and Chan (1997) who claim that L2 learners employ a different method to analyse grammar of English.

Two studies that have been conducted on the acquisition of the English relative clauses by speakers of other first languages are directly related to this study. One study investigates the acquisition of the English relative clauses by L1 Malay speakers (Wong, 2002) and the other investigates the acquisition of the English relative clauses by L1 Chinese speakers (Hawkins & Chan, 1997).

Wong (2002) conducted a study on the acquisition of English relative clauses by L1 Malay speakers. In particular, it addressed the issue of parameter-resetting and the critical period for L2 acquisition. She concentrated on relative clauses with overt operators, and excluded relative clauses with complementizer *that* or null operators. A group of 21 L1 Malay-speaking advanced learners of English and a control group of 13 adult native speakers of English participated in the study. Wong used a GJT to collect data from her subjects. Twenty-three (23) constructions out of a total of 81 items in the GJT were related directly to the target structures of the study, that is, the English relative clause. The GJT was administered using a bimodal technique in which subjects listen to a recording of the items that were paced at a nine-second interval as well as read the items at the same time. Subjects were given four options in judging the items: definitely acceptable, probably acceptable, probably unacceptable and definitely unacceptable. Scoring was based on the subjects' accurate judgement of the items. The results of the study found that Malay speakers in general were not quite proficient in recognizing and in accepting subjacency violations on non-subject positions. Evidence showed that in accordance with construction of relative clauses in Malay which do not permit extraction from non-subject positions, Malay speakers do not possess movement in their L2 grammar representations of relative clauses. Therefore, although they do not have difficulty with the principles of UG, Malay speakers seem to have difficulty with the feature structure of the complementizers in English where they are unable to reset parameters that are associated with functional categories.

Wong (2002) also reported that the subjects of the study did not make use of movement in their mental representation for English but instead they were found to employ a null resumptive pronoun strategy. Therefore, although the subjects can construct mental representations of English relative clauses, they do so using the no-movement resumptive pronoun strategy. The researcher concludes that L2 learners have interlanguage grammars that are UG-constrained but are unable to access some of the features in the functional categories in the L2 and, as a result, manifest properties that are different from both the L1 and L2.

In the other study, Hawkins and Chan (1997) investigated *wh*-operator movement in Chinese and English restrictive relative clauses (RRC). They investigated the availability of UG in SLA. A total of 292 subjects were involved in the study. Subjects comprised of Chinese learners of English as a second language (ESL), French learners of English as a second language and native speakers of English. The native French subjects were controls for the reliability and validity of the test instrument while the native English speakers acted as controls for the reliability of the tokens used in the test. The subjects were divided into seven groups. There were three groups for the Chinese and French learners of elementary, intermediate and advanced proficiency levels. The native speaker group was the control group. A GJT was used in the study as a data collection instrument. There were 101 items in the test with 59 items related directly to the target structure. The items were made up of four different types of sentences which included sentences with grammatical and ungrammatical use of operators and complementizers, ungrammatical sentences with resumptive pronouns, sentences that violated the subjacency condition and ungrammatical sentences involving null subjects in embedded relative clauses. The test was carried out bi-modally and subjects were instructed to follow the pace of the tape while doing the test. The pace was set at nine seconds between each sentence. The subjects read the sentence on the paper given to them while listening to the tape that was played. They then had to write down their judgement for the sentence. They were provided with four options: definitely correct, probably correct, probably incorrect and definitely incorrect.

From this study, Hawkins and Chan (1997) found that the Chinese subjects' accuracy in their intuition about the English RRC increased with exposure to English. However, their accuracy and ability to correct subjacency violations regress with increasing proficiency. In contrast, Hawkins and Chan (1997) found that the French-speaking subjects in the study that were comparable in age and proficiency to the Chinese subjects possess accuracy and ability to correct subjacency violations that increases with proficiency. This result indicated that subjects were constrained by the feature specifications of functional categories in their L1.

This result is due to the fact that Chinese does not allow *wh*-operator movement while French does. It was found that the Chinese subjects did not accept non-UG-licensed constructions despite the possibility of null subject in gap constructions in their native Chinese. This result suggested that their mental representations seemed to be UG constrained as they correctly reject non-UG-licensed [*wh*-phrase …gap] constructions such as *The girl cried [when e lost her way]*.

Hawkins and Chan (1997) hypothesize that the Chinese subjects' mental representations of gap constructions in English RRC is different from those of native speakers; that is, they do not have mental representation of *wh*-operator movement. Instead of a trace, the Chinese subjects' grammars demonstrate the null pronominal *pro*. The *wh*-operators are regarded as topics that are generated *in situ* in CP which do not involve movement regardless of whether the *wh*-operators are overt or null. They attribute this to the "Failed Functional Feature Hypothesis" (FFFH). According to this hypothesis, functional features such as C (Complementizer), Agr (Agreement) and D (Determiner) which are located in the UG lexicon are subject to a critical period, available sometime in childhood. The hypothesis states that beyond this critical period, the functional categories and features not instanted in the L1 of the learners become inaccessible. The findings also show that the IL representations of the learners diverge from the target grammar despite apparent native-like performance.

Another study that is of interest is by Mandell (1999), which investigated the reliability of the GJT to assess learners' linguistic competence on Spanish V(erb)-movement. As a measure for comparison, a dehydrated sentence test (DST)[1] was also included as a research tool. It is the assumption of this study that some form of UG is still accessible among adult L2 learners, in this case, L1 speakers of Spanish. Mandell (1999) predicted that there would be a significant correlation between the data from the GJT and the DST relating to surface-level syntactic properties determined by the V-movement[2] parameter. He also predicted that learners will initially take on the [-raise] value corresponding to their L1 when writing or reading in their L2 which include those properties. However, when the parameters are reset to correspond to the [+raise] value in their L2, there would be a change in the way the learners read or write which reflects a change in parametric value.

A total of 91 adult Spanish university students representing beginning (40), intermediate (24), and advanced (19) L2 learner levels participated in the study. In both tests, the GJT and the DST, were three critical item types: a) Grammatical items whose adverbials were placed between lexical verbs and object NPs (e.g. José explains rapidly the lesson – José rapidly explains the lesson), b) Grammatical *yes/no* questions in which the subjects and the lexical verb were inverted (e.g. reads Juan the newspaper – Does Juan read the newspaper?), and c) Ungrammatical items with fronted thematic *wh*-phrases (without subject-lexical verb inversion) (e.g. what Manolo watches – What is Manolo watching?). In the GJT, subjects had to indicate if the item on the test was "possible" or "impossible." They were also instructed to make the necessary changes to the items that were marked "impossible." Subjects were given 10 seconds to read each item, judge and indicate the changes for the "impossible" items. In the DST, subjects were told to construct an acceptable sentence by combining the given elements. Subjects had 15 seconds to construct each sentence. The scoring method used for both the tests was based on a binary scale. In the GJT, a correct judgement that corresponded to the [+raise] parameter value was scored "1" while incorrect judgements that corresponded to the [-raise] parameter value was scored "0."

Results from both tests of the study showed a significant correlation between the scores obtained for the GJT and the DST on the three critical item types in relation to V-movement in Spanish. A consistent and cross-sectional correlation appeared for all three levels of learners indicating

that the GJT is a reliable measure of learners' L2 competence with regard to V-movement in Spanish. Mandell (1999) stresses that the results of his study indirectly supports the findings from Gass (1994) about the reliability of the GJT which were based upon ESL learners' judgements on the relative clause in English. Mandell (1999) concludes that his study supports the reliability of the GJT as there is a definite relationship between both the GJT and the DST, and although L2 learners' grammaticality judgements are indeterminate, they are indeed consistent.

Since the main purpose of the study is to determine the use of the EIT to measure learners' implicit knowledge of the RRC, three studies and a review on EIT are of relevance here. They are studies conducted by Hamayan et al. (1977), Ellis (2004), Erlam (2006), and a review by Vinther (2002).

In their study, Hamayan et al. (1977) used the EIT to test performance for elicited imitation for L2 learners across ages. Three groups (1, 2 and 3) of participants were involved in the study. Group 1 were native speakers of Arabic of about eight years of age and have had exposure to English since kindergarten. Group 2 consists of participants who were about 11 years of age while the participants in Group 3 were college adults with very low proficiency in English although they had studied English for an average of 7.6 years. A total of seven language structures were selected for the EIT and these were divided into two types: A and B. Type A consisted of 3 "easy" structures, namely the conjunction, complement and number, while Type B consisted of 4 "difficult" structures which are the negative *wh*-question, relative clause, verb auxiliary and adjective. Four sentences for each of the seven structures, two grammatical and two ungrammatical, were devised. In total, the participants were presented with 28 sentences in random order. Scores were given based on four different classifications: accurate repetitions, errors of deviation (sentences with syntax errors), inadequate responses (sentences with distorted meaning) and normalization (ungrammatical sentences that are corrected). It was noted that on the whole, participants were better at correcting Type A sentences than Type B sentences. Scores on correction of ungrammatical sentences were similar to the scores on repetition of grammatical sentences for the Group 1 structures. Further, accuracy on correction of ungrammatical Type A sentences increases with age (and presumably with proficiency) while the accuracy rates remained constant for Type B sentences for all three groups. This indicates that the Type B structures were problematic for the learners even with increased proficiency.

In her review on the elicited imitation technique, Vinther's (2002) aim is to provide a better understanding of the technique for language testing. The focus is on the decoding processes involved in this testing technique and the different aspects it can have that might influence the results of the study. The most challenging aspect of the EIT is to determine whether participants really understand the stimulus that is presented or they simply imitate a chain of sounds without really understanding the meaning. In referring to Eisenstein, Bailey and Madden (1982, p. 391), she states that if learners understand the sentence, they will be able to repeat the sentence without any difficulty. Therefore, it is her assumption that a learner's imitating capacity reflects the learner's ability to produce sentences spontaneously. She draws this assumption from the findings of Smith (1970), Naiman (1974), and Corrigan and Di Paul (1982). In her investigation of other studies that used the EIT, Vinther (2002) observed that most studies that used the EIT as a means of data collection had sentences that were isolated or with minimal contextual clues. Although it has been suggested by Bloom, Hood and Lightbown (1974) that the EIT is not effective due to lack of contextual support, Vinther (2002) argues that this very same format allows students to rely exclusively on the linguistic knowledge of the structure and semantics. She quotes Flynn (1986) who explicitly advised against providing a context. In her study, Flynn (1986) kept the contextual clues in the sentences to a minimum in order to ensure that the participants' reaction showed their underlying knowledge of specific grammatical structures. Vinther (2002) believes that more research has to be done with regard to contextual clues in EIT. She proposed to draw a line between communicative context and redundant context. It is her view that redundant context could be counterproductive as participants may be strained by the tasks. On the other hand, she feels that it was not unreasonable to include communicative context such as the reason the sentence is being uttered

and the function it has in communicative events. This might help participants to forget their role as a subject in an experiment.

In his study, Ellis (2004) investigates valid measures of L2 implicit and explicit knowledge. Ellis employed five different types of tests for this purpose. The tests included an oral imitation test, a narrative test, a timed GJT, an untimed GJT, and a Metalinguistic Knowledge Test. The first three tests were designed with the purpose to measure implicit knowledge and the remaining two to test explicit knowledge. A total of 111 participants were involved in this study, comprising of 20 natives and 91 L2 learners of English. Seventeen (17) English grammatical structures were set as items to be measured in the tests. Results for the tests were obtained from a comparison of performance by the native speakers and the L2 learners of English. The assumption that native speakers would possess a higher level of implicit knowledge was proven with the results showing that the natives performed better than the L2 English language learners in the oral imitation, narrative and timed GJT tests, which were designed to measure implicit knowledge. This, according to Ellis, also confirms the validity of the imitation and timed GJT tests as reliable measures of implicit knowledge. However, Ellis (2004) postulates that the fact that the native speakers have a higher implicit knowledge than the L2 English language learners does not necessarily imply that the L2 learners of English have a higher explicit knowledge. Results from this study show that both the native speakers and the L2 learners of English language performed similarly in the tests of explicit knowledge, namely the untimed GJT and the Metalinguistics Knowledge Test. In addition, the study also reports that test results showed a greater systematicity in learners' performance in the tests for implicit knowledge while the untimed GJT seemed to relate strongly to the Metalinguistics Knowledge Test, confirming it as a measure of explicit knowledge. The timed GJT, as a measure of implicit knowledge, is said to be related to the age learners start to acquire the L2, whereas the untimed GJT, which measures explicit knowledge, is associated with the years of formal instruction that learners have been exposed to. In this study, Ellis (2004) drew on seven hypotheses. These hypotheses state that since the test of explicit knowledge compels learners to use "rule" instead of "feel" and it is not time pressured but focuses on form, it will result in responses that are not consistent and less certain. Further, according to the hypotheses, the test of explicit knowledge employs the use of metalinguistic knowledge and would result in scores which are related more strongly to the years of classroom instruction that learners are exposed to rather than the age at which they started learning the L2. Out of these seven hypotheses in Ellis' (2004) study, only one was proven otherwise where learners were more certain of their answer when accessing explicit knowledge. He attributed this to the learners' low proficiency level and their lack of confidence in their implicit knowledge of the grammatical structures which are mainly known to be acquired late such as the question tags and hypothetical conditionals. He concluded that the overall construct validity of the tests was supported empirically by the scores obtained from the analyses conducted.

In another study, Erlam (2006) investigated the EIT itself as a measure of implicit knowledge. Her EIT focuses on meaning of the utterance. She conducted the study with 115 participants, of whom 20 were native speakers and 95 were SLA learners. In this study, the EIT was designed to test participants' knowledge of specific language structures. The structures concerned were verb complements, 3rd person -*s*, plural -*s*, possessive -*s*, regular past tense, yes/no questions, comparatives, unreal conditional, modal verbs, ergatives, indefinite article, embedded questions, adverb placement, question tags, the adverb *since*, the preposition *for*, relative clauses and dative alternation. Participants were informed that they were responding to a Beliefs questionnaire and were presented with statements, which they could give their views on. Their first task was to point out whether the statements that they have heard were true, not true or they were unsure. There were grammatical and ungrammatical statements. Subsequently, they were instructed to repeat the statement that they were presented with in correct English. The purpose of including ungrammatical statements was to create a task which was reconstructive rather than merely a test or rote repetition. Apart from the EIT, participants were required to take an online narrative task and sit for the speaking and listening components of the International English Language Testing System (IELTS). The EIT results were then compared with the results from the oral narrative task and the speaking and listening components of the IELTS to find out if there is any correlation between the EIT and

the oral narrative task as well as between the EIT and the IELTS. A positive relationship was found between participants' ability to repeat grammatical statements and their ability to correct ungrammatical statements. Native speaking participants scored 97% and 91% respectively for repeating grammatical statements correctly and correcting ungrammatical statements. The same was found of the L2 learners as 61% repeated grammatical statements correctly while 35% corrected ungrammatical statements. Basically, the results from this study showed that participants were not repeating literally as the participants corrected the ungrammatical statements as well. This indicates that the EIT, which centres on a meaning-focus condition during execution, is reconstructive and is possibly accessing implicit language knowledge. In addition, Erlam (2006) found correlation between both the scores obtained from the EIT and the oral narrative task and the IELTS scores. Based on her findings, Erlam (2006) claims that the EIT is a reliable measure of assessing implicit language knowledge. The test is able to assess implicit language knowledge, as it first requires participants to understand the meaning of the statements before repeating the grammatical ones and correcting the ungrammatical ones.

Finally, in an attempt to look at the relationship between EIT and grammaticality judgements, Munnich et al. (1994) conducted a study investigating the English spoken by advanced Japanese learners. Results from the study showed that there was a significant correspondence between the two tests. Munnich et al. (1994) concluded that "there was a need to have multiple investigations of the same linguistic phenomena with distinct methodologies."

3   The study

3.1   Purpose of the study

The purpose of this study is to look at an alternative testing technique apart from the GJT in SLA studies. Since the GJT is an established task and seems to have been more widely used, especially in the SLA studies conducted locally (in Malaysia) using the generative framework, the study set out to explore the possibility of using the EIT to measure L2 learners' underlying knowledge of the English RRC as a manifestation of their IL representation of the said property. The focus in this study is to determine if the EIT is a suitable testing technique to elicit and tap the underlying mental representation of L2 learners. The results obtained from the EIT will be compared with the results obtained from the GJT to determine if the results correlate. In the process, the study also investigates the IL mental representation of L1 Chinese and L1 Malay learners of L2 English with regard to the formation of English RRC. Results from both the EIT and GJT are compared to determine the suitability of the EIT as a measure of the IL representations of English RRC among Malaysian ESL learners. Therefore, this study investigates the areas of what the learners know (language competence) and how to find out what they know (language testing) among L1 Chinese and L1 Malay learners of L2 English in Malaysia as these two groups of learners represent the major ethnic groups in the country.

To address the purpose and objectives of the study, the following questions were formulated to examine the underlying knowledge of English restrictive relative clauses of L1 Chinese and L1 Malay speakers and to determine the suitability of the EIT in comparison to the GJT as a measure of linguistic competence as well as the L2 learners' interlanguage representation of English restrictive relative clauses: a) To what extent does the learners' ability to repeat grammatical structures correctly in the EIT correlate with their ability to judge grammatical structures correctly in the GJT? b) To what extent does the learners' ability to correct and recast ungrammatical structures in the EIT correlate with their ability to judge ungrammatical structures in the GJT?

3.2   Methodology

As reviewed earlier, the GJT has been shown to be a reliable instrument for measuring L2 competence. In this study, we investigate the use of another instrument, the EIT. The research method for this study combined the use of the two different tests. In this, it follows Loew's (1996)

study to investigate the relationship of the GJT and oral production tasks. The GJT is timed at 10 seconds between consequent items, following Mandell (1999). Although the GJT and the EIT differ in their modes, both the timed GJT and the EIT measure implicit knowledge (see Ellis, 2004). As such, there should be a correlation between the scores obtained from the two tasks on the same property. A factor that seems to be crucial is the time interval allocated between item, since a difference is made between Timed GJT and Untimed GJT (Ellis, 2004) intervals. The OPT functions as a tool in the selection process. Only learners who performed at advanced level were chosen for the study. Results from the EIT were analysed to determine if the participants were able to repeat grammatical sentences as well as correct and recast ungrammatical sentences as they repeated them. Results from the two different tests were also analysed and compared. By comparing the results from the two tests, the correlation between the results was established. It was predicted that there would be a positive relationship between participants' ability to repeat sentences in the EIT and their ability to judge the grammaticality of the sentences in the GJT.

The EIT that was used in this study replicated the design, administration and scoring system of that used in Erlam's (2006) study. The statements in the EIT that were used in this study, however, were not a replication of the statements used in her study. This is due to the different target structures being investigated by both studies. Primarily, the EIT in this study was designed as a reconstructive test to measure participants' underlying knowledge, which would reflect their interlanguage representation of the English relative clauses. According to Erlam (2006), the features of a reconstructive test of underlying implicit knowledge of language must include the following:

1) A test which calls for attention on the meaning (not form) of sentences
2) The presence of a delay between the presentation time of the stimuli and the time of repetition
3) A test that is performed with time-pressure (speeded) instead of being self-paced

In addition, the EIT was designed in such a way that participants would believe that they were answering a questionnaire. The reason for this was to focus participants' attention on meaning instead of form. Therefore, the sentences were constructed as statements that may draw differing opinions from participants. Statements were based on current issues in the Malaysian context so that participants were familiar with them. Participants first listened to the stimulus (statement) being played on a tape player. They were then probed on whether they felt that the statement was true or false or they are unsure about the stimulus. The reason for this was to present a delay between the presentation time of the stimulus and the time of repetition in order to correspond to one of the features of reconstructive tests, which would allow learners to focus on meaning rather form. There were statements which were not self-explanatory, for example, "The man who he has a daughter is Mr. Lim." Therefore, pictures relating to certain ambiguous statements were also presented simultaneously with the presentation of some of the statements to aid understanding. Subsequently, after stating whether they felt the statement was true, false or they were unsure of it, participants were asked to repeat the statement that was presented to them. Each respondent's production was recorded. The length of each statement ranged from 13 to 16 syllables with a mean length of 14.5 (see Appendix for examples of sentences used in the EIT).

As the purpose of this study was to determine learners' underlying knowledge of English RRC, the only structure that was being tested in the EIT is the English RRC including those with *that* complementizer. The relative clause was chosen as the target structure as it is a common problematic syntactic property among L2 learners of English. The NPAH (Comrie & Keenan, 1979) has shown the differences of this property in different languages. Due to these differences and the complexity of the structure, it is often a problematic construction for L2 learners.

The items in the GJT and EIT were constructed based on the classification of the types and sub-types of constructions for the target structure in Wong (2002, pp. 222–231). The items are of the following types and the details of each are given in the Appendix.

Grammatical Items
- ***Wh*-restrictive RC**: Extraction from Upper Clause (Short Movement)
- ***Wh*-restrictive RC**: Extraction from Embedded Clause (Long Movement)

- **RC with *that* Complementizer**: Extraction from Upper Clause (Short Movement)
- **RC with *that* Complementizer**: Extraction from Embedded Clause (Long Movement)
- ***Wh*-restrictive RC**: Null Specifier

Ungrammatical Items

- ***Wh*-restrictive RC**: Resumptive Pronouns
- **RC with *that* Complementizer**: Resumptive Pronouns
- ***Wh*-restrictive RC**: Extraction *wh*-Island
- ***Wh*-restrictive RC**: Extraction from Complex Noun (Determiner) Phrase
- ***Wh*-restrictive RC**: Violation of Spec-Head Agreement

Two items were formulated for each relative clause type. The total number of items is 72 (36 grammatical and 36 ungrammatical). Fifty L2 learners of English were selected as participants for the main study after meeting the criterion in the OPT. Participants took the EIT after sitting for the OPT. The administration of the EIT was modelled after Erlam's (2006) study.

As the present study, to a certain degree, was a replication Erlam's (2006) study, the scoring method that was used in her study was adopted and adapted. To begin with, the repeated statements were divided based on whether the participants created an obligatory context for the use of the target structure. The statements where no obligatory context was created were classified as "avoidance" and were later graded as incorrect repetitions in the final analysis of the data. The statements which were repeated with an obligatory context were then identified as to whether they were accurately repeated. Accurate repetitions were scored as correct while inaccurate repetitions were scored as incorrect. The statements that were repeated with lexical errors or repetitions with synonyms were also scored as correct as long as the target structure was produced correctly. The ungrammatical statements which were corrected were scored as correct. The grammatical statements which were correct were given a score of "1" while statements which were incorrect were rated as "0." Similarly, for the ungrammatical statements which were corrected, a score of "1" was given while "0" was given to statements that were incorrect, i.e. repeated with the grammatical error. The statements that were repeated without the obligatory context or avoidance were also rated as "0." A clearer representation of the scoring method is shown in Table 1 below.

	Obligatory Context		**Avoidance**
	Correct (repeated accurately)	*Incorrect (repeated inaccurately)*	
Grammatical Statement	1	0	0
Ungrammatical Statement	Corrected*	Not corrected**	
	1	0	0

* The statement is repeated correctly, with changes made to the ungrammatical portion.
** The statement is not corrected; it is repeated exactly as the original sentence, with the ungrammatical portion still remaining.

Table 1: Scoring method for EIT

The other instrument, the GJT, also comprised 72 items. For each of the *wh*-relative clauses listed for the EIT, two statements were formed for the GJT. The structure and length of the statements in the GJT were similar to that of the statements in the EIT. Similar to the EIT, there were both grammatical and ungrammatical statements in the GJT. In fact, the sentences in both tests were of the same types. The only difference was the vocabulary items used (see Appendix 1 for item types used in both tests). Constructing grammatical and ungrammatical statements in the EIT was important in order to shift the participants' focus onto meaning of the statements. On the whole, the ungrammatical statements in both tests would help us in determining if the participants had native speaker competence; that is, whether their underlying representation was similar to native speakers. Statements in the GJT were constructed based on a context, as according to Birdsong (1989), statements presented with context will be more meaningful for the participants (cited

in Wong, 1999). Therefore, providing contexts was pragmatically acceptable because it aided understanding and participants would be able to make more accurate judgements. In addition, similar to the EIT, the classification of the types and sub-types of constructions for the target structure in the GJT was based on Wong (2002).

The presentation of the GJT was done bi-modally which means participants were able to listen and read the statements simultaneously and subsequently make their judgements. Murphy (1997) states that a bimodal GJT is more effective than a unimodal GJT as a bi-modal presentation makes use of both ocular and auditory senses. Participants first read and listened to the instruction that was played on the tape player. Then, before they attempted the actual test, the participants went through a practice session with six statements. After the practice session, participants were given a five-minute break before they moved on to the actual test. In the GJT, participants carefully read and made judgements on the statements. Participants were instructed to give their judgement on each statement by choosing from the options of "acceptable," "probably acceptable," "probably unacceptable," or "unacceptable" on each statement.

Participants had about ten seconds to read and listen to the test item being played, and make a grammaticality judgement for that item. The time allocation given for participants to make their judgement was based on a grammaticality test that was carried out by Mandell (1999). The scoring method used for the GJT is based on Wong (2002). It is on a scale of "0" to "3." A correct grammatical judgement, both for the grammatical and ungrammatical statements was given a score of "3." Hence, a grammatical statement that was judged as "acceptable" was rated as "3" while an ungrammatical statement that was judged as "unacceptable" was also rated as "3." Meanwhile, a grammatical statement that was judged as "probably acceptable" would be given a score of "2," "probably unacceptable" would be given a score of "1," and "unacceptable" would be given a score of "0." Conversely, an ungrammatical statement that is judged as "probably unacceptable" would be given a score of "2," "probably acceptable," a score of "1" and "acceptable," a score of "0." This information is summarized Table 2 below.

Grammatical Statement	Score	Ungrammatical Statement	Score
Acceptable	3	Unacceptable	3
Probably Acceptable	2	Probably Unacceptable	2
Probably Unacceptable	1	Probably Acceptable	1
Unacceptable	0	Acceptable	0

Table 2: Scoring method for the GJT

The scores of each set of items are presented as percentages. A correlation coefficient between tests was carried out for the different types of the *wh*-relative clauses and *that*-complementizer relative clauses for both groups of participants.

3.3   Results and discussion

The results showed that the subjects involved in this study were able to construct English RRC but the underlying representation of this knowledge does not necessarily parallel that of native speakers' implicit knowledge of the same property. The subjects who were L1 Chinese and L1 Malay learners obtained mean scores of above 70 percent which is above chance level, but below 90 percent for the grammatical items in both tests but this result does not represent the actual competence of the learners based on a closer analysis of the scores on the ungrammatical items. The learners did not perform well in the ungrammatical items for the EIT and the GJT with average mean scores of below 60 percent in each test. In the EIT, it was found that both groups of learners could only perform well for the grammatical statements (refer to Table 4). According to Erlam (2006), learners who do have implicit knowledge of the target structure would not only be able to repeat grammatical statements in the EIT, but they are also expected to be able to correct the ungrammatical statements and recast them appropriately in the test, which is reconstructive in nature.

A reconstructive test would mean that the recording of items that learners hear for the EIT merely function as stimuli where they are required to use their own IL grammar to recast and correctly repeat the ungrammatical items.

Similarly, the learners were unable to identify convincingly the ungrammaticality of the items in the GJT and thus made judgements that did not approximate native speaker knowledge. In other words, they did not reject all the ungrammatical items in the GJT. This clearly shows that the learners have not fully acquired *wh*-movement in English relative clauses, if at all (refer to Table 4). It was however noted that the L1 Malay learners performed particularly well at extraction of restrictive relative clauses from the subject (active) position compared to the L1 Chinese learners (refer to Table 3). Adopting Wong's (2002) explanation, this is due to the fact that subject extraction is the only relativisation operation that is possible in Malay. On the other hand, Chinese does not allow movement in the formation of relative clauses (Hawkins & Chan, 1997). It is therefore not surprising to find that the L1 Malay learners were better than the L1 Chinese learners at the formation of subject relative clauses in English.

	Chinese		Malay	
	EIT	*GJT*	*EIT*	*GJT*
Extraction of Subjects (active)	**82%**	**67%**	**82%**	**87%**
Extraction of Subjects (passive)	**76%**	**68%**	**80%**	**84%**
Extraction of Direct Object	54.5%	81%	62%	80%
Extraction of Prepositional Object	60%	94%	56%	90%
Extraction from Upper Clause (Cumulative)	70.5%	77.5%	70%	85.3%

Table 3: Summary of mean scores (%) for the extraction of restrictive relative clause from the upper clause of L1 Chinese and L1 Malay learners

As stated earlier, the main concern of this study is to actually determine if the EIT test is a suitable measure of implicit knowledge. To investigate this, the widely used test of underlying knowledge, the GJT, was used as comparison. This is addressed in the research questions. Essentially, the study investigated the correlation between the grammatical and ungrammatical items in the EIT and the GJT. It is important to see if the results from the grammatical and ungrammatical items in the EIT correlate with the results from grammatical and ungrammatical items in the GJT as this could help determine if the EIT is a suitable measure of implicit knowledge.

The Pearson Bivariate Correlation Test was used for this purpose. The Bivariate Correlation Test was performed on the data obtained for the grammatical and ungrammatical items from both tests with the prediction that there would be a correlation as both tests are considered to be measures of implicit knowledge thus testing the underlying knowledge of L2 English learners in this syntactic property. The test showed that although there was correlation between the data for the grammatical items in the EIT and the data for the grammatical items in the GJT, there was no correlation between the data for the ungrammatical items in the EIT and the data for the ungrammatical items in the of the GJT. This result seems to be the converse of Ellis' (2004) findings that the results from the timed GJT that he conducted correlated to the results in the EIT in his study.

A more detailed look at the study revealed that time could be a factor that might have skewed the results. This occurred during the administration of the GJT in the present study which was slightly different from Ellis' (2004) as the time which was given to learners in this study was longer, that is, ten seconds, following Mandell (1999). It now seems that the time allocated might have been too long as compared to the time limit given by Ellis in his study. As ten seconds was allocated to learners to judge items in the GJT in this study, they may have had the opportunity to rely, to some extent, on their explicit knowledge of the English RRC. Ten seconds might have been too long for learners to solely depend on their "feel" or implicit knowledge of English RRC formation as they might have had time to think about their judgements which in turn led to a high degree of awareness.

In this case, it seems that the GJT used in this study would be considered as similar to an untimed task as the time given to the learners most probably allowed them to rely on their explicit knowledge to some extent. It was therefore expected that only the responses for the grammatical items in the EIT would correlate with the grammatical items in the untimed GJT as based on prior research conducted by Bialystok (1979) and Hedgcock (1993, cited in Ellis, 2004) that L2 English learners have different responses to grammatical and ungrammatical sentences. According to Hedgcock (1993), it is not impossible to say that L2 learners rely on different data bases when instructed to judge grammatical and ungrammatical sentences. This is supported by the fact that the ungrammatical items in the EIT in Ellis's (2004) study did not correlate with the ungrammatical items in the GJT. According to Ellis (2004), the grammatical and ungrammatical items seem to measure different constructs. Therefore, it can be concluded that the GJT used in this study is equivalent to an untimed test due to the longer time given to the learners during the administration of the test and consequently, it is expected that only the grammatical items in the GJT would correlate with the grammatical items in the EIT which measures implicit knowledge.

	EIT (Mean in %)		GJT (Mean in %)	
Grammatical Items	*M*	*C*	*M*	*C*
Wh-restrictive RC Extraction from Upper Clause (Short Movement)	70	71	85	77
Wh-restrictive RC Extraction from Embedded Clause (Long Movement)	67	68	73	72
RC with *that* Complementizer Extraction from Upper Clause (Short Movement)	80	82	83	85
RC with *that* Complementizer Extraction from Embedded Clause (Long Movement)	76	75	77	76
Wh-restrictive RC Null Specifier	83	87	76	70
Ungrammatical Items	*M*	*C*	*M*	*C*
Wh-restrictive RC Resumptive Pronouns	64	66	63	63
RC with *that* Complementizer Resumptive Pronouns	75	68	76	70
Wh-restrictive RC Extraction *wh*-Island	55	60	54	58
Wh-restrictive RC Extractin from Complex Noun(Determiner) Phrase	54	58	49	50
Wh-restrictive RC Violation of Spec-Head Agreement	76	66	59	47

Table 4: Summary of mean scores obtained by L1 Chinese and L1 Malay speakers for the EIT and the GJT

Thus, the question whether there would be a correlation between the data for the ungrammatical items in EIT and the data for the ungrammatical items in GJT was not supported. This is explained by Ellis (2004) that the grammatical items and the ungrammatical items in an untimed GJT seem to measure different types of knowledge. The grammatical items are more prone to measure implicit knowledge while the ungrammatical items measure explicit knowledge. This could possibly explain the reason why there is no correlation between the ungrammatical items in the EIT and the GJT. This is because while the ungrammatical items in the GJT measure explicit knowledge, the ungrammatical items in the EIT measure implicit knowledge. In other words, the results seem to indicate that it is not so much that the EIT is not a reliable test to measure implicit knowledge. Rather, we propose that this has to do with the administration of the GJT. In this, we claim that it has to do with the time factor involved in the administration of the GJT.

4 Conclusions

One significant conclusion that can be drawn is that there is a need for proper methodology and testing instruments when investigating the acquisition of grammatical properties in SLA studies. As Douglas (2001, cited in Ellis, 2004) says, many researchers are guilty of not exhibiting the validity and reliability of their testing instruments. Based on the findings of this study, it is clear that minor differences in methodology and procedure can affect the results of a study and such differences should therefore be taken seriously.

According to Ellis (2004), the time factor plays a crucial part in the administration of a test in SLA studies particularly the GJT. This is because the difference in the length of time given governs the type of knowledge being tested. He further elaborates that a test measures implicit knowledge if it is performed under time pressure and it measures explicit knowledge if it is carried out without time pressure. A comparative analysis of the findings in Ellis' (2004) study and the present study revealed that there were differences. The results from both tests in Ellis's (2004) study seem to correlate with each other, which contrasts with the findings in this study as the results (the ungrammatical items) from both the GJT and EIT do not correlate. A closer look at the tests showed that the time allocation assigned for the GJT in both studies differed. The learners in Ellis's (2004) study were given up to 6.24 seconds in the administration of the GJT, while the time assigned for the GJT in the present study is 10 seconds, which follows the design of Mandell's (1999) GJT. The difference in the time determined the difference in the type of knowledge being tested. Due to the longer time limit given to the subjects in the present study, it seems reasonable to say that the knowledge tested by the GJT tends to be explicit in nature. In contrast, the GJT carried out by Ellis (2004) tested the learners' implicit knowledge owing to the shorter time limit given. Although the difference of time allocated is less than four seconds, it is clear that minute or minor changes in methodology like the length of time assigned to a test item can influence the results or outcome of a study.

The findings show that it is not the different modes of the tasks that have yielded the results for the ungrammatical items. If the two tasks are supposed to measure implicit knowledge, then they can be compared. However, the factor that is crucial is the time allocated for learners to respond to each item. Thus, in carrying out such tests for the purpose of measuring learners' IL competence, much thought and care need to be given to the procedure in data collection. However, despite its inconclusive findings for the ungrammatical items, this study has nevertheless contributed to the local SLA literature. Crucially, the findings for the ungrammatical items here also show that further similar studies need to be carried out to see if the same findings or otherwise would emerge.

Notes

[1] The DST is a test "composed of a series of constituents separated by slashes. Subjects will have to combine the elements given and form an acceptable sentence," e.g. Hector / corer / regularmente / tre millas (Mandell, 1999, p. 88).

[2] V(erb) movement involves moving the head of a phrase to the head position of another phrase. It is also known as head-movement where move can only take place in a sequence of steps, rather than all at once (Cook & Newson, 1996).

References

Anderson, J. (1983). *The architecture of cognition*. Cambridge, M.A.: Harvard University.

Anderson, J. (1985). *Cognitive psychology and its implication*. (2nd ed.). New York: Freeman.

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

Birdsong, D. (1989). *Metalinguistic performance and interlinguistic competence*. New York: Springer-Verlag.

Bley-Vroman, R. (1983). The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning, 33,* 1–17.

Bloom, L., Hood, L., & Lightbown, P. (1974). Imitation in language development: If, when, and why. *Cognitive Psychology, 6*(3), 389–420.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin and use*. New York: Praeger.

Chomsky, N. (1991). Some notes on economy of derivation and representation. In R. Freidin (Ed.), *Principles and parameters in comparative grammar* (pp. 417–545). Cambridge, M.A.: MIT Press.

Comrie, B., & Keenan, E. L. (1979). Noun phrase accessibility revisited. *Language, 3,* 649–664.

Cook, V., & Newson, M. (1996). *Chomsky's universal grammar: An introduction*. (2nd ed.). Cambridge: Blackwell.

Cook, V., & Newson, M. (2007). *Chomsky's universal grammar: An introduction*. (3rd ed.). Cambridge: Blackwell.

Corrigan, R., & Di Paul, L. (1982). Measurement of language production. *Applied Psycholinguistics, 3*(3), 223–242.

Davies, W. D., & Kaplan, T. I. (1998). Native speaker vs. L2 learner grammaticality judgements. *Applied Linguistics, 19,* 183–203.

Douglas, D. (2001). Performance consistency in second language acquisition and language testing: A conceptual gap. *Second Language Research, 17*, 442–456.

Eichenbaum, H. (1997). Declarative memory: Insights from cognitive neurobiology. *Annual Review of Psychology, 48,* 547–572.

Eisenstein, M., Bailey, N., & Madden, C. (1982). It takes two: Contrasting tasks and contrasting structures. *TESOL Quarterly*, *16*(3), 381–393.

Ellis, R. (2002). Does form-focused instruction affect the acquisition of implicit knowledge? A review of the research. *Studies in Second Language Acquisition, 24,* 223–236.

Ellis, R. (2004). *Measuring implicit and explicit knowledge of a second language: A psychometric study*. Auckland: Department of Applied Language Studies and Linguistics, The University of Auckland.

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

Ferguson, C. A., & Slobin, D. I. (Eds.) (1973). *Studies of child language development*. New York: Holt, Rinehart and Winston, Inc.

Flynn, S. (1986). Production vs. comprehension: Differences in underlying competences. *Studies in Second Language Acquisition, 8*, 135–164.

Gallimore, R., & Tharp, R. G. (1981). The interpretation of elicited sentence imitation in a standardized context. *Language Learning, 31*(2), 369–392.

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

Hamayan, E., Saegert, J., & Larudee, P. (1977). Elicited imitation in second language learners. *Language and Speech, 20*(1), 86–97.

Hawkins, R., & Chan, Y-H. (1997). The partial availability of Universal Grammar in second language acquisition: The "Failed Functional Feature Hypothesis". *Second Language Research, 13*(3), 187–226.

Hedgcock, J. (1993). Well-formed vs. ill-formed strings in L2 metalingual tasks: Specifying features of grammaticality judgments. *Second Language Research, 9*(1), 1–21.

Jong, N. D. (2005). *Learning second language grammar by listening* (Doctoral dissertation). Netherlands Graduate School of Linguistics, Utrecht, The Netherlands.

Karmiloff-Smith, A. (1979). Micro- and macro-developmental changes in language acquisition and other representation systems. *Cognitive Science, 3*(2), 91–118.

Loew, R. (1996). Grammaticality judgment tasks and second-language development. In G. Alatis, C. Straehle, M. Rankin & B. Gallenberger (Eds.), *Georgetown University round table on languages and linguistics* (pp. 126–139). Washington, DC: Georgetown University Press.

Mandell, P. B. (1999). On the reliability of Grammaticality Judgement Tests in second language acquisition research. *Second Language Research, 15*(1), 73–99.

Mitchell, R., & Myles, F. (2004). *Second language learning theories* (2nd ed.). London: Hodder Education.

Munnich, E., Flynn, S., & Martohardjono, G. (1994). Elicited imitation and grammaticality judgment tasks: What they measure and how they relate to each other. In E. Tarone, S. M. Gass & A. Cohen, (Eds.), *Research methodology in second language acquisition* (pp. 227–234). Hillsdale, N.J.: Lawrence Erlbaum.

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

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

Schutze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgements and linguistic methodology*. Chicago: University of Chicago Press.

Slobin, D. I., & Welsh, C. A. (1968). Elicited imitation as a research tool in developmental psycholinguistics. *Working Paper, 10*. Reprinted in C. A. Ferguson & D. I. Slobin (Eds.), (1973), *Studies of child language development* (pp. 485–497). New York: Holt, Rinehart and Winston, Inc.

Smith, C. S. (1970). An experimental approach to children's linguistic competence. In J. R. Hayes (Ed.), *Cognition and development of language*. Reprinted in C. A. Ferguson & D. I. Slobin (Eds.), (1973), *Studies of child language development* (pp. 497–521). New York: Holt, Rinehart and Winston, Inc.

Smith, N. (2004). *Chomsky, ideas and ideals* (2nd ed.). Cambridge: Cambridge University Press.

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

Tarone, E. (1988). *Variation in interlanguage*. London: Edward Arnold.

Vinther, T. (2002). Elicited imitation: A brief overview. *International Journal of Applied Linguistics, 12*(1), 54–73.

White, L. (1998). *Universal grammar in second language acquisition: The nature of interlanguage representation*. Canada: McGill University.

White, L. (2003). *Second language acquisition and universal grammar*. Cambridge: Cambridge University Press.

Wong, B. E. (1999). Acquisition of wh-movement in English questions and relative clauses by speakers of Malay (Doctoral Thesis). University of Essex, UK.

Wong, B. E. (2002) Acquisition of English relative clauses by Malay speakers. *Journal of Pan-Pacific Association of Applied Linguistics, 6*(1), 61–77.

Appendix 1

Sample of items used in the EIT (Similar item types and items were used in the GJT)

Grammatical Stimuli
I. Extraction from the upper clause – Short Movement with *wh*-operator
Extraction of subjects (active)
The girl [who [t scored 17 A's for SPM]] is Nur Amalina.
Extraction of subjects (passive)
The event [which [t [was attended by John]]] was the F1 race.
Extraction of direct object
The car [which [Proton first created t]] is Gen 2.
Extraction of prepositional object
The tourism fair [which [we can get good deals from t]] is the MATTA fair.
II. Extraction from embedded clause – Long Movement with *wh*-operator
Extraction of subjects (active)
The movie [which [you think [t [t bore the audience]]]] is Spiderman.
Extraction of subjects (passive)
The girl [who [it is said [t [t was killed by Ahmad Najib]]]] is Canny.
Extraction of direct object
The nun [who(m) [it is said [t [many loved t]]]] is Mother Theresa.
Extraction of prepositional object
The lady [who(m) [you think [t [David is in love with t]]]] is his wife.
Extraction from the upper clause – Short Movement with complementizer *that*
Extraction of subjects (active)
The man [that [t guards the bank]] is friendly.
Extraction of subjects (passive)
The men [that [t were kidnapped by pirates]] have been released.
Extraction of object
The girl [that [t [someone pushed t]] is badly hurt.
Extraction of preposition
The boys [that [the discipline teacher shouted at t]] were noisy.
II. Extraction from embedded clause – Long Movement with complementizer *that*
Extraction of subject (active)
The people [that [we can say [t love us most]]] are our parents.
Extraction of subject (passive)
The maid [that [we fear [t [t was abused by a woman]]]] is Amy.

Extraction of direct object
The fish [that [you heard [t [many Japanese eat t]]]] is Salmon.
Extraction of prepositional object
The subject [that [you feel [t [students cannot cope with t]]]] is Maths.
Constructions with Null Operator and Null Complementizer
Extraction of direct object
The girl [Op Ø [many of my friends admire]] is rich and pretty.
Extraction of prepositional object
The place [Op Ø [many gamblers are attracted to]] is Genting.

Ungrammatical Stimuli
Relative Clauses with resumptive pronouns
Subject (active)
*The man [who [he has a daughter]] is Mr. Lim.
Extraction of subject (passive)
*The teacher [who(m) [she was praised by the headmaster]] is hardworking.
Direct object
*The comic [which [many adults enjoy it]] is Garfield.
Prepositional object
*The people [who(m) [we donated money to them]] live in Acheh.
Extraction from *Wh*-islands
Extraction of subject (active)
*The boy [who(m) [you wonder [whether [t scored 10 A's in SPM]]]] is Jim.
Extraction of subject (passive)
*The lady [who(m) [we questioned [whether [t was hit by a car]]]] is alive.
Extraction of direct object
*The girl [who(m) [Kim asked [who [John hit t]]]] is Grace.
Extraction of prepositional object
*The thing [which [you ask [whether [children can play with t]]]] is matches.
Extraction from Complex Noun (Determiner) Phrase
Extraction of subject (active)
*The local film [which [I know the story [t [t won an award]]]] is *Sepet*.
Extraction of subject (passive)
*The car [which [they heard the news [t [t was stolen by robbers]]]] is a Kancil.
Extraction of direct object
*The teacher [who(m) [Jane described the way [t [Tim pushed t]]]] is Ms. Tan.

Extraction of prepositional object
*The concert [which [we heard the news [t [Tata sang at t]]]] is *The Forces of Nature*.
Complementizer *that* with resumptive pronouns
Subject (active)
*The fruit [that [it contains a lot of vitamins]] is the orange.
Subject (passive)
*The girl [that [she was scold
Extraction of direct object
*The person [that [recently some dogs bit him]] is a boy.
Extraction of prepositional object
*The lady [that [Jacky Chan is married to her]] is a Taiwanese.
Violation of Spec – Head Agreement
Extraction of subject (active)
*The college [Op Ø [t offers the lowest fees in town]] is TAR College.
*The programme [which that [t everyone watches it]] is Malaysian Idol.