SSLA, **29**, 119–126. Printed in the United States of America.
DOI: 10.1017/S0272263107070052

CONFIRMING THE OPERATIONAL DEFINITIONS OF EXPLICIT AND IMPLICIT KNOWLEDGE IN ELLIS (2005)

Responding to Isemonger

Rod Ellis and Shawn Loewen
University of Auckland

Ellis (2005) and his coresearchers developed a number of tests with a view to providing relatively separate measures of explicit and implicit knowledge. The aim in the development of these tests was to resolve a continuing problem in SLA studies—namely the construct validity of tests used to measure acquisition—and, more specifically, to provide a basis for investigating the relationship between implicit and explicit knowledge (i.e., the strong interface, the weak interface, and the noninterface positions).

The procedure followed was first to provide notional definitions of these constructs and then operational definitions. On the basis of the latter, five tests were designed: The first three were intended to provide measures of implicit knowledge, and the second two were measures of explicit knowledge. For the reader's convenience, the five tests are described in Table 1, together with measures of reliability. These tests were administered to 91 second language (L2) learners (83 in the case of the oral narrative test). These learners were of mixed English language proficiency. A detailed account of the methodology of the study is available in Ellis (2005).

Ellis elected to analyze the data using an exploratory principal components factor analysis,[1] requesting a two-factor solution in accordance with the purpose of the study and the design of the tests. Such exploratory factor analyses have been widely used in applied linguistic studies. For example, Schmidt, Boraie, and Kassabgy (1996) reported a study of the motivational components of 1,554 adult learners of English as a foreign language (EFL) at the American University of Cairo, based on a questionnaire designed to examine eight different motivational factors. They employed an exploratory principal compo-

Address correspondence to: Rod Ellis, Department of Applied Language Studies, The University of Auckland, Private Bag 92019, Auckland, New Zealand; e-mail: r.ellis@auckland.ac.nz.

© 2007 Cambridge University Press 0272-2631/07 $12.00

https://doi.org/10.1017/S0272263107070052 Published online by Cambridge University Press

Table 1. Description of the five tests

Test	Description	Reliability
Oral imitation test	A total of 34 sentences targeting 17 linguistic structures (2 test items per structure: 1 grammatical and 1 ungrammatical). Test-takers listened to each sentence and first indicated whether they agreed or disagreed with the proposition it expressed. They then attempted to repeat the sentence correctly and were audio-recorded. Scoring was based on whether learners successfully repeated/corrected the target structure in each sentence. A percentage accuracy score was calculated.	α = .88 (n = 91)
Oral narrative test	Test-takers read a story twice. The test-takers were then asked to retell the story orally in 3 min. Their narratives were audio-recorded and subsequently transcribed. An obligatory occasion analysis was carried out to establish the percentage of correct suppliance of target structures: regular past tense, modal verbs, third-person -s, plural -s, indefinite article, and possessive -s. Total scores were calculated by averaging the percentage scores for each structure.	r = .85 (interrater agreement) (n = 83)
Timed GJT: total score	This was a computer-delivered test consisting of 68 sentences (4 per structure: 2 grammatical and 2 ungrammatical). Test-takers were required to indicate whether each sentence was grammatical or ungrammatical by pressing response buttons within a fixed time limit. The time limit for each sentence was based on piloting with native speakers. Each item was scored dichotomously as correct or incorrect (based on the responses of native speakers), with items not responded to scored as incorrect. A total percentage accuracy score was calculated.	α = .96 (n = 91)
Untimed GJT: ungrammatical sentences score	This was a computer-delivered test consisting of 68 sentences (4 per structure: 2 grammatical and 2 ungrammatical). Test-takers were required to indicate in their own time whether each sentence was grammatical or ungrammatical. This test provided a percentage judgment accuracy score based on the participants' dichotomous responses. Total accuracy scores as well as separate scores for the grammatical and ungrammatical sentences were calculated. Only ungrammatical scores were included in the analyses.[a]	α = .83 (n = 91)
Metalinguistic test	This consisted of an untimed computerized multiple-choice test in two parts. Part 1 presented test-takers with 17 ungrammatical sentences, based on the 17 structures, and required them to select the rule that best explained each error out of four choices provided. Part 2 consisted of two sections. In section 1, the test-takers were asked to read a short text and then to find examples of 21 specific grammatical features from the text (e.g., preposition and finite verb). In section 2, they were asked to identify the named grammatical parts in a set of sentences. A total percentage accuracy score was calculated.	α = .90 (n = 91)

Note. GJT = grammaticality judgment test.

[a] Ellis (2005) argued on theoretical grounds that in an untimed GJT, ungrammatical sentences provide a clearer measure of explicit knowledge. This was borne out by the analyses reported in Ellis's article.

https://doi.org/10.1017/S0272263107070052 Published online by Cambridge University Press

nent analysis with varimax rotation to extract eight factors. Another example is Skehan and Foster's (1997) study of measures of fluency, complexity, and accuracy. This study also employed an exploratory principal components factor analysis to demonstrate that the nine measures that they examined successfully distinguished the three constructs. A third example is Wintergerst, DeCapua, and Itzen's (2001) study that investigated the construct validity of Reid's (1987) perceptual learning style questionnaire. This study also used an exploratory factor analysis.

In his critique of Ellis's (2005) study, Isemonger (this issue) argued that (a) assuming that an exploratory factor analysis is appropriate, an oblique rather than an orthogonal rotation was required, as the assumption was that the factors would be correlated, and, in any case, (b) given that Ellis's study had a verificational rather than an exploratory agenda (i.e., he was testing an operationalization of a particular model of L2 knowledge), a confirmatory factor analysis should have been employed. Although we feel that Isemonger has taken little cognizance of past practice in applied linguistics and SLA (as reflected in the studies mentioned in the previous paragraph),[2] we accept his criticisms and welcome the opportunity to report the analyses that he (and our own statistical advisors) points out are needed. We will demonstrate that these analyses replicate the results of the analysis reported in Ellis (2005).

PRINCIPAL COMPONENT ANALYSIS WITH OBLIQUE ROTATION

The principal component analysis that we originally conducted employed a varimax orthogonal rotation.[3] The results of the same principal components factor analysis with an oblique rotation are presented in Table 1. As with the rotated solution in Ellis (2005), the oral imitation test, the oral narrative test, and the timed grammaticality judgment test (GJT) load on factor 1, whereas the untimed GJT (ungrammatical) and the metalinguistic knowledge test load on factor 2. In other words, the results of this analysis replicate those of the original analysis. In fact, in some respects they provide even stronger support for Ellis's claim that the two sets of tests provide relatively separate measures of implicit and explicit knowledge, in that the variables load strongly on only one factor, with virtually no loading on the other factor. For example, the oral imitation test's loading on factor 2 in the rotated solution was .253, whereas in the oblique solution, the loading on factor 2 was only .067.

Isemonger also criticized Ellis's (2005) study for the fact that the eigenvalue of the second factor was below one, citing several studies that claimed that even solutions using eigenvalues equal to or greater than one might overestimate the number of factors that can reliably be extracted. We draw the reader's attention to the facts that the eigenvalue of factor 2 in the oblique solution is very close to one and that an inspection of the scree plot clearly justified a two-factor solution. Also, we note that there is divergence of opinion among statisticians regarding the basis for selecting the number of fac-

https://doi.org/10.1017/S0272263107070052 Published online by Cambridge University Press

Table 2. Factor analysis communalities

Test	Initial	Extraction (two factors)	Extraction (one factor)
Imitation	1.000	.744	.630
Oral storytelling	1.000	.653	.452
TGJT total	1.000	.646	.563
UGJT ungrammatical	1.000	.825	.581
Metalinguistic	1.000	.833	.493

Note. Extraction method: principal component analysis.

tors. For example, some studies have argued that an eigenvalue equal to or greater than one is too strict and, furthermore, argued for retaining all factors with eigenvalues greater than 0.7 (Jolliffe, 1972, 1986; cited in Field, 2005). Finally, Field recommended an inspection of the communalities after extraction, with values close to 1.0 being desirable. Table 2 shows that the communalities are considerably higher for the two-factor solution than for the one-factor solution.

The fact that each variable loads primarily on only one factor, as shown in Table 3, arguably removes the need to conduct a confirmatory factor analysis. However, given the doubts that Isemonger has expressed regarding the low eigenvalue of factor 2 as well as his insistence on the need to test alternative solutions, we will now present the results of such an analysis.

CONFIRMATORY FACTOR ANALYSIS

Two confirmatory factor analyses were carried out using AMOS 5.0 (Arbuckle, 2004). The first analysis tested the construct solution predicted by the design of the tests (i.e., that the oral imitation test, the oral narrative test, and the

Table 3. Principal component analysis of the five tests (direct oblimin)

	Component 1	Component 2
Eigenvalue	2.719	.982
% Variance	54.389	19.637
Cumulative %	54.389	74.026
Test		
Imitation	.832	.067
Oral narrative	.847	−.103
Timed GJT	.759	.093
Untimed GJT (ungramm.)	.074	.874
Metalinguistic knowledge	−.045	.932

https://doi.org/10.1017/S0272263107070052 Published online by Cambridge University Press

timed GJT would load on an implicit factor and that the untimed GJT [ungrammatical] and the metalinguistic test would load on an explicit factor). The second analysis tested the alternative method solution proposed by Isemonger (i.e., that the oral imitation test and the oral narrative test would load on a production factor, whereas the timed GJT, the untimed GJT [ungrammatical], and the metalinguistic test would all load on a decision factor). The results are shown in Figures 1 and 2.

Table 4 provides a summary of the model fit for both solutions, with the χ^2 statistic, testing the goodness of fit, presented. A significant value indicates that the model is statistically unlikely to occur, whereas a nonsignificant value indicates an acceptable model (Byrne, 2001). In the current analysis, the χ^2 for the implicit/explicit model is not significant; however, the χ^2 value for the decision/production model is significant. Additionally, Byrne (2001) noted that a normed fit index value of greater than .95 indicates a superior fit and that a root mean square error of approximation value of less than .05 indicates a good fit for the model. Thus, according to these indicators, the implicit/explicit model is an acceptable model, whereas the decision/production model is not.

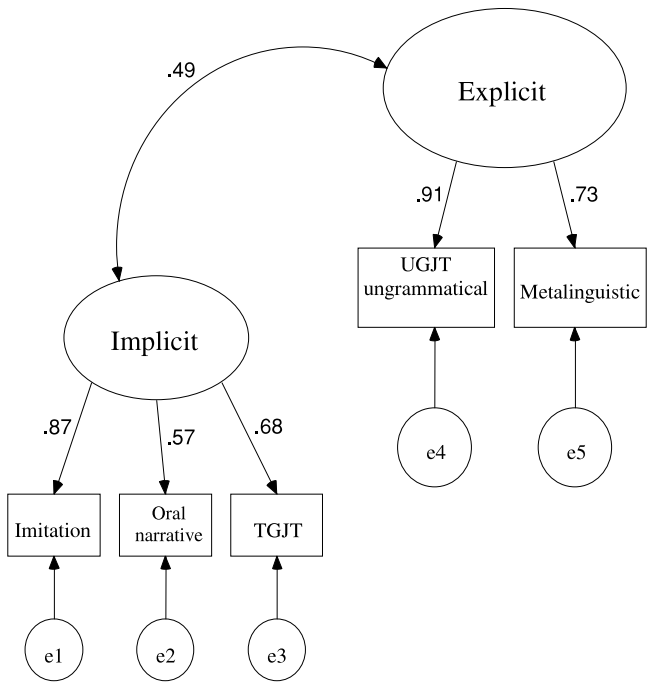


Figure 1. Implicit/explicit model.

https://doi.org/10.1017/S0272263107070052 Published online by Cambridge University Press

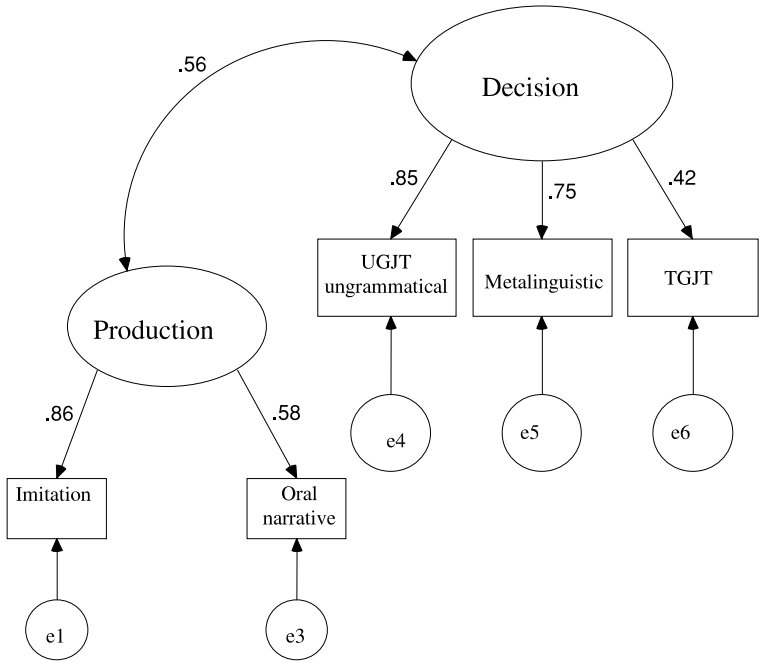


Figure 2. Decision/production model.

OPERATIONALIZING IMPLICIT AND EXPLICIT KNOWLEDGE

Isemonger also addressed our operationalizations of implicit and explicit knowledge. He wondered how the timed GJT could provide a measure of implicit knowledge given that the test invites a judgment of the correctness of sentences—knowledge that he considers to be metalinguistic. Here, Isemonger seems to have missed the point. Speakers of a language are perfectly able to decide whether a particular usage is grammatical without any explicit knowledge of the rule or feature involved. Indeed, the plethora of studies that have utilized GJTs have been based on precisely this assumption. The rationale for

Table 4. Summary of the model of fit for both solutions

Model	χ^2	NFI	RMSEA	*df*
Implicit/explicit	1.191	.991	.000	4
Decision/production	29.150**	.784	.259	4

Note. NFI = normed fit index; RMSEA = root mean square error of approximation.
** $p < .00$.

https://doi.org/10.1017/S0272263107070052 Published online by Cambridge University Press

the timed GJT was that learners would only be able to judge the sentences on the basis of their implicit knowledge because the speed of the response required precluded access to their explicit knowledge.

Second, drawing on Ellis (2004), Isemonger noted that explicit knowledge (as analyzed knowledge) exists independently of a learner's knowledge of metalingual terms, and he then argued that the two operationalizations of explicit knowledge (the ungrammatical sentences in the untimed GJT and the metalinguistic knowledge test) failed to differentiate these two aspects of explicit knowledge. In fact, the untimed GJT was intended as a measure of analyzed knowledge, whereas the metalinguistic knowledge test was designed to measure learners' knowledge of metalanguage. Isemonger's problem with this rests in the fact that neither of these tests measures potentially verbalizable explicit knowledge. Fairly obviously, however, no test is capable of measuring what is potentially verbalizable, but only what is verbalizable. We chose not to include a test that required learners to verbalize rules for two reasons. First, from a practical standpoint, it was easier to assess their receptive knowledge of metalingual terms. Second, when asked to verbalize a rule, learners might construct ad hoc explanations at the time (i.e., they might create explicit knowledge rather than draw on their existing explicit knowledge). In short, whereas it might be helpful to conceptualize explicit knowledge as potentially verbalizable, it might not be possible to operationalize such a notion for testing except in terms of learners' knowledge of metalingual terms.

CONCLUSION

In his critique of Ellis (2005), Isemonger also offered a list of suggestions for avoiding the psychometric shortfalls in SLA studies that both Ellis (2005) and Douglas (2001) warned against. We find this a useful list—one to which all SLA researchers who examine relationships among quantitative datasets should attend. We recognize that Ellis's study was lacking with respect to a number of the points on this list. However, we feel that the soundness of the general conclusion that Ellis reached—namely, that the tests he examined did succeed in providing relatively separate measures of L2 implicit and explicit knowledge—was demonstrated. This is even more clearly demonstrated by the additional analyses presented in this response to Isemonger.

One of Isemonger's suggestions is that higher order models be investigated. In the case of the data available from Ellis's (2005) study, this would be possible (as Isemonger pointed out) by disaggregating the test scores. For example, it would be possible to enter separate scores for the grammatical and ungrammatical sentences in the GJTs and for the three parts of the metalinguistic test. This constitutes an avenue for further research based on the same datasets that we look forward to exploring in the future.

(*Received 30 May 2006*)

https://doi.org/10.1017/S0272263107070052 Published online by Cambridge University Press

NOTES

1. One point not raised by Isemonger (this issue) concerns the use of a principal components analysis versus a factor analysis. There is some disagreement among statisticians regarding the preferred analysis (Field, 2005), and in order to avoid further criticism, a maximum likelihood factor analysis was also performed. The results of the obliquely rotated pattern matrix did not differ from the principal components analysis presented in this article.

2. It is interesting to observe that Isemonger (this issue) cited Wintergerst, DeCapua, and Itzen (2001) in a note. His purpose in doing so was to claim that psychometric research examining Reid's perceptual learning style questionnaire has questioned its validity. We found ourselves wondering how Isemonger could cite the Wintergerst et al. study in this respect, given that it would seem subject to the same criticisms he has leveled at our use of factor analysis.

3. Isemonger (this issue) rightly pointed out that Ellis (2005) neglected to state whether the principal component analysis used in his study was rotated and went on to assume that the solution Ellis presented was not rotated. This assumption is incorrect. It should also be noted that the original manuscript sent to the editor of *Studies in Second Language Acquisition* explicitly indicated that varimax rotation was employed. Unfortunately, this information was deleted in the publishing process. R. Ellis takes complete responsibility for this error.

REFERENCES

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

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

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

Ellis, R. (2004). The definition and measurement of 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.

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). Thousand Oaks, CA: Sage.

Jolliffe, I. T. (1972). Discarding variables in a principal component analysis I: Artificial data. *Applied Statistics*, *21*, 160–173.

Jolliffe, I. T. (1986). *Principal component analysis*. New York: Springer-Verlag.

Reid, J. (1987). The learning style preferences of ESL students. *TESOL Quarterly*, *21*, 87–111.

Schmidt, R., Boraie, D., & Kassabgy, O. (1996). Foreign language motivation: Internal structure and external connections. In R. Oxford (Ed.), *Language learning motivation: Pathways to the new century* (pp. 9–70). University of Hawai'i at Manoa: Second Language Teaching & Curriculum Center.

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, *1*, 185–211.

Wintergerst, A., DeCapua, A., & Itzen, R. (2001). The construct validity of one learning style instrument. *System*, *29*, 385–403.

https://doi.org/10.1017/S0272263107070052 Published online by Cambridge University Press