Development and Validation of a Diagnostic Grammar Test

Running head: Development and Validation of a Diagnostic Grammar Test

Title
Development and Validation of a Diagnostic Grammar Test for Japanese Learners of English

Rie Koizumi
Tokiwa University

Hideki Sakai
Shinshu University

Takahiro Ido
Waseda Junior & Senior High School

Hiroshi Ota
Komazawa Women’s University

Megumi Hayama
Dokkyo University

Masatoshi Sato
Fuefuki High School, Yamanashi

Akiko Nemoto
Atago Junior High School, Ibaraki

Corresponding author:
Rie KOIZUMI, Ph.D
Tokiwa University
College of Applied International Studies
1-430-1 Miwa, Mito, Ibaraki, 310-8585, Japan
e-mail: koizumir@tokiwa.ac.jp
Phone: +81-29-232-2791
Abstract
This paper reports on the development and validation of the English Diagnostic Test of Grammar (EDiT Grammar) for Japanese learners of English. From among the many aspects of grammar, this test focuses on the knowledge of basic English noun phrases (NPs), especially their internal structures, because previous research has indicated the difficulty faced by Japanese learners of English in acquiring these phrases. The results of the examination of the pilot and revised tests suggest that the revised test is generally appropriate in terms of item discrimination, distractor function, reliability, and the difficulty of NP groups. Overall, the EDiT Grammar provided favorable evidence for the validity in interpreting the test scores in the case of a fairly low-stakes diagnostic test.

Keywords
diagnostic test, grammar, noun phrases, validation

Acknowledgement
We are deeply indebted to ELPA, especially Etsuro Ikawa, who granted us permission to publish the results. We also appreciate Yo In’nami and two anonymous reviewers for their constructive comments on our paper.

This is an electronic version of an article published in the Language Assessment Quarterly, Volume 8, Issue 1, 2011 © Taylor & Francis. Language Assessment Quarterly is available online at: www.tandfonline.com with the open access of our article: http://www.tandfonline.com/doi/abs/10.1080/15434303.2010.536868 (see the following, for the permission to post authors’ accepted manuscripts: http://journalauthors.tandf.co.uk/permissions/reusingOwnWork.asp )

Development and Validation of a Diagnostic Grammar Test
Development and Validation of a Diagnostic Grammar Test for Japanese Learners of English

Diagnostic tests have long been recognized as essential for teaching and learning. A diagnostic test is “used to identify test taker’s strengths and weaknesses, by testing what they know or do not know in a language, or what skills they have or do not have” (Davies, Brown, Elder, Hill, Lumley, & McNamara, 1999, p. 43). As several researchers have indicated (e.g., Alderson, 2005; Davies et al., 1999), despite the importance of using diagnostic tests to improve the quality of learning, teaching, and assessment, there appears to be only a limited number of tests available to the public that are specifically for diagnostic purposes. The scarcity of diagnostic tests has arisen mainly because it is rather difficult and time-consuming to construct such tests (Davies et al., 1999) and because decisions based on these tests do not seem to have a large impact on the lives of those taking them, even if the diagnosis is not appropriate (Alderson, 2005). Diagnostic tests, however, have a significant educational value (Alderson, 2005) and are worth constructing despite the high cost involved.

The lack of and need for diagnostic tests for Japanese learners of English was identified by the Association for English Language Proficiency Assessment (ELPA). ELPA is a Japanese non-profit organization that develops and delivers tests in order to objectively assess the outcome of English language education in Japan (ELPA, n.d.). ELPA initiated a series of projects to develop a diagnostic test of grammar, vocabulary, and phonetics for classroom use (Kanatani & EDiT Development Group, 2006). This paper reports on some of the projects in relation to the development and validation of a diagnostic grammar test for Japanese learners of English. In light of the paucity of diagnostic tests and established procedures available so far for effective diagnosis, we will present one approach to addressing this problem.

EDiT Grammar

Internal Structures of Noun Phrase

The English Diagnostic Test of Grammar (EDiT Grammar) is a diagnostic grammar test that mainly targets Japanese secondary school students who have
studied English as a foreign language for more than two years (i.e., 9th- to 12th-year students). The purpose of this test is to identify the grammatical weaknesses of learners that cannot easily be detected by teachers during teaching activities; the test also aims to present information that will be useful for teaching in the future.

From among the many aspects of grammatical knowledge (Purpura, 2004), we selected knowledge of morphosyntactic form and meaning. The test aims to assess mainly morphosyntactic form, especially word order, of basic English noun phrases (NPs). However, morphosyntactic meaning is also indirectly tested since meaning needs to be considered in order to successfully select the appropriate word order.

The knowledge of basic NPs consists of two elements: NP boundaries and internal structures. For example, in the sentence *Yesterday the window near Ken’s desk was broken*, readers need to (a) locate the meaning chunk *the window near Ken’s desk* and the existence of boundaries before *the* and after *desk*, and (b) comprehend that *near Ken’s desk* modifies *the window*, in order to understand the precise sentential meaning. Eventually, we will aim to test the knowledge of both (a) and (b); in order to obtain more detailed diagnosis, however, we currently focus on (b), namely the knowledge of NP internal structures.

We concentrate on the knowledge of NP structures because previous research has found that they seem to be difficult to acquire by Japanese learners of English (e.g., Kimura & Kanatani, 2006; Kimura, Kanatani, & Kobayashi, 2010). To the authors’ knowledge, there seems to be no studies outside Japan dealing with the relative difficulty of acquiring NP internal structures, except for those examining the acquisition of different kinds of relative clauses (see Ellis, 2008), which are not very relevant to our study.

A lack of NP knowledge can lead to difficulty in comprehending and producing sentences in which NP structures are embedded. For example, learners lacking such knowledge may fail to grasp precise sentential meanings—especially those with complex, embedded NP structures—and this, in turn, may hamper reading comprehension (Nuttall, 2005).

Difficulties with NPs for Japanese learners of English may arise for two main reasons. First, there is a difference between the Japanese and English
languages in terms of the order of modifying relationships. In English, both premodification (e.g., my bag and that sleeping dog) and postmodification (e.g., the boy running over there and the house built 100 years ago) are used (Nuttall, 2005), but in Japanese premodification is mostly employed (Ogasawara, 1965). For example, the sentence Look at the book on the desk is translated as tsukue [the desk] + noueniaru [on] + hon [the book] + wo [at] + mite [look]. In English, the book is placed first, followed by a modifier on the desk, but in Japanese, the order is reversed: the desk appears before on, followed by book. Due to the lack of postmodification in Japanese, Japanese learners find it more difficult to grasp postmodification than premodification (Hashimoto & Hirai, 2007; Kimura et al., 2010), and they tend to fail to distinguish the box on the cat from the cat on the box. Kimura and Ota (2006) reported that Japanese secondary school students tend to make errors in distinguishing correct NPs from distractors that conform to the word orders of Japanese.

Another reason for the difficulty in understanding NPs is that despite the difference in word order between English and Japanese, NP internal structure has not been systematically taught in Japan (Kimura & Ota, 2006). Katsuta (2010) reported that the difficulty of NPs with postmodifiers (such as phrases with prepositions and present/past participles) does not reflect well in the differences between textbooks at the novice and intermediate levels and at the advanced level; the exception was NPs with relative clauses, which appeared more frequently at the advanced level. Since textbooks are usually written by English teachers in Japan, the result seems to indicate that the teachers fail to notice the difficulty and that they provide necessary instruction for all postmodification except that which concerns relative clauses. Due to the lack of explicit instruction on NPs, both students and teachers seem to remain unaware of this problematic area. Terada (2009) argued that more instruction vis-à-vis NP internal structures should be provided in Japan, as should that of the relationships between NPs. Nuttall (2005) mentioned the difficulties inherent in understanding NPs as well as the need to teach NPs to L2 learners in general.

**Noun Phrase Groups**

In order to operationalize knowledge pertaining to NP internal structures,
we categorized NPs into five NP groups according to their hypothesized internal complexity (see Table 1). NP Groups 1 and 2 consist of NPs with premodifiers, whereas NP Groups 3 to 5 include NPs with postmodifiers. All the grammatical structures in this table are supposed to be taught at the lower secondary school level. This is specified in the Course of Study, the government guideline for teachers at Japanese secondary schools (Ministry of Education, 1999). Among the five NP groups, the EDiT Grammar intends to assess NP Groups 2 to 5, because the structure of Group 1 is easily mastered by students, and students can be diagnosed as being at the Group 1 level if they do not reach the Group 2 level.

[Insert Table 1 about here]

We tentatively postulated the difficulty order of the NP Groups in ascending order of difficulty (Groups 1 < 2 < 3 < 4 < 5), for three reasons. First, we used Pienemann’s (1998) processability theory. Pienemann attempted to explain and predict the order and sequence of morphosyntactic acquisition by postulating that L2 learners need to acquire language-specific processing procedures necessary to produce and comprehend the target structures. For example, in order to process NP Group 1 (determiner + head noun), L2 learners must store and access lexical items (e.g., *two* and *sweaters*) in their lexicon. They also need to identify the grammatical categorization of these lexical items (e.g., quantifier and noun). Further, they have to realize diacritic features (e.g., plural) and combine the lexical items into a phrase (e.g., *two sweaters*). At this point, the diacritic feature of *two* (plural) must be matched with that of *sweaters* (plural). In other words, three processing procedures—which Pienemann calls the word/lemma access, the category procedure, and the phrase procedure, respectively—are involved in the process of NP Group 1. Groups 2 and 3 similarly require these three procedures. Group 4 involves the S-procedure, by which the syntactic roles of a sentence (e.g., subject and object) are assigned to phrases. For example, to process the phrase *a couple having a dinner time*, L2 learners need to not only process the constituent phrases (e.g., *a couple* and *a dinner time*) but also realize their syntactic roles (e.g., subject for *a couple* and object for *a dinner time*). Group 5 requires the subordinate procedures, by which L2 learners distinguish
subordinate clauses from their main clauses. Pienemann (1998) hypothesized that these procedures are implicationally related to each other in the above order; in other words, the acquisition of one procedure requires the acquisition of the previous procedures. For example, the acquisition of the word/lemma access and the category procedure is a precondition for the acquisition of the phrasal procedure; similarly, the acquisition of the S-procedure requires the acquisition of the first three procedures. Thus, according to the processability theory, the difficulty of NP groups can be predicted as follows: Groups 1, 2, 3 < 4 < 5.

The second justification of the difficulty prediction of NP groups is provided by Hashimoto and Hirai (2007) and Kimura et al. (2010), who suggest that premodification is easier than postmodification; thus we predicted the order of groups as 1, 2 < 3, 4, 5. A last rationale for the difficulty order is that, due to the more complex structure of Group 2 because of the presence of premodifiers, Group 2 is more difficult than Group 1. The difficulty order of the groups as 1 < 2 < 3 < 4 < 5 was predicted by combining these three justifications.

**Test Format and Diagnosis**

The EDiT Grammar adopts a multiple-choice format of matching a Japanese phrase with one of four English alternatives. Each test item belongs to one of the four NP groups and is dichotomously scored to determine whether test-takers understand the targeted NP group structure. Although several item formats relevant to the domain were initially contrived and examined based on Purpura (2004), only the multiple-choice format was retained, mainly due to its clear advantage of utilizing distractors to identify the difficulties experienced by learners.

One of the distinct characteristics of our diagnostic test lies in its distractors, each of which can indicate learners’ error patterns, or tendencies to misunderstand NP structures. Here is an example.

**Instruction:** Choose the most appropriate English phrase for the underlined part of the Japanese sentence from 1 to 4 (Originally written in Japanese).

**Example item 1 (NP Group 3)**

**Stem:** 丘の上の白い建物は、私の学校です。
Development and Validation of a Diagnostic Grammar Test

*Oka noueno shiroi tatemono wa watashi no gakko desu.*
(The white building on the hill is my school.)

Options:
1. the hill on the white building  (Japanese word order)
2. the white building on the hill*
3. the white on the hill building  (Premodification)
4. on the hill the white building  (Postmodification)

Three distractors have been formulated to indicate the points that test-takers are unable to recognize as errors. For example, if students select Option 3, they may not understand premodification structure, such as the white building, which belongs to NP Group 2; if they choose Option 4, they are likely to lack the understanding of postmodification, with on the hill modifying the white building; if they select Option 1, they may rely on the Japanese word order, because Option 1 follows the Japanese word order: oka [the hill] + noueno [on] + shiroi-tatemono [the white building].

Table 2 displays all five distractor types. In addition to the three mentioned above, there are two other types of distractors: phrase structure and clause structure. Assume that Example item 1 has the following distractor; the choice of this option by test-takers indicates a problem in understanding the phrase structure on the hill; if they select this option, it suggests that they lack knowledge of phrase structures.

5. the hill the white building on  (Phrase)

[Insert Table 2 about here]

The final type, clause structure, is presented in the following example. In Option 3 of Example item 2, the clause *which was built in 2004* is not well structured. If students select this option, it indicates that they lack knowledge of correct clause structures (see Appendix A for more example items).

Example item 2 (NP Group 5)
2004年に建てられた図書館は、とても近代的です。
2004 nen ni taterareta toshokan wa totemo kindaiteki desu.
(A library which was built in 2004 is very modern.)

1. a library which was built in 2004*
2. which was built in 2004 a library (Postmodification)
3. a library was built in 2004 which (Clause)
4. 2004 in was built which a library (Japanese word order)

Some may wonder if deliberately using distractors that learners have difficulty in can entice them into selecting incorrect options. We argue, however, that failing to distinguish incorrect options from answers and consistently choosing one type of option, such as the Japanese word order, suggests that they are likely to have problems with processing correct and distracting structures. In our test, several items belong to the same option type; when test-takers select the same type to a certain proportion (currently 50%), they are diagnosed as having a problem. Additionally, using the common mistakes of learners that are seen in their performances is a recommended technique for creating effective distractors in multiple-choice test construction (Downing, 2006; Hughes, 1987). What makes our approach distinct from common multiple-choice format is the use of distractors for diagnosis.

Others may be concerned that the three distractors in each item represent some of the ways in which test-takers can misunderstand NP internal structures, which we have to admit is the case. The multiple-choice format may affect the process and response patterns by, for example, unintentionally presenting extra information and inviting test-takers to guess at answers. These possibilities of construct underrepresentation and construct-irrelevant variance should be considered in score interpretation; future research should attempt to explore whether other formats can be included.

The development of a method for identifying error patterns by using distractors is still in progress. Nonetheless, we currently calculate the percentage of each distractor type selected by students and we consider 50% or above as indicating a specific error pattern. This percentage is determined by setting two cut-off points (i.e., 50% and 70%) and by asking teachers to judge the relative quality of error pattern diagnosis in comparison with their students’ classroom
performances.

A diagnostic profile (see Table 3) consists of (a) a student’s level of basic grammatical knowledge (Group 1 to 5 levels), (b) the type of NP groups the student has mastered, and (c) an error pattern. The level of basic grammatical knowledge is mainly provided for students, and the rest primarily for use by teachers. (a) The grammatical knowledge level is a numerical expression of (b) the type of NP groups a student has mastered. This level is determined by the NP group with the largest number in which a student responded correctly to 80% or more of the items.

[Insert Table 3 about here]

In the test development stage, we completed the test specifications and wrote items for the EDiT Grammar. This was followed by the iterative process of refining construct definition, test format, and distractor types; we conducted small-scale trials of items in schools and had intensive discussions with secondary school teachers and researchers specializing in SLA and language testing.

We faced two main difficulties in this process. First, as mentioned in Alderson (2005), we failed to find any in-depth theory that enabled us to specify the construct, test format, and error patterns. Second, we experienced difficulty in creating items with appropriate answers and distractors on the basis of the test specifications; this was especially the case when pinpointing adequate words to use in the test. The size of the usable lexicon was limited on account of our strict selection guidelines: We needed to select words that were easy for target learners (so that they would have no difficulty in comprehending the vocabulary) and we needed to avoid providing unintended “hints” for answers. We eventually set up an explicit procedure by which we could select an appropriate vocabulary by using a database and concordances for revised tests (see Revision of the Pilot Test). Still it took collaboration among seven test creators, during an intensive, five-hour session, to choose the vocabulary and write items. The difficult and time-consuming nature of constructing the diagnostic test was evident.

In order to investigate the validity of the inferences of the pilot and revised tests based on the test scores and to provide empirical evidence for the
validity argument based on Chapelle, Enright, and Jamieson (2008), we posited four research questions (RQs).

RQ1: How is item discrimination in the test?
RQ2: To what degree do test-takers select different types of distractors?
RQ3: How reliable are scores on the test? How many items are required to achieve high reliability?
RQ4: Does NP group difficulty increase in the order of the Groups 2 < 3 < 4 < 5?

RQ1 and RQ2 employed item analysis; RQ3 was answered using multivariate generalizability theory; and RQ4 was investigated using analysis of variance (ANOVA). Two studies were conducted: One used a pilot test (Study 1); the other used a revised test (Study 2); each of these studies addressed all the four RQs.

**Study 1: Analyses Using the Pilot Test**

**Study 1 Method**

Participants consisted of 387 ninth-year students attending three public lower secondary schools. While all the participating schools followed the Course of Study (Ministry of Education, 1999), they used different textbooks authorized by the Ministry. The test-takers’ teachers volunteered to administer the EDiT Grammar in order to assess their students’ weaknesses. Their students took the pilot EDiT Grammar between November 2005 and March 2006 at their own schools as part of their English lessons. The pilot test consisted of five or six items for each NP group and 23 items in total (see Kanatani and EDiT Development Group, 2006, for the whole test).

For item discrimination (RQ1), point-biserial correlation coefficients were computed. To compute descriptive statistics, the coefficients were transformed using Fisher’s z and inverse z formulas. For distractor analysis (RQ2), we calculated the percentage of selection in an option type (i.e., the number of options a student selected in an option type / the number of options available in the option type).

For RQ3, the multivariate generalizability theory (Brennan, 2001) was employed with a single-random-facet design (p x i [person by item]) with four
dependent variables (NP Groups 2 to 5) analyzed separately. This was used in order to examine current reliability (Φ) for each NP group and the number of items necessary to achieve high reliability. Phi coefficients (Φ), measurement for an absolute decision, were used for the purpose of interpretation instead of generalizability coefficients, which are for a relative decision, since the EDiT Grammar is a criterion-referenced test. The phi coefficient of above .70 was judged to be high for a low-stakes diagnostic test. The phi lambda (Φ (λ)) was also calculated in order to obtain the agreement of decisions. After comparison between the cut-off points of .60, .70, and .80, we reported only the result of .80, which had the highest agreement.

For RQ4, the proportion correct (the number of correct options selected / the number of all items) was calculated for each student and for each NP group. After testing the assumption of normality, we performed a univariate one-way repeated measures ANOVA to evaluate the effects of NP group (four levels) on the EDiT Grammar scores (one dependent variable). All analyses were conducted using SPSS Version 15.0J (2006) or mGENOVA Version 2.1 (CASMA, 2007).

**Study 1 Results and Discussion**

*Discrimination analysis.* The mean of item discrimination was .45, with an SD of .08 (minimum = .28, maximum = .55; skewness = –.71; kurtosis = .45), indicating that there was a fairly high level of discrimination among target learners.

*Option analysis.* Table 4 provides a descriptive summary of the types of distractors that the test-takers selected when we analyzed (a) all test-takers, (b) those who passed an NP group by obtaining 80% correct or more, and (c) those who failed. The correct options were selected by 51% of all the examinees on average. Options with incorrect postmodifiers and in the Japanese word order were selected by 13% each, which was a little higher than options with split premodification, phrase, and clause structures (7%, 6%, and 9%, respectively). This tendency was more evident among students who failed (e.g., 16% each for incorrect postmodification and Japanese-word-order options). This finding suggests that the difficulties inherent in NP structures derive more from a lack of
Development and Validation of a Diagnostic Grammar Test

postmodification structures in Japanese and the difference between Japanese and English word orders.

[Insert Table 4 about here]

**Reliability analysis.** The reliability coefficients, as presented in Table 5, were rather low, especially for NP Group 2 (.55). It was also found that if the number of items was increased up to 8 for Groups 3 and 4, up to 10 for Group 2, and 6 items are retained for Group 5, reliability of .70 could be achieved. Furthermore, the agreement of decisions at the cut-off point .80 was low for NP Group 2 (.54) but sufficient for the other groups (.73 to .79). The low agreement in NP Group 2 occurred because the criterion of .80 is the same as the mean of Group 2 (see Table 6 and Bachman, 2004). Whether this low agreement is observed after the revision is examined later.

[Insert Table 5 about here]

**Comparisons between NP groups.** Table 6 sets out the descriptive statistics for the four NP groups. A notable result was that the mean proportions correct for students were 80% for NP Group 2 and less than 80% for Groups 3, 4, and 5 (61%, 56%, and 59%, respectively). This suggests that students at the average level can pass only Group 2 and tend to have a poor mastery over the knowledge of basic NP structures, which shows the need for and usefulness of the development of the EDiT Grammar.

[Insert Table 6 about here]

We expected that the NP group means would decrease in the order of Groups 2 > 3 > 4 > 5. We found a significant effect of NP groups on scores with a fairly strong degree, \(F(3, 1158) = 126.96, p < .001, \eta_p^2 = 0.25\). The post-hoc Bonferroni multiple comparisons also showed significant differences, as follows: Groups 2 > 3 > 4. This indicated that the NP group difficulty of the EDiT Grammar increased in the order of Groups 2, 3, and 4. Significant differences were
not found between NP Groups 3 and 5 and between 4 and 5.

In order to determine why Group 4 had an unexpectedly lower mean (as low as Group 5) Group 4 items were divided into three subgroups: present participle, past participle, and to-infinitive. Another ANOVA was conducted to examine the differences between Groups 2, 3, 5, and the three subgroups of Group 4. The analysis showed that there was a significant effect of NP groups with a relatively strong degree, \( F(4.02, 1551.48) = 137.53, p < .001 \) (Greenhouse-Geisser adjustment), \( \eta_p^2 = 0.26 \). The post-hoc tests showed that most groups had significantly different means: Groups 2 > 4 (past participle) > 3 > 5 > 4 (to-infinitive); Group 4 (present participle) showed no differences from Groups 3 and 5. This result seems to indicate that the subgroups of Group 4 have a discrete level of difficulty.

One possible reason why that Group 4 (past participle) turned out to be easier than Group 3 is that some test-takers may have been able to answer correctly, not with the knowledge of NP Group 4 structures, but with the knowledge of chunks. For example, one item in Group 4 (past participle) contained the phrase made in ... and had the high mean and a relatively large SD (\( M = .76, SD = .43 \)). This chunk is a common phrase that has been co-opted into the Japanese language and may have been processed as a familiar chunk by the learners; thus the option a game made in Canada chosen from the four alternatives must have been simple for some students. This finding led us to avoid familiar chunks in test items in constructing the EDiT Grammar. It should be noted that each NP G4 subgroup (e.g., G4, to-infinitive) had only two items, thereby necessitating further analysis.

**Revision of the Pilot Test**

Based on the results of Study 1, we revised the test specifications, with the following four main points added. First, based on the rather low reliability of each NP group, we decided to increase the total number of items. Second, we clarified the procedure for selecting words, hoping to reduce the construct-irrelevant variances. We decided to use simple words and to avoid employing familiar chunks and using words with similar spellings (e.g., tree vs. three) in order to exclude unintended clues as to vocabulary. At the same time, we sought to employ
a variety of words and structures where possible. For this purpose, we developed a database that included three major government-authorized textbooks used for Japanese lower secondary school students (seventh to ninth). Each time we selected words for the test items, we checked if any of the words appeared in the database and if the chunks did NOT appear. Third, we attempted to order items and options for each NP group at random, because the ordering can also serve as a clue that could help some learners. Fourth, we made it mandatory to ask native speakers of English to check the naturalness of the expressions selected (see Appendix B for the revised test structure).

We then constructed a test based on revised specifications with all the items newly constructed, which we hereafter refer to the “revised test.” Although the pilot study contained items that functioned well, we decided to retain them for future item banking and not use them in the revised test. The pilot and revised tests were both based on the same four NP groups and test format, although there were differences in the number of items and the procedures for constructing items. The revised test consisted of 10 or 12 items for each group and 46 items in total. We added more items to Groups 3, 4, and 5 than was indicated by the reliability analysis because additional items could possibly provide the reliable results needed to determine the relative difficulty of the NP Groups. We then investigated the quality of the revised test.

**Study 2: Study Using the Revised Test**

**Study 2 Method**

The test was conducted on 107 ninth-year students in December, 2008. Their proficiency levels are considered to be similar to those in Study 1 since there was no elimination process for the entrance to public schools used for both studies. The test was implemented in a manner similar to Study 1, as part of a normal English lesson, and was conducted by a teacher who was willing to understand the difficulties that her students face (see Kanatani, 2008 for the whole test). In Study 2, the same statistical analyses as Study 1 were conducted.

**Study 2 Results and Discussion**

*Discrimination analysis.* There were items with fairly high
Development and Validation of a Diagnostic Grammar Test

Discrimination with a mean of .45, with an SD of .13 (minimum = .21, maximum = .64; skewness = .08; kurtosis = -.48). Across Studies 1 and 2, we obtained high discrimination values.

Option analysis. As can be seen from Table 7, when we analyzed all the test-takers, the correct option was selected by 52% of the test-takers on average; the percentages of selected options were higher with incorrect postmodifiers (20%) and those in Japanese word order (19%). These percentages were higher than other option types (10% to 14%). The result accords with Study 1 and can be predicted from differences between English and Japanese word orders.

Furthermore, the percentages of selection in the distractor types were higher than 10% of all the students (10% to 20%). Compared to Study 1, the percentages of selection were higher in all the distractor types (e.g., for postmodification structures, 13% in Study 1 and 20% in Study 2). This trend is the same as when test-takers who received less than 80% correct and who failed on average were included in the analysis (e.g., for postmodification structures, 16% in Study 1 and 21% in Study 2). This may be because the number of items for each distractor type increased, which may have shown stable results, or because the revised test had an improved ability to identify test-takers who experience problems with NP internal structures. Additionally, if there were distractor types with extremely low or high percentages, we would need to think of revising the distractor type classification, perhaps by abandoning and separating categories. However, the result that options with each distractor type were used to some degree seems to indicate that the current five distractor type classification is appropriate to detect error patterns for diagnosis.

[Insert Table 7 about here]

Reliability analysis. The reliability coefficients, as presented in Table 7, indicate a sufficient level of reliability for all the NP groups and an improvement from Study 1. The result showed that the number of items can be reduced up to 6 for NP Group 2, up to 8 for Group 3, and up to 7 for Group 5, while the number for Group 4 should remain at 12. Additionally, the agreement of decisions at the
Development and Validation of a Diagnostic Grammar Test

Cut-off point of .80 was sufficiently high, and was much higher than in Study 1. In particular, the agreement of Group 2 increased from .54 in Study 1 to .85 in Study 2. Thus, the revision process seemed to be successful in improving reliability.

[Insert Table 8 about here]

Comparisons between NP groups. Table 9 provides the descriptive statistics for the four NP groups, with all the means at less than .80 (42% to 67%), which means that students on average could not pass any NP groups. The repeated measures ANOVA indicated that there was overall a significant difference between the NP groups with a relatively large degree, $F(2.58, 273.69) = 45.13$, $p < .001$ (Greenhouse-Geisser adjustment), $\eta_p^2 = 0.30$. The Bonferroni post-hoc comparisons indicated that NP Group 2 was easier than Groups 3 and 5, which was easier than Group 4 (Groups 2 > 3 = 5 > 4, when non-significant differences were expressed as =). Similar to Study 1, the difficulty order of some of the NP groups was not the same as that predicted. In particular, NP Group 4 was found to be significantly more difficult than Group 5.

[Insert Table 9 about here]

We thus classified 12 Group 4 items into three: present participle, past participle, and to-infinitive. Table 9 indicated that means of Groups 4 (present participle) and 4 (past participle) were very similar, so we combined these into one group: Group 4 (present/past participle). This change, in which the present and past participle items now possess a similar difficulty level, may be that we attempted to avoid including familiar chunks (e.g., made in …) in the test.

Again the ANOVA was conducted to compare between Groups 2, 3, 5, 4 (present/past participle), and 4 (to-infinitive). The result provided a significant difference between the NP groups with a fairly large degree, $F(3.19, 337.72) = 50.17$, $p < .001$ (Greenhouse-Geisser adjustment), $\eta_p^2 = 0.32$. The multiple comparisons showed this order: Groups 2 > 3 = 4 (present/past participle) > 5 > 4 (to-infinitive). Across Studies 1 and 2, postmodification (seen in Groups 3 to 5) was more difficult than premodification (Group 2); moreover, to-infinitive items in
NP Group 4 were consistently more difficult than other items in Group 4 as well as in Groups 2, 3, and 5. A difference between the two studies is that while Study 1 showed Group 3 was significantly easier than Group 5, there was no such significant difference in Study 2. This distinction can be explained by virtue of a difference in the number of participants. The effect sizes, which show the degree of mean difference, were similar and negligible across the studies (corrected Hedges’ $g = 0.07$ and 0.10 for Studies 1 and 2, respectively). However, Study 1 had 387 test-takers, and the larger number must have led to statistical significance, even with the very small mean difference, since test statistics are affected by not only effect size but also sample size (see Kline, 2004). One note here is that the interpretation of Study 2 results does not change, since the effect sizes of the mean difference between two groups were negligible when there were no statistically significant differences ($g = 0.00$ to 0.11), and the effect sizes were moderate or large when there were statistically significant differences ($g = 0.58$ to 1.25).

We predicted that Group 2 (premodification) is easier than the other NP groups, which was confirmed as intended, in line with previous studies (e.g., Hashimoto & Hirai, 2007). However, the result also showed two differences from the predicted difficulty order. First, no statistically significant differences were observed between NP Groups 3, 4 (present/past participle), and 5. This indicates that difficulty may not differ much between phrase-modified NPs (Groups 3, 4) and clause-modified NPs (Group 5), which suggests that the difference in the structural complexity between modifying phrases and clauses may not be reflected in the changes in difficulty. The result seems to indicate that the prediction derived from Pienemann’s (1998) theory did not function as expected in our test. One reason may come from the test format: We employed a multiple-choice format in which test-takers basically needed to judge the appropriate word order. Thus, the processes Pienemann assumes may not be the same as those required in the test, and the format may not be discriminatory enough to distinguish between Groups 3, 4, and 5.

The second result divergent from the prediction was that NP Group 4 (to-infinitive) was the most difficult of all the groups. The order of the presentation of grammatical structures in textbooks seems unlikely because past participles (Group 4) and relative clauses (Group 5) are often presented in
textbooks in the ninth year, whereas *to*-infinitive and present participles (both Group 4) are usually taught in the eighth year. Japanese textbooks tend to adopt grammar syllabi and the teachers normally teach grammatical structures according to the textbooks.

A possible explanation for Group 4 having the highest difficulty is that there is a distractor that is acceptable as English but that is incorrect due to incongruence with the Japanese stem. An example is Option 2 (*to buy a book*) below, which is correct as an English phrase but wrong as an option.

Example item 3 (NP Group 4)

これは、買うべき本です。

*Kore wa kaubeki hon desu.* (This is a book to buy.)

1. a book to buy* (35%)
2. to buy a book (46%; Postmodification)
3. buy to a book (12%; Japanese word order)
4. to a book buy (4%; Phrase structure)
No response (4%)

This incorrect option attracted more students (46%) than did the correct option (Option 1; 35%). Moreover, a similar tendency could be observed with respect to the other items in Group 4 (*to*-infinitive). This suggests that students who are attracted by these powerful distractors are unlikely to possess sufficient knowledge of NP Group 4 internal structures to be able to eliminate this option as wrong and that they may select it based on the naturalness and/or their intuitions regarding English. Since such items appear able to separate those who possess the necessary knowledge from those who do not, the items should therefore be regarded as useful discriminatory items.

Another possible explanation is that our interpretation may not have been correct regarding the processability theory (Pienemann, 1998), on which our predication was based. Pienemann assumed that the *to*-infinitive used as a complementizer (e.g., *want to go*) requires the S-procedure, but, in fact, he did not discuss other types of *to*-infinitive. Thus, it is possible that the structures of *to*-infinitive used for postmodification (in Group 4) require more processing than
the subordinate procedure (in Group 5). To sum, the difficulty of NPs with the to-infinitive needs close examination in terms of test format and theory.

Although our findings in terms of NP difficulty order may be related to English learners’ NP developmental order, we need to refrain from generalizing the results to other contexts for two reasons. First, the comparatively high difficulty in Group 4 (to-infinitive) is partly due to the difficulty of eliminating strong distractors. As Green and Weir (2004) demonstrated, test formats may change the difficulty order of grammatical structures. Second, our results are based on the percentage of correct answers, which Purpura (2004) names “accuracy-based scores”; we did not use “developmental-based scores,” which consider interlanguage development, for example, by “assigning partial-credit to scores” (p. 37). Whether our results can be applied to other contexts needs to be further examined.

Our results shows that Group 2 is the easiest of all the NP groups, whereas Group 4 (to-infinitive) is the most difficult, and that Groups 3 (prepositional phrase), 4 (present/past participle), and 5 (relative clause) have similar difficulty levels. Thus, in the EDiT Grammar, NP Group 4 (to-infinitive) can be re-categorized as NP Group 6 for more precise diagnosis, which will lead to a further revision of the test specifications. As for the number of items needed to achieve high reliability (see Table 8), NP Group 4 (present/past participle) requires 12 items and Group 6 (renamed, to-infinitive) needs 6 items.

Conclusion

In Studies 1 and 2, we have reported the empirical evidence that can be used in support of the validity argument. More interpretable and stronger evidence came from Study 2, where the revised test was employed; we thus summarize only the Study 2 results here. In the first research question (RQ), we found item discrimination high, which was thus deemed appropriate. The percentages of selected options in the distractor types were investigated in RQ2; the result showed that all distractor types were selected by at least 10% of all the test-takers, and especially by test-takers who failed to pass the NP groups overall. In our diagnostic grammar test, the quality of distractors is essential to provide error pattern profiles. The evidence that a certain percentage of examinees select
options in each distractor type points to the need for and effectiveness of the error detection system.

RQ3 dealt with the reliability of the test items. It was found that the reliability was high, with the minimum number of items required varying from 6 to 12; the agreement of decision at the cut-off point of .80 was also high. RQ4 assessed whether NP group difficulty increased in the order of Groups 2 < 3 < 4 < 5. When NP Group 4 was divided into two, the result showed that NP group difficulty increased as follows: Groups 2 < 3 = 4 (present/past participle) = 5 < 4 (to-infinitive); Group 4 (to-infinitive) was the most difficult. In general, the EDiT Grammar provided favorable evidence for the validity in interpreting the test scores in the case of a fairly low-stakes diagnostic test.

In order to reinforce the validity evidence, we will continue (a) to investigate other aspects of validity such as the impact of the test and the usefulness of diagnostic profiles in teaching NP internal structures, (b) to enhance the quality of the diagnostic reports by creating a systematic procedure for producing diagnosis based on test scores, by using cognitive diagnostic assessment frameworks (e.g., Lee & Sawaki, 2009), (c) to develop remedial activities for students experiencing problems in understanding NPs, and (d) to investigate the effectiveness of such activities empirically.

The current study is significant in mainly two ways. First, we constructed a diagnostic grammar test to reveal students’ strengths and weaknesses and suggest future direction in providing instruction. Through this test development process, we demonstrated one approach to constructing a diagnostic test. Second, the present research confirmed previous research findings that indicate the difficulty of acquiring NP structures—specifically, those with postmodifiers—by Japanese learners of English (e.g., Hashimoto & Hirai, 2007). We also presented empirical evidence that Pienemann’s (1998) theory is not very useful in predicting the difficulty of the EDiT Grammar, perhaps due to the effects of the multiple-choice format, attractive distractors, or a grammatical structure that was not dealt with in his theory. Moreover, the result that average students failed to pass any NP groups supports the claim that greater emphasis must be placed on the need for a systematic teaching of NP internal structures in Japan. Teachers would need to consider potentially difficult points (e.g., word order of the head NP and
Development and Validation of a Diagnostic Grammar Test

pre-/postmodifiers, especially *to*-infinitive phrases) and provide some assistance in overcoming them. The EDiT Grammar would help teachers identify students in need of instruction and their specific remedial areas.

References


Kanatani, K., & English Diagnostic Test (EDiT) Development Group. (2006). *Eigo shindan test kaihatsu heno michi* [Road to the development of the EDiT: Trajectory of the ELPA English Diagnostic Project]. Tokyo: ELPA.


foreign languages]. Tokyo shoseki.
Appendix A

Sample Test Items From the Revised Test of English Diagnostic Test of Grammar (EDiT Grammar)

Instruction: Choose the most appropriate English phrase for the underlined part of the Japanese sentence from 1 to 4 (Originally written in Japanese).

1. 私は、アジアの国を3カ国訪れました。 (NP Group 2)
   "Watashi wa Ajia no kuni wo sankakoku otozureta koto ga arimasu."
   (I have visited three Asian countries.)
   1. three Asian countries*
   2. Asian countries three (Japanese word order)
   3. three countries Asian (Premodification)
   4. countries Asian three (Postmodification)

2. パーティーの食事はおいしかったです。 (NP Group 3)
   "Party no shokuji wa oishikatta desu."
   (The food at the party was good.)
   1. the party at the food (Japanese word order)
   2. the food at the party*
   3. the food the party at (Phrase)
   4. at the party the food (Postmodification)

3. これは、インターネットに使われるコンピュータです。 (NP Group 4)
   "Kore wa internet ni tsukawareru computer desu."
   (Here is the computer used for the internet.)
   1. the computer used for the internet*
   2. used for the internet the computer (Postmodification)
   3. the internet for used the computer (Japanese word order)
   4. used the computer for the internet (Phrase)

4. あの人は、私がテレビで見た歌手です。 (NP Group 5)
   "Anohito wa watashi ga terebi de mita kashu desu."
   (That person is the singer who I saw on TV.)
1. I saw the singer on TV who   (Clause)
2. who I saw on TV the singer   (Postmodification)
3. the singer who I saw on TV* 
4. I saw on TV who the singer   (Japanese word order)
### Appendix B

#### Detailed Test Structure of the Revised Test of the EDiT Grammar

<table>
<thead>
<tr>
<th>Group</th>
<th>Item type</th>
<th>Option type [Example]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 2</td>
<td>D + A + N</td>
<td>Option 1: Answer [one popular CD]</td>
</tr>
<tr>
<td>(12 items)</td>
<td>(a) articles: 4 items</td>
<td>Option 2: Premodification (D + N + A) [one CD popular]</td>
</tr>
<tr>
<td></td>
<td>(b) numerals: 4 items</td>
<td>Option 3: Premodification (N + A + D) [CD popular one]</td>
</tr>
<tr>
<td></td>
<td>(c) interrogative: 4 items (Use what and which mainly. Don’t use how many)</td>
<td>Option 4: Japanese (A + N + D) [popular CD one]</td>
</tr>
<tr>
<td>Group 3</td>
<td>NP1 + Pre + NP2</td>
<td>Option 1: Answer [the boy in the room]</td>
</tr>
<tr>
<td>(10 items)</td>
<td></td>
<td>Option 2: Phrase (NP1 + NP2 + Pre) [the boy in the room]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Option 3: Japanese (NP2 + Pre + NP1) [the room in the boy]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Option 4: Postmodification (Pre + NP2 + NP1) [in the room the boy]</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Note.</strong> The preposition of should be used for a few items.</td>
</tr>
<tr>
<td>Group 4</td>
<td>NP1 + present/past participle (or to-infinitive) phrase</td>
<td>Option 1: Answer [some work to finish]</td>
</tr>
<tr>
<td>(12 items)</td>
<td></td>
<td>Option 2: Phrase (~ing [-ed or to] + NP1 + the rest) [to some work finish]</td>
</tr>
<tr>
<td></td>
<td>(a) present participle (~ing): 4 items</td>
<td>Option 3: Japanese (the rest + ~ing [-ed or to] + NP1) [finish to some work]</td>
</tr>
<tr>
<td></td>
<td>(b) past participle (~ed): 4 items</td>
<td>Option 4: Postmodification (~ing [-ed or to] + the rest + NP1) [to finish some work]</td>
</tr>
<tr>
<td></td>
<td>(c) to-infinitive (to V): 4 items</td>
<td></td>
</tr>
<tr>
<td>Group 5</td>
<td>NP1 + (RP) + C</td>
<td>Option 1: Answer [the music which all my family members enjoy; the language she used]</td>
</tr>
<tr>
<td>(12 items)</td>
<td></td>
<td>Option 2: Japanese (a, b, c) C + RP + NP1 [all my family members enjoy which the music]</td>
</tr>
<tr>
<td></td>
<td>(a) who: 4 items</td>
<td>(d) the + C + N [the she used language]</td>
</tr>
<tr>
<td></td>
<td>(b) which: 4 items</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(c) that: 2 items</td>
<td>Option 3: Postmodification (a, b, c) RP + C + NP1 [which all my family members enjoy the music]</td>
</tr>
<tr>
<td></td>
<td>(d) contact clause: 2 items</td>
<td>(d) C + NP1 [she used the language]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Option 4: Clause (a, b, c) C + NP1 + RP [all my family members enjoy the music which]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(d) NP1 + V + Subject [the language used she]</td>
</tr>
</tbody>
</table>

**Note.** A = adjective; C = clause; D = determiner; N = noun; NP1 = head noun phrase; NP2 = non-head-noun phrase; Pre = preposition; RP = relative pronoun; V = verb. See Table 2 for the meanings of distractor types. We can create options involving other patterns, within each distractor type.
## Tables

### Table 1

<table>
<thead>
<tr>
<th>Noun Phrase Groups According to the Internal Complexity</th>
<th>Structure</th>
<th>Example</th>
</tr>
</thead>
</table>
| Group 1 (not tested)                                    | Determiner (article, possessive, demonstrative, quantifier, numeral, interrogative) + head noun | *my brother*  
*two sweaters* |
| Group 2                                                 | (Determiner) + premodifier (adjective, NP, NP’s) + head noun | *a green sweater*  
*the boy’s name*  
*which beautiful glass* |
| Group 3                                                 | NP (Groups 1 or 2) + prepositional phrase | *a surprise party for my brother* |
| Group 4                                                 | NP (Groups 1 or 2) + present/past participle (or to-infinitive) phrase | *a couple having a dinner time*  
*books written in English*  
*chances to go to parties* |
| Group 5                                                 | NP (Groups 1 or 2) + relative clause | *things (that) I bought yesterday* |

*Note. NP = noun phrase. aWe did not include premodifying present/past participles (e.g., *a singing boy*), because it is rather different from other premodifying types (e.g., *a happy boy*) in terms of structure and period when students are exposed in the syllabus.*
Table 2

*Classification of Distractor Types*

<table>
<thead>
<tr>
<th>Problematic areas</th>
<th>Description (Diagnosis)</th>
<th>Examples</th>
</tr>
</thead>
</table>
| Premodification structure | When students select an option with a split premodification structure.  
(Students do not understand premodification structures of Group 2.) | the name boy's [the boy's name]  
Item 1. Option 3 |
| Phrase structure       | When students select an option with a split phrase structure.  
(Students do not understand phrase structures.) | the boy the room in [the boy in the room]  
Item 1. Option 5 |
| Clause structure       | When students select an option with a split clause structure.  
(Students do not understand clause structures.) | any friends who the guitar play can [any friends who can play the guitar]  
Item 2. Option 3 |
| Postmodification       | When students select an option with an incorrect postmodification (i.e., a structure in which a modifier does not modify their head NP).  
(Students do not understand relationships between the head noun phrase {e.g., the boy in Example item 1} and a modifier {e.g., in the room}.) | in the room the boy [the boy in the room]  
Item 1. Option 4  
Item 2. Option 2 |
| Japanese word order    | When students select an option having the Japanese word order.  
(Students are affected by Japanese word order and do not understand English word order.) | the room in the boy [the boy in the room]  
Item 1. Option 1  
Item 2. Option 4 |

*Note.* [ ] = Answer.
Table 3

*Example of Diagnostic Profiles*

<table>
<thead>
<tr>
<th>Basic grammar level</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Types of noun phrase groups acquired</td>
<td>Your level is two out of five.</td>
</tr>
</tbody>
</table>

- This student seems to have the knowledge to understand the internal structure of basic noun phrases of Group 2, such as *Tom’s beautiful car*. However, she seems to have difficulty understanding the internal structure of more difficult groups, such as *the boy in the room* and *some work to finish*.

<table>
<thead>
<tr>
<th>Error pattern</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>She may rely on Japanese word orders; in other words, she probably does not understand the English word order of the modifier and the noun. She may find it difficult to distinguish <em>a cat on the desk</em> from <em>a desk on the cat</em>. Possible remedial activities include translating similar phrases and exercises in reordering words.</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* With this diagnostic profile, further explanations and examples of each NP group structure and some exercises for the student are attached.

---

Table 4

*Descriptive Statistics of Option Types in the Pilot Test*

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k$</td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Answer</td>
<td>23</td>
<td>.51</td>
<td>.27</td>
<td>.93</td>
<td>.06</td>
</tr>
<tr>
<td>Internal</td>
<td>13</td>
<td>.07</td>
<td>.07</td>
<td>.01</td>
<td>.03</td>
</tr>
<tr>
<td>Phrase</td>
<td>12</td>
<td>.06</td>
<td>.10</td>
<td>.01</td>
<td>.03</td>
</tr>
<tr>
<td>Clause</td>
<td>6</td>
<td>.09</td>
<td>.13</td>
<td>.03</td>
<td>.08</td>
</tr>
<tr>
<td>Postmodification</td>
<td>18</td>
<td>.13</td>
<td>.13</td>
<td>.04</td>
<td>.05</td>
</tr>
<tr>
<td>Japanese</td>
<td>20</td>
<td>.13</td>
<td>.13</td>
<td>.02</td>
<td>.03</td>
</tr>
<tr>
<td>Others</td>
<td>--</td>
<td>.01</td>
<td>.06</td>
<td>.00</td>
<td>.01</td>
</tr>
</tbody>
</table>

*Note.* $N = 387$. $k =$ Number of items. Answer = When students selected correct options; Others = Students did not provide an answer or selected two options for one question. See Table 2 for the meanings of labels. This applies to Tables 5 to 9.
Table 5

*Reliability and Agreement of Decisions at the Cut-Off Point in the Pilot Test*

<table>
<thead>
<tr>
<th></th>
<th>G2</th>
<th>G3</th>
<th>G4</th>
<th>G5</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>23</td>
</tr>
<tr>
<td>$\Phi$</td>
<td>.55</td>
<td>.63</td>
<td>.64</td>
<td>.71</td>
<td>.86</td>
</tr>
<tr>
<td>Minimum $k$ for $\Phi = .70$</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>--</td>
</tr>
<tr>
<td>$\Phi (.80)$</td>
<td>.54</td>
<td>.73</td>
<td>.77</td>
<td>.79</td>
<td>.91</td>
</tr>
</tbody>
</table>

Table 6

*Descriptive Statistics by NP Group in the Pilot Test*

<table>
<thead>
<tr>
<th></th>
<th>$M$</th>
<th>$SD$</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>G2</td>
<td>.80</td>
<td>.24</td>
<td>.00</td>
<td>1.00</td>
<td>−1.15</td>
<td>.62</td>
</tr>
<tr>
<td>G3</td>
<td>.61</td>
<td>.29</td>
<td>.00</td>
<td>1.00</td>
<td>−.20</td>
<td>−.91</td>
</tr>
<tr>
<td>G4</td>
<td>.56</td>
<td>.29</td>
<td>.00</td>
<td>1.00</td>
<td>−.04</td>
<td>−.94</td>
</tr>
<tr>
<td>G5</td>
<td>.59</td>
<td>.31</td>
<td>.00</td>
<td>1.00</td>
<td>−.18</td>
<td>−1.06</td>
</tr>
<tr>
<td>G4 (present participle)$^a$</td>
<td>.62</td>
<td>.40</td>
<td>.00</td>
<td>1.00</td>
<td>−.45</td>
<td>−1.32</td>
</tr>
<tr>
<td>G4 (past participle)$^a$</td>
<td>.73</td>
<td>.35</td>
<td>.00</td>
<td>1.00</td>
<td>−.91</td>
<td>−.42</td>
</tr>
<tr>
<td>G4 (to-infinitives)$^a$</td>
<td>.35</td>
<td>.41</td>
<td>.00</td>
<td>1.00</td>
<td>.62</td>
<td>−1.24</td>
</tr>
</tbody>
</table>

*Note.* $N = 387$. $^a k = 2$.

Table 7

*Descriptive Statistics of Option Types in the Revised Test*

<table>
<thead>
<tr>
<th></th>
<th>$k$</th>
<th>All</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
<td></td>
</tr>
<tr>
<td>Answer</td>
<td>46</td>
<td>.52</td>
<td>.14</td>
<td>.88</td>
<td>.07</td>
<td>.45</td>
<td>.18</td>
<td></td>
</tr>
<tr>
<td>Internal</td>
<td>32</td>
<td>.10</td>
<td>.09</td>
<td>.02</td>
<td>.03</td>
<td>.12</td>
<td>.09</td>
<td></td>
</tr>
<tr>
<td>Phrase</td>
<td>22</td>
<td>.12</td>
<td>.09</td>
<td>.05</td>
<td>.06</td>
<td>.15</td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td>Clause</td>
<td>12</td>
<td>.14</td>
<td>.04</td>
<td>.03</td>
<td>.05</td>
<td>.16</td>
<td>.13</td>
<td></td>
</tr>
<tr>
<td>Postmodification</td>
<td>34</td>
<td>.20</td>
<td>.13</td>
<td>.06</td>
<td>.07</td>
<td>.21</td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td>38</td>
<td>.19</td>
<td>.07</td>
<td>.03</td>
<td>.04</td>
<td>.22</td>
<td>.11</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>--</td>
<td>.02</td>
<td>.02</td>
<td>.00</td>
<td>.00</td>
<td>.02</td>
<td>.10</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* $N = 107$. 

33
### Table 8

*Reliability and Agreement of Decisions at the Cut-Off Point in the Revised Test*

<table>
<thead>
<tr>
<th></th>
<th>G2</th>
<th>G3</th>
<th>G4</th>
<th>G5</th>
<th>Overall</th>
<th>G4 (present/past participle)</th>
<th>G4 (to-infinitive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>12</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>46</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>$\Phi$</td>
<td>.83</td>
<td>.77</td>
<td>.72</td>
<td>.82</td>
<td>.93</td>
<td>.62</td>
<td>.65</td>
</tr>
<tr>
<td>Minimum $k$</td>
<td>6</td>
<td>8</td>
<td>12</td>
<td>7</td>
<td>--</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>$\Phi (.80)$</td>
<td>.85</td>
<td>.89</td>
<td>.92</td>
<td>.92</td>
<td>.97</td>
<td>.85</td>
<td>.90</td>
</tr>
</tbody>
</table>

### Table 9

*Descriptive Statistics by NP Group in the Revised Test*

<table>
<thead>
<tr>
<th></th>
<th>$M$</th>
<th>$SD$</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>G2</td>
<td>.67</td>
<td>.27</td>
<td>.00</td>
<td>1.00</td>
<td>−.52</td>
<td>−.60</td>
</tr>
<tr>
<td>G3</td>
<td>.51</td>
<td>.28</td>
<td>.00</td>
<td>1.00</td>
<td>.20</td>
<td>−1.20</td>
</tr>
<tr>
<td>G4</td>
<td>.42</td>
<td>.24</td>
<td>.00</td>
<td>1.00</td>
<td>.44</td>
<td>−.22</td>
</tr>
<tr>
<td>G5</td>
<td>.48</td>
<td>.29</td>
<td>.08</td>
<td>1.00</td>
<td>.36</td>
<td>−1.16</td>
</tr>
<tr>
<td>G4 (present participle)$^a$</td>
<td>.48</td>
<td>.29</td>
<td>.00</td>
<td>1.00</td>
<td>.07</td>
<td>−.76</td>
</tr>
<tr>
<td>G4 (past participle)$^a$</td>
<td>.48</td>
<td>.32</td>
<td>.00</td>
<td>1.00</td>
<td>.09</td>
<td>−1.01</td>
</tr>
<tr>
<td>G4 (present/past participle)$^b$</td>
<td>.48</td>
<td>.26</td>
<td>.00</td>
<td>1.00</td>
<td>.15</td>
<td>−.80</td>
</tr>
<tr>
<td>G4 (to-infinitives)$^a$</td>
<td>.30</td>
<td>.32</td>
<td>.00</td>
<td>1.00</td>
<td>.98</td>
<td>−.16</td>
</tr>
</tbody>
</table>

*Note.* $N = 107$. $^a k = 4$; $^b k = 8$. 

---

Development and Validation of a Diagnostic Grammar Test©34