Studies in Second Language Acquisition, 2017, **39**, 431–457.
doi:10.1017/S0272263115000510

MEASURES OF IMPLICIT KNOWLEDGE REVISITED

Processing Modes, Time Pressure, and Modality

Jeong-eun Kim
Korea University, Seoul, Korea

Hosung Nam
Department of English Language and Literature, Korea University, Korea, Haskins Laboratories, New Haven, CT, USA

Timed grammaticality judgment tests (TGJT) and oral elicited imitation tests (OEIT) are considered reliable and valid measures of implicit linguistic knowledge, but studies consistently observe better performances on the TGJT than the OEIT due to the different types of processing they require: comprehension for the TGJT and production for the OEIT. This study examines whether degree of access to implicit knowledge is a function of processing type. Results from a series of factor analyses suggest that the OEIT requires greater access to implicit knowledge—implying that it measures stronger implicit knowledge—than the TGJT. Furthermore, the study examines effects on construct validity of time pressure in the OEIT (uncontrolled vs. controlled) and modality in the TGJT (written vs. aural). The results indicate that the tests reached higher construct validity, or measured *stronger* implicit knowledge, when the OEIT employed controlled time pressure and the TGJT used aural stimuli.

Correspondence concerning this article should be addressed to Hosung Nam, Department of English Language and Literautre, Korea University, Korea Haskins Laboratories, New Haven, CT, USA. E-mail: hnam@korea.ac.kr

© Cambridge University Press 2016

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

INTRODUCTION

Second language acquisition (SLA) researchers generally agree that implicit linguistic knowledge is different from explicit linguistic knowledge, and language competence primarily consists of implicit knowledge of *how* to use a language. Therefore, implicit linguistic knowledge is regarded as the primary source of speech production and comprehension. In SLA, implicit and explicit knowledge are often distinguished in terms of their representation and processing in L2 learning and use (Ellis, 2006). According to Ellis, implicit knowledge "involves unconscious awareness" and allows for "automatic processing" while explicit knowledge involves conscious awareness and requires more "controlled processing" (pp. 432–434).

Sparked by Ellis's (2005) psychometric work, a range of SLA studies have attempted to ascertain the validity and reliability of a test battery designed to measure L2 implicit and explicit knowledge (e.g., Bowles, 2011; Erlam, 2006; Gutiérrez, 2013; Spada, Shiu, & Tomita, 2015; Zhang, 2014). The studies have obtained consistent results from exploratory and/or confirmatory factor analyses, which indicate that timed grammaticality judgment tests (TGJT) and oral elicited imitation tests (OEIT) are better measures of L2 implicit knowledge, whereas untimed grammaticality judgment tests (UGJT) and metalinguistic knowledge tests (MKT) are better measures of L2 explicit knowledge. The design features of the measures of implicit knowledge that distinguish them from the measures of explicit knowledge include (a) experimental manipulation to draw learners' primary attention to L2 meaning and (b) time constraints to reduce the possibility of accessing metalinguistic knowledge. While previous studies support the construct validity of the TGJT and the OEIT, little research has investigated differences in the strength of implicit knowledge the tests assess. Both tap into the same type of knowledge; however, the primary processing modes required for the two tests are fundamentally different: Grammaticality judgment requires receptive responses while elicited imitation draws productive responses. Researchers generally assume that L2 productive ability lags behind L2 receptive ability, and that productive ability is gained incrementally during the process of L2 learning. The incremental nature of development suggests that the two measures might not tap into implicit knowledge that is equal in quality. This study refers to the qualitative difference in implicit knowledge as the *strength* of knowledge.

Furthermore, test administration methods (e.g., written vs. aural sentence stimuli for timed grammaticality judgment; controlled vs. uncontrolled time pressure for elicited imitation) might affect the strength of the implicit knowledge they assess. Although previous studies have consistently shown that both timed grammaticality judgment and elicited imitation tap into implicit knowledge, little attention has been paid to how administration methods affect construct validity. Graham, McGhee,

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

and Millard (2010) pointed out that test administration can be a factor in test results and noted a need for research that could contribute to refining the procedures used for L2 knowledge measurement. But little progress has been made in refining test administration methods to reduce the likelihood of inconsistency from different test procedures. This study attempts to address these issues by examining (a) whether the OEIT and the TGJT assess implicit knowledge differing in strength; and (b) how manipulations of time pressure and modality affect the construct validity of the OEIT and the TGJT, respectively.

LITERATURE REVIEW

Measures of L2 Implicit Knowledge: Grammaticality Judgment and Elicited Imitation

Timed grammaticality judgment and oral elicited imitation are frequently used to observe learners' development of L2 implicit knowledge. The construct validity of these two measures was first explored by Ellis (2005) in a psychometric study that attempted to design relatively separate measures of L2 implicit versus explicit knowledge. Twenty English NSs and 91 L2 learners of English participated in (a) an OEIT, (b) an oral narration test, (c) a TGJT, (d) an UGJT, and (e) a MKT. A principal component factor analysis found that the L2 learners' scores on tests (a), (b), and (c) loaded on Factor 1, whereas the scores on tests (d) (ungrammatical sentences, in particular) and (e) loaded on Factor 2. Based on the study's subsequent analyses, these two factors were interpreted as corresponding to implicit and explicit knowledge, respectively.

Ellis's (2005) findings have been replicated in different languages and with different learner populations (e.g., Bowles, 2011; Zhang, 2014). For example, Bowles (2011) included heritage language (HL) learners, who, unlike L2 learners, may have little explicit knowledge because they learned their HL without formal instruction. Bowles hypothesized that HL learners would gain higher scores on the implicit measures (i.e., TGJT, OEIT, and the oral narration test), and that L2 learners would achieve higher scores on the explicit measures (i.e., UGJT and MKT). The results of a factor analysis replicated Ellis's two-factor model. The HL group scored highest on the TGJT, OEIT, and oral narration test, and lowest on the UGJT and MKT. The opposite pattern was found for the L2 group. More recently, Zhang (2014) conducted another validation study in an English as a foreign language (EFL) context. One hundred Chinese EFL students completed four tests: OEIT, TGJT, UGJT, and MKT. The results of a factor analysis, again, replicated Ellis's model. Zhang found that all items in the OEIT and the grammatical items in the TGJT and UGJT heavily loaded on the implicit factor while the MKT and the ungrammatical items in the TGJT and UGJT

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

heavily loaded on the explicit factor. The same results were replicated by Gutiérrez (2013) in an analysis restricted to the TGJT, UGJT, and MKT, indicating that grammatical and ungrammatical sentences constitute measures of implicit and explicit knowledge, respectively. Taken together, these studies provide support for the construct validity of the test battery they investigate, and they have brought SLA research one step closer to establishing separate measures for implicit and explicit linguistic knowledge. In SLA, these tests are frequently used to examine the differential effects of various instructional interventions on the development of implicit versus explicit knowledge (e.g., Akakura, 2012; Ellis, Loewen, & Erlam, 2006; Erlam & Loewen, 2010; Loewen & Erlam, 2006; Monteiro, 2014).

Processing Modes: Comprehension Versus Production

The previous research (e.g., Bowles, 2011; Ellis, 2005; Erlam, 2006; Gutiérrez, 2013; Spada et al., 2015; Zhang, 2014) has demonstrated that L2 learners' responses to the TGJT and OEIT reflect underlying implicit knowledge in similar ways. Yet they have also reported that learners tend to perform better on grammaticality judgment than elicited imitation. As illustrated in Table 1, L2 learners tend to gain higher scores on the TGJT than the OEIT. Spada et al. (2015) found this pattern regardless of modality; that is, higher scores on both written (M = .72, SD = .14) and aural TGJTs (M = .52, SD = .15) than on an OEIT (M = .15, SD = .16). Unlike learners, native speakers (NSs) and HLs show little difference in their performances on the two test types. L2 learners' different performances on the TGJT and OEIT can be attributed to the different types of processing involved in the two tests: a receptive process (reading/listening) in the TGJT versus a productive process (speaking) in the OEIT.

Table 1. Mean Scores and Standard Deviations across Four Validity Studies (standard deviations in parentheses)

	Ellis, 2005 (n = 111)		Bowles, 2011 (n = 30)		Zhang, 2014 (n = 100)		Spada et al., 2015 (n = 150)		
	$TGJT_W$	OEIT	$TGJT_W$	OEIT	$TGJT_W$	OEIT	$TGJT_W$	$TGJT_A$	OEIT
NSs	.96 (.02)	.94 (.04)	.99 (.01)	.96 (.03)	—	—	.96 (.03)	.98 (.04)	.94 (.06)
L2ers	.82 (.11)	.51 (.17)	.67 (.10)	.46 (.10)	.86 (.08)	.44 (.12)	.72 (.14)	.52 (.15)	.15 (.16)
HLs	—	—	.81 (.11)	.79 (.15)	—	—	—	—	—

Note. $TGJT_W$ = TGJT with written stimuli; $TGJT_A$ = TGJT with auditory stimuli.

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

Studies in SLA have conventionally utilized both receptive and productive tests to measure L2 learning outcomes. However, few attempts have been made to develop a systematic interpretation of learners' performances on the two test types. Some SLA researchers have assumed a shared knowledge base for L2 comprehension and production, arguing that they make use of the same system of representation (Bates & MacWhinney, 1989; Krashen, 1985; Schwartz, 1993). Nevertheless, they have also pointed out that the real-time exigencies of L2 processing for comprehension and production might be quite different (Bates & MacWhinney, 1989). Izumi (2003) also claimed that "learners' productive knowledge lags behind their receptive knowledge" and their "productive ability is much more susceptible than their receptive ability to the processing difficulty" (p. 315).

Following this argument, it can be assumed that access to existing knowledge is constrained by the mode of language processing. L2 vocabulary research provides some evidence for this assumption. Several studies (e.g., Burns, 1951; Eyckmans, Van de Velde, Van Hout, & Boers, 2007; Stoddard, 1929; Waring, 1997) have found that learners tend to show better performance on receptive (e.g., yes/no vocabulary size, checklist, matching form-meaning) than on productive types of vocabulary tests (e.g., translation, fill-in-the blanks). Fitzpatrick (2007) attributed the difference to the different processing operations required for receptive and productive tests. According to him, a receptive test requires learners to access only their L2 orthographic and/or phonological receptive lexicons, while a productive test requires learners to access not only the L2 productive lexicon but also the L1 lexicon, using semantic, collocational, orthographic, and/or phonological routes. If a test requires activation of multiple routes, then it could place a higher processing burden on learners, resulting in poorer performances. Considering the higher processing burden of production, Laufer and Goldstein (2004) introduced the notion of "strength of knowledge" (p. 408), implying that stronger knowledge is required, and therefore assessed, by production tests. According to this notion, productive knowledge is more advanced than receptive knowledge, and thus, more difficult to achieve.

Due to their different natures, the OEIT and TGJT might also impose different degrees of processing burden in accessing implicit knowledge. While the primary processing operation for the TGJT matches L2 grammar against a learner's interlanguage system (i.e., grammatical decoding), the OEIT involves not only grammatical decoding but also grammatical encoding for oral reproduction. According to Craik and Lockhart (1972), information processing "proceeds through a series of sensory stages to levels associated with matching or pattern recognition (i.e., shallow processing) and finally to semantic associative stages of stimulus enrichment (i.e., deep processing)" (p. 675). They referred to this processing hierarchy as "depth of processing" (ibid.); a deeper level implies more

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

processing burden and cognitive effort. The OEIT involves L2 conceptual processing, at a deeper level than the TGJT's shallower, perceptual level, leading learners to make a greater cognitive effort. If the strength of implicit knowledge is what allows it to remain intact even during deeper processing that requires more cognitive effort, the OEIT would be better at tapping into stronger implicit knowledge than the TGJT.[1]

Drawing on this asymmetry of processing burden and cognitive effort in the OEIT and TGJT, this study proposes, first, that poorer performance on the OEIT might be due to the productive nature of the test, which creates a greater processing burden and, second, that the OEIT might therefore assess stronger implicit knowledge than the TGJT. Few studies have discussed how the two measures might tap into implicit knowledge of different strengths. Hence, a more fine-grained observation of the test constructs seems to be needed to understand how and why learners' performances on the OEIT and TGJT might not be equivalent.

Time Pressure on the OEIT: Uncontrolled Versus Controlled

Although the construct validity of the OEIT as a measure of implicit knowledge has been established by many studies (e.g., Bowles, 2011; Ellis, 2005; Erlam, 2006; Spada et al., 2015; Zhang, 2014), several (e.g., Erlam, 2006; Graham et al., 2010; Vinther, 2002) have reported that learners' OEIT performance can be influenced by other factors, such as sentence length, time pressure, lexical difficulty, and sentence complexity. Therefore, more refined OEIT procedures are needed to ensure test validity and reliability. As Ellis (2005) explained, the design of the OEIT encourages learners' responses based on intuition (not rules) by keeping their primary focus on L2 meaning (not form) by asking participants about their belief in statements. Time pressure, which inhibits careful planning and self-monitoring, also helps keep their focus on meaning (Krashen, 1985). In other words, learners' intentional efforts at form-focused self-correction can be reduced (although not completely inhibited) by time constraints during L2 production.

The importance of time pressure in the OEIT can be inferred from previous studies on the relationship between on-line planning and L2 accuracy. On-line planning is "the process by which speakers attend carefully to the formulation stage during speech planning and engage in preproduction monitoring of their speech acts" (Yuan & Ellis, 2003, p. 6). Drawing on Baddeley's model of working memory (Baddeley & Hitch, 1974; Baddeley & Logie, 1999), Yuan and Ellis (2003) suggested that, when learners have time to plan, the central executive of working memory operates, enabling learners to search their long-term memory for grammatical encoding. This operation involves greater attention to

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

form, which leads to enhanced accuracy. Indeed, in their empirical study, Yuan and Ellis found that a group with on-line planning opportunities exhibited greater grammatical accuracy than a group that could not plan. A similar result was reported by Hulstijn and Hulstijn (1984), who found that when learners producing L2 oral narratives directed their focus to form, they showed higher accuracy if they were not pressured for time. These findings illustrate that time pressure may inhibit learners from accessing grammatical knowledge and monitoring their output during production. Therefore, the use of time pressure during a test measuring learners' production can be understood to enhance test validity.

While researchers are aware of the importance of time pressure in eliciting learners' implicit knowledge, very few studies have explicitly discussed how much time pressure is necessary or what factors (e.g., sentence length, number of syllables) to consider in estimating the amount of time pressure for the OEIT. Many studies employing OEITs (e.g., Akakura, 2012; Bowles, 2011; Ellis, 2005; Ellis et al., 2006; Monteiro, 2014; Soleimani, Jahangiri, & Gohar, 2015; Zhang, 2014) do not explain how they operationalized time pressure. Others (Erlam, 2006; Erlam & Loewen, 2010) administered one-on-one tests, in which time pressure comes from the need to quickly respond in a face-to-face situation. Some of these studies note that their coding methods exclude productions that show self-correction, which may indicate the use of explicit knowledge (Ellis et al., 2006; Erlam, 2006). More recently, Spada et al. (2015) employed time pressure by imposing an eight-second limit for the oral reproduction of each test item. Graham et al. (2010) employed sentence-length dependent time pressure in their OEIT. Based on sentence length, the learners were given from six seconds (4–6 syllables) to 12 seconds (16–19 syllables). Although these studies controlled time pressure, their choice of estimated time seems somewhat arbitrary rather than systematic.

In current practices of measuring implicit knowledge, there is no agreement on methods of estimating time constraints on the OEIT. Strictly controlled time pressure may improve construct validity by eliciting stronger implicit knowledge because it can more effectively inhibit learners from accessing explicit knowledge.

Modality of the TGJT: Written Versus Aural

Research in SLA provides empirical evidence that learners' attentional capacity is differentially constrained in aural versus written processing (e.g., Leow, 1993, 1995; Lund, 1991; Wong, 2001). Studies have found higher recall rates of visually presented information (Lund, 1991; Wong, 2001) and increased L2 intake in the written mode, even with nonsalient target forms (Leow, 1993, 1995). Cognitive psychology studies

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

(e.g., Anderson, 1980; Clifton, Meyer, Wurm, & Treiman, 2013; Danks, 1980; Rost, 1990; Sticht & James, 1984) have noted that processing speech sounds (listening comprehension) is more challenging than processing written text (reading comprehension) because of the evanescent nature of aural input. Unlike text processing, aural processing forces learners to segment and comprehend continuous speech sounds while perceiving new material. Various cognitive strategies facilitating language comprehension, such as backtracking, skimming, and focusing on/ignoring certain features of input, are not available in real-time processing. In contrast, written text remains available for reprocessing, giving readers relatively more control over the input and rate of processing and allowing them to use various cognitive strategies for effective reading. The differences in control over aural and written input have been demonstrated by L2 learners' poorer performances in aural conditions.

In SLA, the TGJT is one of the most commonly used measures of L2 learners' implicit knowledge. However, there is no consensus on the use of modality (written vs. aural) in the TGJT. It is plausible that a TGJT employing aural input might need more effort and thus use more cognitive capacity. If the aural TGJT requires learners to process meaning before form, it could be a better candidate to assess implicit knowledge than the written TGJT.

A few studies (e.g., Johnson, 1992; Murphy, 1997) have examined the influence of UGJT modality (aural vs. written) on learners' access to L2 grammatical knowledge, but not on their access to L2 implicit knowledge. For example, Johnson (1992) investigated the influence of the written versus aural mode in UGJTs assessing L2 learners' grammatical competence. In an earlier study, Johnson and Newport (1989) employed an aural UGJT to examine the validity of the critical period hypothesis in L2 acquisition, and found that a late L2 learner group scored lower than an early L2 learner group and showed a high degree of interlearner variability. To examine whether its poor performance was due to the nature of the aural processing the test required, Johnson (1992) further tested the same group of late L2 learners with the written UGJT, and found that they made 2.5 times fewer errors in the written version. The results suggest that some source of L2 knowledge is more readily accessible during written input processing than during aural input processing. Another study, by Murphy (1997), tested modality's role in the accessibility of Universal Grammar in adult SLA. Assuming that aural input is more effort demanding than written input, Murphy predicted less accurate and slower responses in the aural GJT. L1 English–L2 French and L1 French–L2 English learners completed both aural and written GJTs in both languages. The participants were instructed to judge grammaticality as quickly as possible. Accuracy rates and reaction times were the dependent variables. The results found a modality effect in the L2 but not the L1: Participants completed the written GJT more

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

accurately and quickly than the aural one in their L2, and the effect was more pronounced for ungrammatical sentences. The learners' better performance on the written GJT was attributed to the lower burden of processing written input than aural input in an L2, and Murphy posited that a lower processing burden might facilitate access to Universal Grammar. Although these previous studies have reported consistent findings regarding the effect of modality, they used untimed GJTs. Therefore, the construct validity of the TGJT as a function of modality still remains questionable.

In current practices of measuring implicit knowledge, there is no consensus on the choice of modality for stimuli presentation in the TGJT (i.e., aural vs. written). Spada et al. (2015), however, provided some evidence of interaction between the modality and construct validity of the TGJT. Using an exploratory factor analysis, the researchers found that the TGJT with aural stimuli loaded more heavily on the factor corresponding to implicit knowledge (.77) than the TGJT with written stimuli (.59). By contrast, the TGJT with written stimuli appeared to load with a similar weight (.51) on the factor corresponding to explicit knowledge, while the TGJT with aural stimuli loaded on the same factor at nearly zero (.04). Brown (1990) pointed out that aural processing can be particularly challenging for L2 learners who have learned the target language in written form and not been frequently exposed to speech. Because aural input requires a higher processing burden and more cognitive effort than written input, it is possible that the TGJT in the aural mode could better lead learners to maintain a primary focus on meaning rather than form. This would, in turn, result in greater construct validity of the aural TGJT as a measure of stronger implicit knowledge.

As this review suggests, the three variables of (a) processing mode, (b) time pressure, and (c) modality may affect the test construct of implicit knowledge in terms of its strength. Previous studies have demonstrated that the OEIT and TGJT are reliable and valid measures of implicit knowledge. The current study aims to further examine whether manipulating these three variables can enhance the tests' construct validity by leading the measures to tap into stronger implicit knowledge. The effect of the three independent variables (i.e., processing mode, time pressure, and modality) on the strength of implicit knowledge was examined by a series of factor analyses.

This study posited the following three research questions:

1. Processing mode: Does the implicit knowledge tapped by the OEIT and the TGJT differ in strength?
2. Time pressure: For the OEIT, does the use of uncontrolled versus controlled time pressure tap into implicit knowledge differing in strength?
3. Modality: For the TGJT, does the use of written versus aural sentence stimuli tap into implicit knowledge differing in strength?

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

METHOD

Participants

A total of 75 participants, nine NSs of English and 66 adult EFL learners, took part in this study. The NSs (n = 9: male = 1, female = 8) were undergraduate exchange students at a Korean university. They were all in their early twenties (M = 20.88, SD = 1.96). They were enrolled in introductory linguistics courses, and received extra credit as compensation for participating. Their majors were economics, Asian studies, international studies, or English language and literature. Seven came from the United States and two from Singapore, and all identified English as their first language. The nonnative group (n = 66: male = 19, female = 47) were all regular (n = 64) or exchange (n = 2) students enrolled in the same university in Korea, (double) majoring or minoring in English language and literature. They were aged between 19 and 23 (M = 21.29, SD = 1.60). Sixty-one students identified their first language as Korean, three as Chinese, one as Danish, and one as Arabic. Thirty-one students reported having been to English-speaking countries (e.g., the United States, Canada, Australia, England, Hong Kong, and New Zealand) for purposes of tourism, family visits, or study abroad; 23 had lived in such countries for more than six months. Thirty-five students reported that they had learned English only in Korea in formal language learning contexts (e.g., public and/or private classes taught by native/nonnative instructors). According to their self-reported English proficiency on a six-point Likert scale (1: Beginner, 2: Intermediate, 3: High-intermediate, 4: Advanced, 5: Near-native, 6: Native), the learners appeared to have advanced English proficiency in the four language skills: listening (M = 4.86, SD = 1.17), speaking (M = 4.06, SD = 1.32), reading (M = 4.79, SD = 1.07), and writing (M = 4.21, SD = 1.17).

Test Content

The present study replicated the tests and test content used by Ellis (2005). Ellis employed 17 English grammatical structures to measure learners' implicit and explicit L2 knowledge. These structures were chosen by considering (a) target difficulty (universally problematic to L2 learners in general), (b) timing of acquisition within the developmental process (acquired earlier or later), (c) levels of L2 proficiency when the targets are introduced (beginner, lower intermediate, upper intermediate, and advanced), and (d) linguistic aspects of the targets (morphological and syntactic features; see Ellis, 2005, p. 155 for the summarized properties of the 17 target structures).

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

Test Battery

Three of the five tests utilized by Ellis (2005) were used in this study: (a) the OEIT; (b) the TGJT; and (c) the MKT. The present study hypothesized that different administration methods would influence construct validity as a measure of implicit knowledge. Therefore, the study manipulated time pressure methods for the OEIT and modality for the TJGT, creating two different versions of each: an OEIT with uncontrolled time pressure ($OEIT_{UNCTRL}$) versus an OEIT with controlled time pressure considering sentence length ($OEIT_{CTRL}$), and a TGJT with written stimuli ($TGJT_{WRITTEN}$) versus a TGJT with aural stimuli ($TGJT_{AURAL}$). No manipulation was made on the MKT, a measure of explicit knowledge. The auditory stimuli for the tests were recorded by a male NS of English. The same target forms were used for all five tests, but sentences were not reused. An independent samples *t*-test to check the equivalence of the sentences in the GJTs and OEITs found no statistical difference in terms of number of characters (GJT: $t(66) = .000$, $p > .05$; OEIT: $t(66) = -.813$, $p > .05$), phonemes (GJT: $t(66) = .230$, $p > .05$; OEIT: $t(66) = .147$, $p > .05$), syllables (GJT: $t(66) = -.050$, $p > .05$; OEIT: $t(66) = -.554$, $p > .05$), or words (GJT: $t(66) = -.130$, $p > .05$; OEIT: $t(66) = -.801$, $p > .05$). The tests were individualized and computerized using MATLAB R2014a and Psychtoolbox (ver. 3.0.12).

OEIT with Uncontrolled Time Pressure ($OEIT_{UNCTRL}$). This test comprised 34 statement sentences employing 17 target structures (two sentences per target structure: one grammatical and one ungrammatical; see Appendix 1). The sentences were presented aurally to the participants (e.g., **Young women like cigarettes and fast car.*), who were then directed to a dichotomous comprehension question (e.g., *Do young women like slow cars?*).[2] The questions were presented visually on a screen, and the participants answered them by pressing designated keys on a keyboard (right shift key for *yes* and left shift key for *no*). The comprehension questions were used to draw learners' primary attention to L2 meaning over form. This use of comprehension questions differs from that of Ellis (2005), who asked participants whether they agreed with, disagreed with, or were not sure about the content of statements. Because all answers to such questions are correct, they cannot ensure learners' attention to meaning, and there is no way to exclude random responses. Hence, the current study asked comprehension questions with only one correct answer; learners could not answer correctly if they responded randomly without processing meaning. The participants were next asked to repeat the sentences orally in correct English as quickly as possible (e.g., *Young women like cigarettes and fast cars.*). Their responses were automatically recorded as wave files.[3] As in previous studies, this task

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

had no controlled time pressure. However, the learners were pressured *emotionally* by the instruction to "repeat the sentence as quickly as possible." The OEIT was designed to assess whether the learners could reproduce the target sentences correctly while focusing on meaning. Therefore, given a target sentence, it was scored as correct if and only if the comprehension question was answered correctly and the target sentence was reproduced in correct English. All correct answers were given one point and all incorrect answers were given zero points. Scores on all items, grammatical items, and ungrammatical items were then expressed as percentages.

OEIT with Controlled Time Pressure ($OEIT_{CTRL}$). This test comprised 34 new statement sentences employing the 17 target structures (two sentences per target structure: one grammatical and one ungrammatical; see Appendix 1). The administration and scoring were identical to those used for the $OEIT_{UNCTRL}$ except that the participants were limited to a computed time period to repeat the sentences. While the participants were repeating the sentences, a time bar appeared on the screen, indicating the amount of time they had left. The participants were directed to the next item automatically when the allotted time was reached. The time limit for each sentence varied as a function of the number of syllables. The procedures for estimating the time limits were as follows: First, production times from the nine NSs with no time pressure were measured by identifying the offset of the produced utterance using Praat (ver. 5.3.65). Second, the median of the production times (t) for each sentence across speakers was computed to exclude outliers. Third, the median production times (T) were averaged across sentence items. Fourth, the mean of the median production times ($\overline{T} = 3.04$) was used to estimate the scaling coefficient (μ) for syllable-dependent time limit as in equation (1). The mean of the syllable numbers ($\overline{S}$) for all sentences was 13.97. The mean of the median production times from the NSs was increased by 20% for the non-NSs (see Ellis, 2005, p. 156, for a justification for this adjustment).

(1) $\mu = \overline{T}(1+0.2)/\overline{S}$
$\mu = 3.04(1+0.2)/13.97$
$\mu = 0.2611$

Finally, using the estimated coefficient (μ) of .2611, a time limit was calculated for each sentence by multiplying the coefficient by the number of syllables (S). Equation (2) was used to compute the time limit for each sentence.

(2) Time pressure for i-th sentence = $S_i \times 0.2611$

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

TGJT with Written Stimuli ($TGJT_{WRITTEN}$). This test consisted of 34 sentences (two for each target structure), evenly divided between grammatical and ungrammatical sentences (see Appendix 2). Each sentence was presented in written form on a computer screen. The participants were asked to judge whether the sentence was grammatical or ungrammatical by pressing the designated keys on the keyboard within a fixed time limit. Response times from the NSs were used to establish the time limit for each sentence. As mentioned, median values were used to exclude outliers, and an additional 20% was added due to the slower processing speed of L2 learners (Ellis, 2005). The time limit for individual sentences ranged from 1.55 to 6.43 seconds. Scoring of the written TGJT was dichotomous (correct or incorrect). All unanswered items were scored as incorrect. The accuracy rates on all items, grammatical items, and ungrammatical items were then calculated as percentages.

TJGT with Aural Stimuli ($TGJT_{AURAL}$). This test employed 34 new sentences (two for each target structure), evenly divided into grammatical and ungrammatical sentences (see Appendix 2). The sentences were presented aurally, and the learners were asked to judge the grammaticality within a fixed time limit after listening to each sentence. The time limit ranged from .37 to 2.99 seconds. The scoring method was identical to that used for the written TGJT.

Metalinguistic Knowledge Test. This study used the first of the two sections of the MKT used by Ellis (2005), administering it in the same manner as in Ellis's study. The test consisted of 17 multiple-choice items. Each item was an ungrammatical sentence, and the participants were instructed to select one out of four options that best explained the ungrammaticality. There was no time pressure. One point was credited to correct answers while no point was credited to incorrect answers. A total percentage accuracy score was then calculated.

Procedure

The experiment in this study involved two rounds of data collection. The first round with English NSs provided the information for the time pressure estimations for the timed tests. The second round with English non-NSs provided the data for the construct validity analyses. Prior to participating in the experiment, all participants completed an online background questionnaire asking about their personal information, linguistic background, and language learning experiences.

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

For the first round of data collection, each English NS participant completed three tests in the following order: (a) untimed $GJT_{WRITTEN}$ (34 items), (b) untimed GJT_{AURAL} (34 items), and (c) untimed OEIT (64 items). Each test was preceded by six practice items. They were asked to respond as quickly as possible based on their native intuition. Because the results from the native participants were used to estimate the time pressure for the nonnative participants, the tests were untimed. The completion of all tests took approximately 30 minutes: 10 minutes for both GJTs and 20 minutes for the OEIT. The tests were computerized and individually administered in a recording booth. The NS participants did not take the MKT because no time pressure estimation was necessary for the MKT.

For the second round of data collection, the non-NSs of English completed five tests in the order of (a) $TGJT_{WRITTEN}$ (34 items), (b) $TGJT_{AURAL}$ (34 items), (c) $OEIT_{UNCTRL}$ (34 items), (d) $OEIT_{CTRL}$ (34 items), and (e) MKT (17 items). Each test started with six practice items. The first four tests were timed. The completion of all tests took approximately 45 minutes for each participant: 10 minutes for the TGJTs, 20 minutes for the OEITs, and 15 minutes for the MKT. The tests were computerized and individually administered in a recording booth. All of the participants' responses were automatically saved and recorded.

RESULTS

First, the reliability of each test was assessed, based on the nonnative participants' performance, by Cronbach's alpha (α) coefficient (Table 2). These ranged between .76 and .95, indicating acceptable levels of internal consistency. (Values greater than .70 indicate that results from individual test items are not caused by specific test items; Field, 2009.)

Descriptive statistics (Table 3) demonstrate that the NSs showed nearly or more than 95% accuracy on all tests: $OEIT_{UNCTRL}$ (M = 92.12, SD = 5.68), $OEIT_{CTRL}$ (M = 95.75, SD = 2.14), $TGJT_{WRITTEN}$ (M = 96.04,

Table 2. Reliability Measures for the Five Tests (nonnative participants only: n = 66)

Test types	Version	Items	Reliability
Oral elicited imitation test	Uncontrolled time pressure	34	α = .81
	Controlled time pressure	34	α = .90
Timed grammaticality judgment test	Written	34	α = .80
	Aural	34	α = .76
Metalinguistic knowledge test	—	17	α = .95

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

Table 3. Descriptive Statistics for the Five Tests

		Native speakers (n = 9)			Nonnative speakers (n = 66)		
Tests	Grammaticality	Count	Mean (%)	SD (%)	Count	Mean (%)	SD (%)
$OEIT_{UNCTRL}$	All (*n* = 34)	31.32	92.12	5.68	21.98	64.66	15.62
	G (*n* = 17)	16.33	96.08	5.88	12.61	74.17	17.11
	UG (*n* = 17)	14.99	88.15	7.13	9.37	55.14	17.08
$OEIT_{CTRL}$	All (*n* = 34)	32.56	95.75	2.14	15.16	44.58	20.79
	G (*n* = 17)	15.89	93.46	1.96	8.56	50.37	22.52
	UG (*n* = 17)	16.67	98.04	2.94	6.59	38.79	21.11
$TGJT_{WRITTEN}$	All (*n* = 34)	32.65	96.04	5.12	20.88	61.40	16.01
	G (*n* = 17)	16.56	97.39	7.84	12.94	76.10	17.34
	UG (*n* = 17)	16.08	94.61	6.49	7.94	46.69	20.21
$TGJT_{AURAL}$	All (*n* = 34)	32.00	94.12	3.89	19.11	56.20	14.67
	G (*n* = 17)	16.44	96.73	5.96	12.67	74.54	12.13
	UG (*n* = 17)	15.56	91.50	5.19	6.44	37.87	22.35
MKT[a]	UG (*n* = 17)		—	—	11.94	70.22	15.43

Note. G = grammatical sentences; UG = ungrammatical sentences.
[a] Only non-NSs completed the MKT.

SD = 5.12), and $TGJT_{AURAL}$ (*M* = 94.12, *SD* = 3.89). In contrast, the nonnative group showed accuracy rates of approximately 50% for the $TGJT_{AURAL}$ (*M* = 56.20, *SD* = 14.67) and the $OEIT_{CTRL}$ (*M* = 44.58, *SD* = 20.79), and 60% for the $TGJT_{WRITTEN}$ (*M* = 61.40, *SD* = 16.01) and the $OEIT_{UNCTRL}$ (*M* = 64.66, *SD* = 15.62). The nonnative group's greater standard deviations for each test indicate a higher degree of interlearner variability. The nonnative group scored highest on the MKT (*M* = 70.22, *SD* = 15.43).

A correlation analysis was conducted to examine whether a linear relationship existed between any two test variables (Table 4), and revealed significant correlations between every one of the four test variables and any other one ($OEIT_{UNCTRL}$, $OEIT_{CTRL}$, $TGJT_{WRITTEN}$, and $TGJT_{AURAL}$), with meaningful positive coefficient values between .363

Table 4. Relationships between the Five Tests (nonnative participants: n = 66)

Tests	$OEIT_{UNCTRL}$	$OEIT_{CTRL}$	$TGJT_{WRITTEN}$	$TGJT_{AURAL}$	MKT
$OEIT_{UNCTRL}$	—	.664**	.363**	.517**	.236
$OEIT_{CTRL}$		—	.383**	.521**	–.019
$TGJT_{WRITTEN}$			—	.529**	.182
$TGJT_{AURAL}$				—	.055
MKT					—

** $p < .001$.

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

and .664 ($p < .001$ for all). Specifically, there was a strong positive correlation between $OEIT_{UNCTRL}$ and $OEIT_{CTRL}$, $r(64) = .664$, $p < .001$, $OEIT_{UNCTRL}$ and $TGJT_{AURAL}$, $r(64) = .517$, $p < .001$, $OEIT_{CTRL}$ and $TGJT_{AURAL}$, $r(64) = .521$, $p < .001$, and $TGJT_{WRITTEN}$ and $TGJT_{AURAL}$, $r(64) = .529$, $p < .001$. The relationship between the OEITs and the $TGJT_{WRITTEN}$ appeared to be moderately positive: $OEIT_{UNCTRL}$ and $TGJT_{WRITTEN}$, $r(64) = .363$, $p < .001$; $OEIT_{CTRL}$ and $TGJT_{WRITTEN}$, $r(64) = .383$, $p < .001$. The MKT, by contrast, was not significantly correlated with any of the other four tests, indicating that level of metalinguistic knowledge had no relationship with the other tests. These results may suggest that the OEIT and the TGJT (but not the MKT) do partially tap into the same construct of L2 knowledge.

A series of principle component factor analyses was conducted utilizing a Varimax rotation for the nonnative participants' test scores. The appropriate number of factors was first decided by the following criteria: eigenvalue (greater than one), variance (at least 70% of total variance; Table 5), and scree plot (components within the sharp descent). These criteria indicated that three components should be investigated. The inclusion of three components increased the model fit as it decreased the number of residuals exceeding the .05 criteria. After rotation, the first component accounted for 35.95%, the second one for 28.70%, and the third one for 21.24%. Thus, the three components accounted for 85.90% of the total variance.

Table 6 presents the loadings of the five tests for each component. Factors considered significant were based on criteria proposed by Comrey and Lee (1992): A factor score higher than .71 is considered excellent, .63–.70 is very good, .55–.62 is good, .45–.54 is fair, and .32–.44 is poor. Scores under .32 are discarded. A cutoff of .71 (*excellent*) was used as the threshold to determine the test constructs. As shown in Table 6, the $OEIT_{UNCTRL}$ and the $OEIT_{CTRL}$ loaded heavily on Factor 1 (.779 and .871, respectively). The $TGJT_{WRITTEN}$ loaded heavily on Factor 2 (.921). The $TGJT_{AURAL}$ reached the threshold on Factor 2 (.696) and, at the same time, loaded fairly strongly on Factor 1 (.495). The MKT loaded heavily on Factor 3 (.984).

The data were then analyzed using the software R and the Lavaan package for the confirmatory factor analysis and the structural equation model (Rosseel, 2011). The analysis provided support for a three-factor

Table 5. Principle Component Factor Analysis: Eigenvalues of the Components (before rotation)

Component	Eigenvalue	Variance (%)	Cumulative (%)
1	4.018	44.648	44.648
2	1.422	15.797	60.445
3	1.072	11.906	72.351

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

Table 6. Loadings for Principal Component Factor Analysis

	Factor 1	Factor 2	Factor 3
$OEIT_{UNCTRL}$	.779	.286	.139
$OEIT_{CTRL}$	.871	.151	-.118
$TGJT_{WRITTEN}$	.140	.921	.139
$TGJT_{AURAL}$	.495	.696	-.058
MKT	.047	.073	.984

model (Figure 1), as indicated by an acceptable model fit: χ^2 = .717, *df* = 2, *p* = .332; root mean square error of approximation (RMESA) = .01; normed fit index (NFI) = .908; standardized root mean square residual (SRMR) = .02; comparative fit index (CFI) = .932.

Ellis (2005), in his psychometric study, advised researchers to carefully distinguish grammatical and ungrammatical sentences in analyzing test scores because they can influence what is measured. Following this suggestion, another factor analysis was conducted to examine whether the factor loadings would differ according to sentence grammaticality.

Again, a three-factor solution was specified with the components (eigenvalue > 1.0 for all and 79.46% of total variance; Table 7). The analysis (Table 8) revealed that both grammatical (G) and ungrammatical (UG) sentences of the $OEIT_{UNCTRL}$ (G = .743 and UG = .784) and the $OEIT_{CTRL}$ (G = .879, UG = .875) loaded heavily on Factor 1. As for the $TGJT_{WRITTEN}$, both grammatical and ungrammatical sentences loaded heavily on

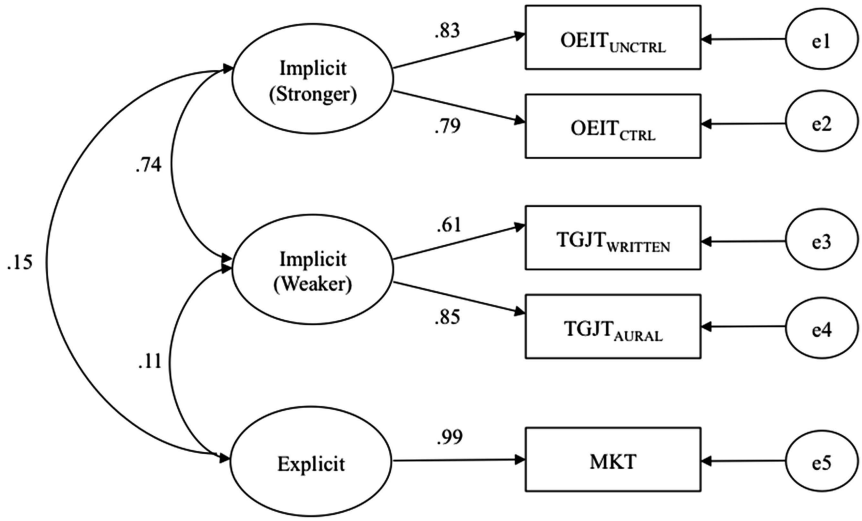


Figure 1. CFA for implicit versus explicit knowledge, testing the three-factor mode.

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

Table 7. Principle Component Factor Analysis: Eigenvalues of the Components (before rotation)

Component	Eigenvalue	Variance (%)	Cumulative (%)
1	2.796	46.598	46.598
2	1.422	18.699	65.297
3	1.039	14.158	79.455

Factor 2 (G = .801, UG = .734). The $TGJT_{AURAL}$ differentially loaded according to grammaticality. While the grammatical sentences of the $TGJT_{AURAL}$ loaded heavily on Factor 2 (.742), the ungrammatical sentences of the $TGJT_{AURAL}$ loaded heavily on Factor 1 (.624). The MKT still loaded heavily on Factor 3 (.886), on which the grammatical sentences of the $OEIT_{UNCTRL}$ also loaded fairly heavily (G = .444).

The data were once again analyzed using the software R and the Lavaan package for the CFA and the structural equation model. The analysis, again, provided support for a three-factor model (Figure 2), as indicated by an acceptable model fit: χ^2 = .558, *df* = 8, *p* = .257; RMESA = .04; NFI = .786; SRMR = .09; CFI = .890.

DISCUSSION

Processing Modes and Strength of Implicit Knowledge

This study dealt first with possible differences in the strengths of implicit knowledge accessed by different processing modes in the OEIT (production) and TGJT (reception). To this end, the nonnative participants' scores on the OEIT and the TGJT were analyzed by a principal component factor analysis. If the two tests tap into implicit knowledge with no

Table 8. Loadings for Principal Component Factor Analysis

	Grammaticality	Factor 1	Factor 2	Factor 3
$OEIT_{UNCTRL}$	G	.743	.095	.444
	UG	.784	.219	.253
$OEIT_{CTRL}$	G	.879	.104	-.167
	UG	.875	.171	-.024
$TGJT_{WRITTEN}$	G	.004	.801	.184
	UG	.380	.734	-.126
$TGJT_{AURAL}$	G	.201	.742	.180
	UG	.624	.412	-.233
MKT	-	-.001	.157	.886

Note. G = grammatical sentences; UG = ungrammatical sentences.

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

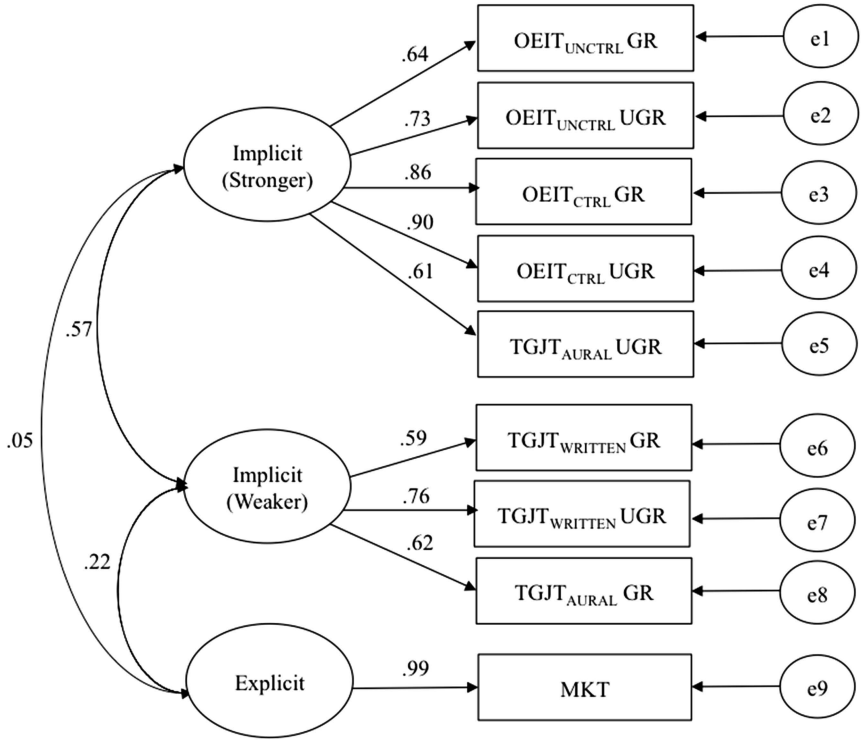


Figure 2. CFA for grammatical versus ungrammatical sentences, testing the three-factor mode.

difference in quality, they should load heavily on the same factors. In contrast, if the OEIT and the TGJT tap into implicit knowledge with a difference in quality, the two tests should load on different factors among the identified components. In any case, their factor(s) should not be shared with the MKT to maintain their construct validity as measures of implicit knowledge.

The scores of the tests loaded largely on different factors: the $OEIT_{UNCTRL/CTRL}$ on one factor (Factor 1), the $TGJT_{WRITTEN/AURAL}$ on another (Factor 2), and the MKT on a third (Factor 3). The OEIT and the TGJT loaded heavily on the factors (Factors 1 and 2) that were not heavily loaded on by the MKT. Interpreting Factor 3, on which the MKT did load heavily, as associated with explicit knowledge, this finding confirms the construct validity of the OEIT and the TGJT as relative measures of implicit knowledge. By contrast, the different factor loadings of the OEIT (Factor 1) and the TGJT (Factor 2) can be interpreted as associated with the quality of implicit knowledge tapped into by these two tests: that is, implicit knowledge that differs in strength. Much as in the previous studies (e.g., Bowles, 2011; Ellis, 2005; Spada et al., 2015; Zhang, 2014),

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

the learners in this study scored significantly higher on the TGJTs than the OEITs, $t(63) = 2.315$, $p < .05$. In contrast, the native group's performances were not statistically different on the two, $t(8) = .558$, $p > .05$. Given the L2 learners' lower accuracy rates on the OEITs than on the TGJTs, Factor 1 seems to correspond to more advanced implicit knowledge, and therefore stronger implicit knowledge, than Factor 2. On the continuum of the strength of implicit knowledge (weak to strong), the implicit knowledge assessed by the OEIT seems to be nearer to the strong end than the implicit knowledge assessed by the TGJT.

The differences of access to implicit knowledge in the two tests can be accounted for by the notion of "depth of processing" (Craik & Lockhart, 1972). The OEIT taps into stronger implicit knowledge because it involves L2 processing at a deeper level, which requires greater cognitive effort. The performance gap between the OEIT and the TGJT should decrease as learners gain more proficiency in processing L2 information automatically. According to the information processing account, L2s are learned through the "gradual integration of subskills as controlled processes initially predominate and then become automatic" (McLaughlin, 1987, p. 139; McLaughlin, Rossman, & McLeod, 1983). This implies that a learner's reliance on controlled processing gradually shifts to the automatic use of implicit procedures, and hence, the amount of L2 information processed at a time gradually increases. In this sense, the accessibility of implicit knowledge may be indicative of the degree of L2 information a learner can process automatically without cognitive effort. Thus, the learners' superior TGJT performance can be accounted for by the fact that the TGJT items are more likely to be completed by automatic processes that learners develop at the earlier stages of L2 learning.

The results, however, could also be interpreted differently: Learners' performance on the OEIT might not represent automatic processing of implicit knowledge but the use of automatized explicit knowledge (DeKeyser, 2003, 2009). The validity of the OEIT as a measure of implicit knowledge was challenged by Suzuki and DeKeyser (2015); they found that OEIT scores were positively correlated with metalinguistic knowledge, but not with a serial reaction time task (an aptitude measure for implicit learning). However, Suzuki and DeKeyser's study has a similar limitation as previous studies in terms of time pressure estimation. To limit the L2 Japanese participants' access to explicit knowledge, they allowed them eight seconds to reproduce each stimuli sentence, whose average length was 24.1 morae. The average speech rate of modern Japanese is 7.48 morae per second (Fukumori, 2008 cited in Suzuki & DeKeyser, 2015), so eight seconds represents a 270% increase of the normal time for NSs. The present study allowed only 20% more than the normal native-speaker speed, for an average time pressure of 3.60 seconds ($SD = .67$). Suzuki and DeKeyser's finding of a positive correlation between the OEIT and metalinguistic knowledge might be

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

a result of their participants having time to access explicit knowledge. Due to this discrepancy in time pressure methods, therefore, Suzuki and DeKeyser's findings do not provide strong grounds for interpreting the learners' scores on the OEIT in the present study as evidence of automatized explicit knowledge.

Given these considerations, this study's findings suggest that the OEIT can be expected to be capable of assessing stronger implicit knowledge that is stable and robust enough not to be susceptible to processing burden. Note that the native group's performance did not differ across tests. For this reason, along with the results of the factor analysis, it can be argued that the OEIT taps into a more advanced and stronger type of implicit knowledge than does the TGJT.

Time Pressure, Modality, and Strength of Implicit Knowledge

The second and third research questions concerned the effects of time pressure and modality on the construct validity of the OEIT and the TGJT in relation to the strength of implicit knowledge. Specifically, the second research question addressed the influence of time pressure (uncontrolled vs. controlled) for the OEIT, and the third addressed the influence of the modality of stimuli presentation (written vs. aural) for the TGJT. The effects of test administration methods were explored with the principal component factor analysis on the nonnative participants' test scores. If the two different versions of each test type separately loaded on different factors, it would indicate that the variables of test administration (i.e., time pressure and modality) affect the primary construct of the test.

Factor 1 was interpreted to indicate more advanced, stronger implicit knowledge and Factor 2 to indicate less advanced, weaker implicit knowledge. While both versions of the OEIT tended to load heavily on Factor 1, the factor score of the $\text{OEIT}_{\text{CTRL}}$ was greater than that of the $\text{OEIT}_{\text{UNCTRL}}$ (.871 and .779, respectively). Comparing the L2 learners' performances on the two versions, a paired-sample *t*-test revealed that the learners tended to score significantly lower on the $\text{OEIT}_{\text{CTRL}}$ than on the $\text{OEIT}_{\text{UNCTRL}}$, $t(64) = -10.261$, $p < .001$. This finding might suggest that the construct validity of the OEIT is stronger when it employs controlled time pressure computed for each sentence.

Because the $\text{OEIT}_{\text{UNCTRL}}$ also exhibited a high factor score on Factor 1, one might argue that uncontrolled time pressure is sufficient to be operationalized in the OEIT. However, the current data partially suggest that the construct validity of the $\text{OEIT}_{\text{UNCTRL}}$ can be impacted by learners' use of explicit knowledge. Notably, the grammatical sentences of the $\text{OEIT}_{\text{UNCTRL}}$ loaded fairly strongly on Factor 3 (.444), which was interpreted

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

as corresponding to explicit knowledge. It is plausible that L2 learners under no time pressure would rely more on explicit knowledge, which would in turn negatively affect the construct validity of the OEIT. This result conforms to the claim that time pressure is critical to reduce the possibility of learners accessing explicit knowledge (Bialystok, 1979; Ellis, 2004).

Unlike the OEITs, the TGJTs tended to load heavily on Factor 2. The $TGJT_{WRITTEN}$ exhibited a higher factor score on Factor 2 than the $TGJT_{AURAL}$ (.921 and .696, respectively). In addition, the $TGJT_{AURAL}$ loaded fairly strongly on Factor 1 (.495) whereas the $TGJT_{WRITTEN}$ did not (.140). A separate analysis of test scores according to sentence grammaticality found that the loading of the $TGJT_{AURAL}$ on Factor 1 was driven by the ungrammatical sentences, which strongly loaded on Factor 1 (.624). A further paired-sample *t*-test showed that the L2 learners tended to score significantly lower on the $TGJT_{AURAL}$ than on the $TGJT_{WRITTEN}$, $t(64) = -2.781$, $p < .01$. These results indicate that, while both TGJTs generally tap into a type of implicit knowledge, the test can more readily tap into more advanced, stronger implicit knowledge when the stimuli are presented aurally.

Previous studies also provide some support for the differential effects of modality on the construct validity of the TGJT. Leow (1993, 1995) reported that L2 learners of Spanish tended to take in salient (present perfect) and less salient (present subjunctive) forms equally well from written input, but took in the salient form significantly better from aural input. Wong (2001) found that L2 comprehension can be negatively affected when learners pay extra attention to a specific function word (e.g., English article *the*) in an aural mode, but not in a written mode. These researchers concluded that learners have difficulty processing both form and meaning and thus tend to focus on meaning at the sacrifice of form during aural input processing. Likewise, the TGJT with aural input may hinder learners from maintaining attention to L2 form due to the need to process meaning. In fact, in the current study, the $TGJT_{WRITTEN}$ still appears to tap into an implicit type of knowledge. According to the factor scores of the TGJTs, however, these tests tend to tap into more advanced, stronger implicit knowledge when they are operationalized with aural stimuli.

Grammaticality and Strength of Implicit Knowledge

As mentioned, the $TGJT_{AURAL}$ loaded strongly on Factor 1, particularly when the sentences were ungrammatical (.624). These results suggest that rejection of ungrammatical sentences required stronger implicit knowledge than endorsement of grammatical sentences in the $TGJT_{AURAL}$.

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

Previous studies (Ellis, 2005; Gutiérrez, 2013; Zhang, 2014) also have showed different factor loadings for grammatical versus ungrammatical sentences on the TGJT, claiming that grammatical and ungrammatical sentences in the same test may correlate with different test constructs. Unlike the current study, however, they found the ungrammatical items to load relatively heavily on the explicit knowledge construct. One explanation for these contradictory findings could be differences in modality and time constraints. The previous studies utilized only written TGJTs and allowed more time (6–9 seconds) for each sentence than the $TGJT_{AURAL}$ in the present study (.37–2.99 seconds). Greater time pressure and the nature of aural processing (focusing more on meaning) could have led the learners in this study to rely largely on their implicit knowledge to detect ungrammaticality.

Three-Factor Versus Two-Factor Solutions in Principal Component Factor Analysis

Unlike the two-factor solution identified in the previous studies (e.g., Bowles, 2011; Ellis, 2005; Spada et al., 2015), the criteria (eigenvalue, variance, and scree plot) of the present study yielded a three-factor solution. The higher number of component factors found in this study can be attributed to the different levels of correlation of the two different TGJTs with the OEITs. For example, the correlation coefficients between the $TGJT_{WRITTEN}$ and the $OEIT_{UNCTRL}$ and $OEIT_{CTRL}$ were relatively lower ($r = .363$, $p < .01$ and $r = .383$, $p < .01$, respectively) than those reported in previous studies (e.g., correlation coefficients between the $TGJT_{WRITTEN}$ and the OEIT: $r = .58$, $p < .01$ in Ellis, 2005; $r = .64$, $p < .01$ in Bowles, 2011; and $r = .42$, $p < .002$ in Spada et al., 2015). By contrast, the correlation coefficients between the $TGJT_{AURAL}$ and the $OEIT_{UNCTRL}$ and $OEIT_{CTRL}$ ($r = .517$, $p < .01$ and $r = .521$, $p < .01$, respectively) were similarly high. In this study, the difference in the two TGJTs with reference to the OEITs allowed the identification of more underlying latent factors, which may have enabled the more elaborate explanation (i.e., the three-factor solution) of the variance of the test results.

CONCLUSIONS

This study had two main goals. The first was to evaluate the construct validity of the OEIT and the TGJT as measures of L2 implicit knowledge by identifying the different strengths of implicit knowledge that each type of test draws on. The second was to provide information that would enable the refining of administration methods for these two tests.

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

Although the current study confirmed that both the OEIT and the TGJT are reliable and valid measures of implicit knowledge, it showed that the OEIT taps into more advanced and stronger implicit knowledge than the TGJT. The differential accessibility of implicit knowledge in the two tests is due to the different degrees of automatic processing and depth of processing that they require.

As for the effects of test administration methods, the present study found that time pressure methods for the OEIT (e.g., uncontrolled vs. controlled) and modality for the TGJT (written vs. aural) significantly affected L2 learners' performance. The learners scored significantly higher when the time pressure was controlled rather than uncontrolled in the OEIT and when the stimuli were presented visually rather than aurally in the TGJT. In addition, the results from a series of factor analyses showed that the TGJT with aural stimuli and the OEIT with controlled time pressure tapped into more advanced, stronger implicit knowledge than their counterparts. This is explained by the fact that aural modes and fixed time constraints provide more optimal conditions for inhibiting learners' use of metalinguistic knowledge, which allows learners to maintain their primary attention to L2 meaning, rather than form.

The study's findings suggest that using multiple measurements for implicit knowledge is more useful for fine-grained analysis of the developmental status of learners' implicit knowledge than using a single measurement. As Bowles (2011) suggested, one difficulty in measuring L2 learners' implicit and explicit knowledge is that "interlanguage contains both types of knowledge, making measurement tricky" (p. 248). Because learners inevitably access different types of knowledge during L2 processing, establishing valid measures for implicit knowledge is particularly challenging. Although no test can completely block learners' access to explicit knowledge, specified administration methods and consistent implementation procedures could at least ensure test reliability, which in turn would contribute to both internal and external research validity. In this regard, the present study makes a potentially significant contribution to refining widely used measurements of L2 implicit knowledge and ensuring the construct validity of these tests.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit https://doi.org/10.1017/S0272263115000510.

Received 18 May 2015
Accepted 1 December 2015
Final Version Received 30 December 2015

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

NOTES

1. Strength of implicit knowledge can also be understood in a connectionist model as the strength of connections among units (Rumelhart, Hinton, & Williams, 1986). Connectionism sees language learning as acquiring associations among linguistic systems (e.g., forms and meanings). Networks connect processing units in each system to those in other systems, becoming stronger as learning proceeds. According to Rumelhart and McClelland (1987), implicit knowledge of language may be stored as strengths of connections among processing units.

2. An anonymous *SSLA* reviewer noticed that the comprehension question containing *slow cars* could have cued participants to the correct target form, *fast cars*. Two stimuli sentences had this issue: (1) **Young women like cigarettes and fast car*; *do young women like slow cars?* ($OEIT_{UNCTRL}$); (2) *People in love usually want getting married as soon as possible*; *do they want to get married soon?* ($OEIT_{CTRL}$). While this is a limitation of the materials, an analysis excluding these items ($OEIT_{UNCTRL}$ original: $M = 64.66$, $SD = 15.62$; postexclusion: $M = 65.72$, $SD = 23.92$ and $OEIT_{CTRL}$ original: $M = 44.58$, $SD = 20.79$; postexclusion: $M = 43.56$, $SD = 24.22$) found no significant difference between the mean scores ($OEIT_{UNCTRL}$: $t(130) = .3014$, $p > .05$; $OEIT_{CTRL}$: $t(130) = .2596$, $p > .05$), indicating that exposure to the correct form had no or minimal effect.

3. An anonymous *SSLA* reviewer pointed out that explicitly instructing the participants to reproduce the target sentences in *correct* English might draw their attention to formal correctness, thus creating a potential obstacle to eliciting implicit knowledge. Whether OEIT instructions should be explicit is frequently debated; no study to date has offered conclusive evidence supporting one or the other position. Although acknowledging this potential limitation, this study used explicit instruction to maintain comparability with previously published studies.

REFERENCES

Akakura, M. (2012). Evaluating the effectiveness of explicit instruction on implicit and explicit L2 knowledge. *Language Teaching Research*, *16*, 1–29.

Anderson, J. (1980). *Cognitive psychology and its implications*. New York: W. H. Freeman.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, *8*, 47–89.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple component model. In A. Miyake & P. Shah (Eds.), *Models of working memory* (pp. 28–61). New York: Cambridge University Press.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The cross-linguistic study of sentence processing* (pp. 77–117). Cambridge, UK: Cambridge University Press.

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

Bowles, M. A. (2011). Measuring implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition*, *33*, 247–271.

Brown, G. (1990). *Listening to spoken English*. London: Longman.

Burns, G. (1951). An investigation into the extent of first-year vocabulary in French in boys' grammar school. *British Journal of Educational Psychology*, *21*, 36–44.

Clifton, C., Jr., Meyer, A. S., Wurm, L. H., & Treiman, R. (2013). Language comprehension and production. In A. F. Healy & R. W. Proctor (Eds.), *Handbook of psychology* (pp. 523–547). New York: Wiley.

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

Craik, F., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.

Danks, J. (1980). Comprehension in listening and reading: Same or different? In J. Danks & K. Pezdek (Eds.), *Reading and understanding* (pp. 1–39). Newark, DE: International Reading Association.

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 312–348). Oxford, UK: Blackwell.

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

DeKeyser, R. M. (2009). Cognitive-psychological processes in second language learning. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 119–138). Oxford, UK: Wiley-Blackwell.
Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, *54*, 227–275.
Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, *27*, 141–172.
Ellis, R. (2006). Modelling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge. *Applied Linguistics*, *27*, 431–463.
Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, *28*, 339–368.
Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics*, *27*, 464–491.
Erlam, R., & Loewen, S. (2010). Implicit and explicit recasts in L2 oral French interaction. *Canadian Modern Language Review*, *66*, 877–905.
Eyckmans, J., Van de Velde, H., Van Hout, R., & Boers, F. (2007). Learners' response behaviour in yes/no vocabulary tests. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 59–76). Cambridge, UK: Cambridge University Press.
Field, A. (2009). *Discovering statistics using SPSS*. London: Sage.
Fitzpatrick, T. (2007). Word association patterns: Unpacking the assumptions. *International Journal of Applied Linguistics*, *17*, 319–331.
Graham, C. R., McGhee, J., & Millard, B. (2010). The role of lexical choice in elicited imitation item difficulty. In M. T. Prior et al. (Eds.), *Selected proceedings of the 2008 Second Language Research Forum* (pp. 57–72). Somerville, MA: Cascadilla Proceedings Project.
Gutiérrez, X. (2013). The construct validity of grammaticality judgment tests as measures of implicit and explicit knowledge. *Studies in Second Language Acquisition*, *35*, 423–449.
Hulstijn, J. H., & Hulstijn, W. (1984). Grammatical errors as a function of processing constraints and explicit knowledge. *Language Learning*, *34*, 23–43.
Izumi, S. (2003). Processing difficulty in comprehension and production of relative clauses by learners of English as a second language. *Language Learning*, *53*, 285–323.
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.
Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60–99.
Krashen, S. (1985). *The input hypothesis: Issues and implications*. New York: Longman.
Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*, 399–436.
Leow, R. P. (1993). To simplify or not to simplify: A look at intake. *Studies in Second Language Acquisition*, *15*, 333–355.
Leow, R. P. (1995). Modality and intake in second language acquisition. *Studies in Second Language Acquisition*, *17*, 79–89.
Loewen, S., & Erlam, R. (2006). Corrective feedback in the chatroom: An experimental study. *Computer Assisted Language Learning*, *19*, 1–14.
Lund, R. J. (1991). A comparison of second language listening and reading comprehension. *The Modern Language Journal*, *75*, 196–204.
McLaughlin, B. (1987). *Theories of second language acquisition*. London: Edward Arnold.
McLaughlin, B., Rossman, T., & McLeod, B. (1983). Second language learning: An information processing perspective. *Language Learning*, *33*, 135–158.
Monteiro, K. (2014). An experimental study of corrective feedback during video-conferencing. *Language Learning & Technology*, *18*, 56–79.
Murphy, V. (1997). The effect of modality on a grammaticality judgment task. *Second Language Research*, *13*, 34–65.
Rosseel, Y. (Ed.). (2011). *Lavaan: An R package for structural equation modeling and more*. Retrieved from http://www.webcitation.org/68OFlEhit.
Rost, M. (1990). *Listening in language learning*. London: Longman.

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press

Rumelhart, D. E., & MacClelland, J. L. (1987). Learning the past tense of English verbs: Implicit rules or parallel distributed processes? In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Lawrence Erlbaum.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations* (pp. 318–362). Cambridge, MA: MIT Press.

Schwartz, B. D. (1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition*, *15*, 147–163.

Soleimani, H., Jahangiri, K., & Gohar, M. J. (2015). Effect of explicit and implicit instruction on implicit knowledge of English past simple tense. *International Journal of Asian Social Science*, *5*, 257–265.

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

Sticht, T. G., & James, H. (1984). Listening and reading. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 293–317). London: Longman.

Stoddard, G. D. (1929). An experiment in verbal learning. *Journal of Educational Psychology*, *20*, 452–457.

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

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

Waring, R. (1997). A study of receptive and productive vocabulary learning from word cards. *Studies in Foreign Languages and Literature*, *21*, 94–114.

Wong, W. (2001). Modality and attention to meaning and form in the input. *Studies in Second Language Acquisition*, *23*, 345–368.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, *24*, 1–27.

Zhang, R. (2014). Measuring university-level L2 learners' implicit and explicit knowledge. *Studies in Second Language Acquisition*, *37*, 1–30.

https://doi.org/10.1017/S0272263115000510 Published online by Cambridge University Press