http://e-flt.nus.edu.sg/

Electronic Journal of Foreign Language Teaching
2008, Vol. 5, No. 1, pp. 13–40
© Centre for Language Studies
National University of Singapore

Acquisition of English Grammatical Features by Adult Japanese EFL Learners: The Application of Item Response Theory in SLA Research[1]

Kenichi Yamakawa
(kyamakaw@yasuda-u.ac.jp)
Yasuda Women's University, Japan

Naoki Sugino
(gwisno@is.ritsumei.ac.jp)
Ritsumeikan University, Japan

Hiromasa Ohba
(hohba@juen.ac.jp)
Joetsu University of Education, Japan

Michiko Nakano
(nakanom@waseda.jp)
Waseda University, Japan

Yuko Shimizu
(yukos@ec.ritsumei.ac.jp)
Ritsumeikan University, Japan

Abstract

Although the last couple of decades have seen great theoretical developments in Second Language Acquisition (SLA), many studies have focused on single grammatical features, and have assumed interrelationships among features are granted by a given linguistic theory. One of the weaknesses of the grammaticality judgment task, the most common measurement tool in SLA, is its reliability. Since the data obtained from different studies are sample- and test-dependent, it is difficult to compare the findings. In order to compensate for this weakness, we employ IRT (Item Response Theory) to equate the data from different grammaticality judgment tests. Since the parameters of IRT models provide a theoretical justification for equating scores, comparison of acquisition among different grammatical items on the same scale is possible with a high degree of reliability. Approximately 1,200 adult native speakers of Japanese learning EFL in Japan participated in our study. These participants were given grammaticality judgment tests in order to examine acquisitions of various English grammatical features such as unergative verbs, unaccusative verbs, psych verbs, relative clause constructions, *wh*-question constructions, to infinitives, and dative alternation. On the basis of our findings, we discuss the appropriateness and benefits of using IRT in SLA research, and explore interrelationships among these grammatical features in participants' acquisition process.

1 Introduction

1.1 Item Response Theory in language testing

Many researchers in Second Language Acquisition (SLA) as well as practitioners in language teaching are still widely using Classical Test Theory (CTT) in their analysis, whose foundation was provided by Charles Spearman's conception of an observed test score consisting of the true score and a measurement error component (Baker, 1997). Based on the theory, several indices such as item facility, item discrimination and test reliability estimation are calculated and used for many educational purposes. Application of CTT, however, poses some limitations and problems. That is, test taker characteristics and test characteristics cannot be separated: each can be interpreted only in the context of the other (Hambleton, Swaimnathan & Rogers, 1991). Therefore, item analysis such as item facility and item discrimination in CTT, for example, does not provide information about how test takers of different levels have performed on the item. In order to address the shortcomings of CTT and to yield a more complete image of how each item functions to test developers, a new measurement system, Item Response Theory (IRT) or latent trait theory has been advanced over the past 40 years in the fields of educational measurement and psychometrics.

IRT relates characteristics of individual traits and characteristics of items, which we call item parameters to the probability of a correct response. Crocker and Algia (1986) describe that the "heart" of the theory is a mathematical model of how examinees at different ability levels for the trait should respond to an item, which allows us to compare the performance of examinees who have taken different tests (p. 339). As the basis for IRT, there are two postulates. First, the performance of an examinee on a test item can be predicted by a set of factors called traits. Secondly, the relationship between examinees' item performance and the set of traits underlying item performance can be described by a monotonically increasing function called item characteristic function. This function specifies that as the level of the trait increases, the probability of a correct response to an item increases (Hambleton et al., 1991). Various IRT models have been developed for dichotomous and polytomous data. Along with the development of models, a variety of programs are now available for estimating IRT parameters, particularly for dichotomous unidimensional models. It is easy to conceive that language testers set their eyes on IRT, which has been in the spotlight since the late 1980's in the area of language testing for designing tests, test calibration and item bank construction.

1.2 Background of the present study

Bachman and Cohen (1998) describe that, in applied linguistics, relatively small amounts of research have been incorporated between the fields of second language acquisition and language testing. In order to share knowledge and information obtained in each field, the authors have been conducting joint research over the past five years to investigate the development of grammatical competence of Japanese EFL learners. As the details of the process will be described in Section 2, we first developed a set of standardized tests called MEG (Measure of English Grammar) (Shimizu et al., 2003; Shimizu, Yamakawa, Sugino, Ohba & Nakano, 2006) in order to measure learners' general proficiency of the English grammar. Then, we also developed seven test units (six grammatical judgment tasks and one multiple-choice task), each targeting a particular grammatical feature, and administered the seven units one by one to Japanese EFL learners. Different learners were given different sets of units although MEG was administered to all the learners. Basically, in the analysis of the results of each test unit, MEG was utilized to divide the learners into groups according to different English proficiency levels, and ANOVA was conducted in order to demonstrate a developmental stage of a given grammatical feature or examine particular linguistic theories that the present authors rested on. The findings of each test unit were reported individually in our previous papers (see Nakano, Sugino, Ohba, Yamakawa & Shimizu, 2005; Ohba, Yamakawa, Sugino, Shimizu & Nakano, 2005; Ohba et al., 2006; Sugino et al., 2005; Sugino, Yamakawa, Ohba, Nakano & Shimizu, 2006; Yamakawa et al., 2003, 2005).

Many studies in SLA have focused on single grammatical items using a grammaticality judgment task, as we also did above, and have assumed that interrelationships among items are granted by a given linguistic theory. The data obtained from different studies, however, are sample- and test-dependent, which makes it difficult to interpret the results of different tasks aiming to investigate different grammatical features. For example, although our previous studies dealt with seven grammatical features, we simply could not compare the results of each test unit and make any claims about interrelationships among the grammatical features, because the seven test units were taken by different groups of learners.

In order to compensate for this weakness and obtain a higher reliability, the present study will carry out a process of equating scores using IRT so as to compare the results on the seven test units that we obtained in our previous studies and explore interrelationships among targeted grammatical features in learners' acquisition processes. MEG will be utilized as anchor items in the equation design because it was taken by all the learners.

2 The study

2.1 Purpose

The present study, which is an interim report of an ongoing research project, aims to utilize IRT in SLA research and examine interrelationships among different grammatical features in the acquisition process of Japanese EFL learners in a wider context. The application of IRT to SLA research, to our knowledge, has scarcely been investigated in the relevant literature, and therefore it is expected that our research findings, though they are descriptive at this stage, will have the potential to explore the overall development of grammatical competence of Japanese EFL learners, and to make a contribution to methodological advancements in SLA studies.

2.2 Test battery and test administration

At the present phase of our research, a test battery consisting of seven test units (six grammatical judgment tasks and one multiple-choice task) has been developed thus far, as mentioned in Section 1.2, in order to shed light on the development of grammatical competence of Japanese EFL learners in a more extensive manner. The seven units are called Units S, Y, KM, N, G, O, and H. Each of the seven units was based on a particular theoretical framework that each author rested on, and was administered individually to examine a specific grammatical feature. The purpose of Unit S, as will be explained in greater detail later, is slightly different from the other test units in that it is not to see the acquisition of a particular grammatical feature but to see how learners' cue dependency (i.e. semantic, pragmatic and syntactic cues) affects their grammaticality judgments. Units Y and KM basically have the same purpose of investigation, and they both target the acquisition of unaccusative/unergative verbs. Unit N aims to investigate learners' knowledge of dative alternation in English. The purpose of Unit G is to examine the acquisition of English psych verbs, taking the animacy effect into account. Units Y, KM, N and G, in a sense, can be categorized in the same group because they aim to scrutinize acquisition processes of argument structures of English verbs. Units O and H, on the other hand, are different from those tests in that they intend to investigate structural sides of the English grammar: Unit O focuses on relative clause constructions and Unit H focuses on *wh*-question constructions. More detailed descriptions of the seven units will be provided in Section 3.

All the units except Unit S, which instructs participants to choose the agent of a to infinitival phrase, use a format of grammaticality judgment task with a five-point (1 to 5) scale. In the process of analysis, we measured the "distance" of learners' judgments from the correct answers and converted their judgments to points (0 to 4). For example, if a learner judged a grammatical sentence as "5", s/he was given four points, and if s/he judged it as "1", s/he was given no point.

The participants in our previous studies were Japanese EFL learners at university level. The seven units were administered to them in regular classes at seven universities in Japan from 2002 to 2006. The participants were given 15-20 minutes to complete each unit. The numbers of participants who took Units S, Y, K, M, N, G, O and H were 767, 369, 133, 153, 414, 328, 444 and 399, respectively, totaling 1,185 different learners. The research findings regarding the acquisition process of the grammatical items targeted in each unit have been accumulated in our previous publications separately (for more elaborate descriptions, see Nakano et al., 2005 for Unit N; Ohba et al., 2005 for Unit H, 2006 for Unit O; Sugino et al., 2005 for Unit N, Sugino et al., 2006 for Unit G; Yamakawa et al., 2003, 2005 for Units U and KM). In the present study, all the previous data were combined and analyzed using IRT in order to make it possible to compare the results of different units given to different samples of learners on the same scale of analysis.

As also mentioned in Section 1.2, in addition to the seven units described above, we also developed MEG (Measure of English Grammar), which consisted of three test sets (MEG 1, MEG 2 and MEG 05) aiming to measure learners' general proficiency of the English grammar (for more details, see Shimizu et al., 2003, 2006). MEG 1, MEG 2 and MEG 05 contained 54 discrete-point items, 56 items and 35 items, respectively. All the items were multiple-choice type, where test-takers were to choose one correct response. The total number of participants of MEG was 1,185 university students gathered from the seven universities mentioned above; 472 students took MEG 1, 475 took MEG 2 and 716 took MEG 05. The number of students who took all the three sets was 105. MEG was administered in regular classes and they were given 20-30 minutes to complete MEG. In the present study, where IRT was adopted, MEG was utilized to equate the scores of the seven test units as common anchor items in an equation design. Further details will be explained in Section 2.3.

2.3 Data analysis

The seven units and MEG had different levels of difficulty, and the ability distributions of the participants were found to be different. Since we used different test units as measuring instruments, we needed to equate scores to compare the data obtained on these tasks. It was our concern to construct a single scale that would make it possible to compare the abilities of participants at different levels. We therefore used IRT to equate scores, using MEG as common anchor items in our equation design.

Generally speaking, equating can be carried out with any models such as the Rash model, the two-parameter logistic model and the three-parameter logistic model. In the present study, we used the two-parameter logistic model, whose elements are the item discrimination parameter (*a*-parameter) and the item difficulty parameter (*b*-parameter). The data processing was delegated to the Japan Institute for Educational Measurement, Inc. (JIEM). Because we needed to equate the binary data from S and MEG and the polytomous data from the other test units, we first used the statistical software called PARSCALE for the equating process. However, a JIEM expert suggested that direct comparison of the equated *b*-parameter values from the binary data and the equated *b*-parameter values from the polytomous data might be problematic, because the *b*-parameter values of the polytomous data may be distorted in the equating process. In order to avoid the danger of distortion, we converted the polytomous data into binary: Polytomous responses of "0", "1" and "2" were converted to "0", and "3" and "4" to "1" for estimation. We used the statistical software called BILOG-MG instead, and analyzed the converted data in the same dimension using the graded response mode. After the equating process, we compared the *b*-parameter values based on the non-converted data and those based on the converted data, and confirmed that sufficiently high correlation was achieved. This would mean that the foreseen distortion was not actually in effect. Since it was also suggested that the parameter values based on the converted data are more stable, we used the converted data throughout the present study.

3 Results

Analysis of all the data using the two-parameter logistic model yielded *a*-parameter (item discrimination parameter) values and *b*-parameter (item difficulty parameter) values of all the test items in the seven test units. However, our focus will be confined only to *b*-parameter values because our main concern is learners' acquisition processes, which could presumably be inferred from the difficulty orders of test items. The range of *b*-parameter values is typically from about -2 to 2. Items with values near -2 can be considered very easy. Overall descriptions of the interrelationships among the different grammatical features, along with other findings, will be made in Section 4. Analysis of the *a*-parameter values will also be made briefly in Section 4. In Section 3, the results on each of the seven test units will be presented one by one, following the theoretical background and detailed descriptions of each unit.

3.1 Unit S

According to the Competition Model (Bates & MacWhinney, 1982; MacWhinney, 1987a, 1987b, 1992; MacWhinney, Bates & Kliegl, 1984), it is claimed that (i) a sentence contains "cues" of varied nature, such as syntactic cues (e.g. word order), morphological cues, and/or semantic cues (e.g. animacy), (ii) the relative salience of these cues varies among languages, and (iii) a language user bases her/his interpretation of a sentence on these various cues.

A number of studies in the L2 context have been carried out to investigate the "processing transfer" which can be predicted by the Competition Model. For instance, in one of Gass' experiments (Gass, 1984, 1986, 1987, 1989), Italian speakers of EFL/ESL (the ESL/EFL group) and English learners of Italian as a SL/FL (the ISL/IFL group) were instructed to choose an "actor" of the verb in three types of word sequences, NVN, NNV and VNN. In this experiment, while the ISL/IFL group displayed greater consistency in choosing the first noun in NVN and second in NNV and VNN, the ESL/EFL group was more affected by the animacy of the two nouns. Similarly, studies by Harrington (1987) and Sasaki (1991, 1994), with Japanese learners of EFL and English learners of JFL as participants, report that while English learners showed greater reliance on the word order cue, Japanese learners employed the animacy cue in choosing the "actor." Furthermore, it has also been shown that, as learners' proficiency in the target language develops, their use of the cues shifts from those they employed in their L1 to the cue dependency in L2, and this shift appears to be easier from the syntax-dependent strategy to the lexical-semantics dependent strategy than vice versa. Supporting results have been obtained with various L1-L2 combinations (Gass, 1984, 1986, 1987, 1989; Heilenman & McDonald, 1993; Kilborn & Cooreman, 1987; Koda, 1993; McDonald, 1987; McDonald & Heilenman, 1992).

In order to verify these findings with the participants of the present study, a test set called Unit S was prepared. This test set contains 36 target sentences, all of which have the identical sentence structure "NP1+ask/promise/tell+NP2+to infinitival phrase", but the order of the two NPs was controlled in terms of the likelihood of the event. The participants were instructed to choose the "agent" of the to infinitival phrase. Consider the following examples.

(1) a. The doctor told the patient to take the medicine.
    b. The doctor promised the patient to take the medicine.
    c. Ken asked Peter to join the baseball club.

In sentence (1a), both the word order cue and the likelihood of the event in which someone takes a medicine indicate that the second NP, that is, the patient, is the agent of the to infinitival phrase; thus these two kinds of cues converge. On the other hand, in sentence (1b), the word order indicates that it is the doctor who takes the medicine, which is in conflict with our general knowledge. In sentence (1c), the likelihood cue was neutralized by using proper nouns.

The Competition Model would predict that those target sentences with conflicting cues may be more difficult than those with neutral or with conflicting cue conditions. Table 1 shows the *b*-parameter values of each item and the average *b*-parameter values of the three cue conditions. Generally, the items with the conflicting cues were more difficult than those where the likelihood cue was neutralized, or those with the converging cues which turned out to be the easiest, with only a couple of items contradicting this general tendency.

Item #	Cue Cond.	*b*-parameter value	Target Sentence
S44	CONV.	-1.93	The boy promised his mother to study harder for the exam.
S47	CONV.	-1.46	The baseball player promised the boy to hit a homerun for him.
S13	CONV.	-1.41	The student promised the teacher to hand in his report by tomorrow.
S20	CONV.	-1.34	The police officer told the driver to stop the car immediately.
S11	CONV.	-1.30	The mother told her son to clean up the room.
S42	CONV.	-1.27	The customer asked the waitress to bring a new spoon.
Avg.	CONV.	-1.27	Average of the Converging Cue Sentences
S46	CONV.	-1.24	The doctor told the patient to have some rest.
S06	CONV.	-1.16	The patient promised the doctor to take the medicine.
S39	CONV.	-1.13	The boy asked his mother to bake a cake for his birthday.
S33	CONV.	-1.10	The interviewer asked the actor to talk about his new film.
S05	CONV.	-1.08	The teacher told the student to look up the word in a dictionary.
S10	NEUT.	-1.04	John promised Ken to show his new computer.
S27	NEUT.	-1.03	Ken promised John to show his new computer.
S29	NEUT.	-1.02	John promised Mary to attend the lecture.
S36	NEUT.	-0.91	Peter asked Ken to come and see his new computer.
S15	NEUT.	-0.88	Ken asked Peter to come and see his new computer.
S25	CONV.	-0.85	The boss asked his secretary to reserve a hotel room for his business trip.
S12	NEUT.	-0.74	Ken asked Lucy to bake a birthday cake for their father.
S21	NEUT.	-0.68	Mary promised John to attend the lecture.
Avg.	NEUT.	-0.66	Average of the Neutral Cue Sentences
S26	NEUT.	-0.50	Peter told Ken to join the baseball club.
S23	NEUT.	-0.43	Lucy asked Ken to bake a birthday cake for their father.
S32	NEUT.	-0.35	Ken told Peter to join the baseball club.
S24	CONF.	-0.32	The student told the teacher to look up the word in a dictionary.
S45	CONF.	-0.29	The boy told his mother to clean up the room.
S07	NEUT.	-0.28	Lucy told Sam to visit Peter's house at six.
S37	CONF.	-0.28	The driver told the police officer to stop the car immediately.
S48	CONF.	-0.17	The mother asked her son to bake a cake for his birthday.
S30	NEUT.	-0.02	Sam told Lucy to visit Peter's house at six.
S14	CONF.	0.10	The patient told the doctor to have some rest.
S09	CONF.	0.27	The secretary asked his boss to reserve a hotel room for his business trip.
S38	CONF.	0.29	The doctor promised his patient to take the medicine.
Avg.	CONF.	0.31	Average of the Conflicting Cue Sentences
S35	CONF.	0.36	The boy promised the baseball player to hit a homerun for him.
S19	CONF.	0.39	The actor asked the interviewer to talk about his new film.
S02	CONF.	0.64	The waitress asked the customer to bring a new spoon.
S04	CONF.	1.31	The teacher promised the student to hand in his report by tomorrow.
S18	CONF.	1.39	The mother promised her son to study harder for the exam.

*CONV.: Converging; NEUT.: Neutralized; CONF.: Conflicting

Table 1: The equated *b*-parameter values in Unit S

Also shown in Table 1 is a rather wide range of *b*-parameter values of items within each category. However, of particular interest here is the interaction of the cue conditions and the main

verbs in the target sentences. When the two types of cues converged, the sentences with the verb promise is the easiest (the average *b*-parameter value of -1.49), followed by those with the verb tell (-1.24) and those with the verb ask (-1.09). On the other hand, when the cues conflicted, the sentences with the verb promise was the most difficult (0.84) while those with ask and tell are easier (0.28 and -0.20, respectively). When the likelihood cue was neutralized, the targets with the verb tell is much more difficult (-0.29) than the promise sentences (-0.94) and the ask sentences (-0.74). This indicates that, when cues are in conflict, the Japanese EFL learners resort to their L1 cue dependency of employing the likelihood as the basis of identifying the agent.

3.2 Unit Y and Units KM

The "Unaccusative Hypothesis" (Perlmutter, 1978) claims that there are two distinct classes of intransitive verbs known as unergatives and unaccusatives, which exhibit different argument structures:

(2) a. Unergatives: [NP_1 [$_{VP}$ V]] (e.g. [Mary [$_{VP}$ danced]])
b. Unaccusatives: [empty [$_{VP}$ V NP_2]] (e.g. [empty [$_{VP}$ happen the accident]])
c. The accident happened 15 years ago.

The unergatives (e.g. cry, dance, laugh) originally have a logical subject (NP_1) as an external argument (i.e. an argument outside the VP), which bears the participant role "Agent" (the instigator of an event) (2a). The unaccusatives (e.g. die, fall, happen), on the other hand, originally lack a logical subject (i.e. "empty"), and have only a logical object (NP_2) as an internal argument, which assumes the participant role "Theme" (a participant affected by an event) (2b). The internal argument (NP_2) is then moved to the surface subject position in order to satisfy the English syntactic requirement which stipulates that the subject position must be filled with a lexical item (2c). As a result, the grammatical subject of the unaccusative verb originates as the logical object. At first sight it becomes difficult to make the unergative/unaccusative distinction because both have the same surface structure (i.e. NP + V).

Many researchers have noted that L2 learners of English often extend passive formation rules to unaccusatives and produce the following types of ungrammatical sentence (Zobl, 1989, p. 204):

(3) a. *Most of people are fallen in love and marry with somebody. (Japanese L1)
b. *My mother was died when I was just a baby. (Thai L1)

Unlike unaccusatives, unergatives rarely undergo this inappropriate passivization process. In addition, these nontarget sentences are observed in L2 English with various L1 backgrounds (e.g. Hirakawa, 2003; Oshita, 1997; Shomura, 1996; Sorace, 1997; Yip, 1995), and are particularly noticeable among intermediate/advanced learners (Oshita, 2001).

Two major accounts of the nontarget phenomena have been advanced so far: the NP movement account and the lexical causativization account. The former account points out that the argument structures of an unaccusative (2b) and a passive construction (4a) are almost identical in that both lack an external argument (logical subject) and that an internal argument (logical object) is moved to the surface subject position. One difference is that only the passive construction can take the be + p.p. marker to signal the NP movement (cf. 2c and 4b):

(4) a. [empty [$_{VP}$ V NP_2]] (e.g. [empty [$_{VP}$ be spoken English]]
b. English is spoken in many countries.

However, some learners notice the similarity of the two and also apply the passive formation rules to unaccusatives in order to signal NP movement, which results in inappropriate passives as in (3). According to the lexical causativization account, L2 learners treat an unaccusative verb as

transitive and temporarily create a causer of the event (5a). Then the verb is passivized with the suppression of the nonce causer (5b):

(5) a. *The driver happened the accident 15 years ago.
b. *The accident was happened 15 years ago.

With the use of the Longman Learners' Corpus, Oshita (2000) examined these two accounts and demonstrated the superiority of the NP movement account over the lexical causativization account. In order to further explore the acquisition of unaccusative/unergative verbs by Japanese EFL learners, grammaticality judgment test sets called Unit Y, Unit K and Unit M were prepared. Unit Y consists of 48 items, which fall under eight sentence categories. Six unaccusative verbs (appear, arrive, die, exist, fall, happen) and six unergative verbs (cry, dance, laugh, play, sing, work) are placed in sentence categories such as NP + V, NP + be + p.p. and NP + V + NP, in addition to six ergative verbs (break, burn, close, dry, grow, melt), which can be used both as transitive and as intransitive verbs, placed in the NP + be + p.p. and NP + V + NP sentence categories. On the other hand, Unit K and Unit M (henceforth, Units KM), consisting of 36 items each, were designed to duplicate the experiment which had already been done with Unit Y (Yamakawa et al., 2003) adding a different perspective (i.e. transitive alternation, which will not be explored here). The total 72 test items in Units KM are composed of 46 items identical to the items in Unit Y and 26 additional new items. In Units KM, four new sentence categories were added: the NP + V pattern with the same six ergative verbs as in Unit Y (except boil, which was used instead of burn), and the three sentence patterns (NP + V, NP + be + p.p. and NP + V + NP) with six transitive verbs (build, cut, find, invite, paint, write). Sample sentences in Unit Y and Units KM are provided in Table 2.

Unit	Category	Verb	Construction	Example
Y, KM	A	unaccusative	NP+V	Your letter arrived yesterday.
Y, KM	B	unergative		Her father cried at her wedding ceremony.
Y, KM	C	unaccusative	*NP+be+p.p.	*Because of the rain, the train was arrived late.
Y, KM	D	unergative		*He was cried when he heard of his mother's death.
Y, KM	E	unaccusative	*NP+V+NP	*Finally the waitress arrived the salad to us.
Y, KM	F	unergative		*The boy hit his little sister and cried her.
Y	G	ergative	NP+be+p.p.	The door was broken by the police.
KM	G	ergative	NP+V	The refrigerator broke because it was so old.
Y, KM	H	ergative	NP+V+NP	I broke a glass in the kitchen.
KM	I	ergative	NP+be+p.p.	The door was broken by the police.
KM	J	transitive	*NP+V	*The book writes in easy English.
KM	K	transitive	NP+V+NP	She wrote her telephone number on the paper.
KM	L	transitive	NP+be+p.p.	The sign was written in Spanish.

Table 2: Sample target sentences in Unit Y and Units KM

Analysis of Unit Y and Units KM using IRT yielded *b*-parameter values for all of the test items (Tables 3 and 4 below). The average *b*-parameter value of the items in Categories A, C and E (i.e.

unaccusatives) is higher than that of the items in Categories B, D and F (i.e. unergatives) both in Unit Y (-0.45 and -0.79, respectively) and Units KM (-0.32 and -0.52, respectively). This indicates that the acquisition of English intransitives does not manifest a unanimous process. Learners notice the different underlying argument structure in unaccusatives and unergatives, and experience greater degrees of difficulty with unaccusatives than with unergatives, which lends support to the Unaccusative Hypothesis.

Cat	**Item**	***b*-value**	**Cat**	**Item**	***b*-value**	**Cat**	**Item**	***b*-value**	**Cat**	**Item**	***b*-value**
a	Y06	-1.64	b	Y23	-1.98	c	Y07	-0.44	d	Y18	-0.55
a	Y13	-1.24	b	Y42	-1.95	c	Y34	-0.35	d	Y14	-0.36
a	Y20	-1.19	b	Y32	-1.47	c	Y02	0.22	d	Y29	-0.35
a	Y36	-1.09	b	Y46	-1.47	c	Y43	0.37	d	Y47	-0.26
a	Y16	-0.83	b	Y09	-1.18	c	Y22	0.86	d	Y38	-0.20
a	Y28	-0.71	b	Y03	-0.83	c	Y27	0.88	d	Y12	0.22
	Ave.	**-1.12**		**Ave.**	**-1.48**		**Ave.**	**0.26**		**Ave.**	**-0.25**
	Range	**0.93**		**Range**	**1.15**		**Range**	**1.32**		**Range**	**0.77**
Cat	**Item**	***b*-value**	**Cat**	**Item**	***b*-value**	**Cat**	**Item**	***b*-value**	**Cat**	**Item**	***b*-value**
e	Y04	-1.11	f	Y08	-1.18	g	Y19	-2.13	h	Y30	-1.74
e	Y24	-0.93	f	Y26	-0.78	g	Y25	-2.09	h	Y37	-1.60
e	Y10	-0.67	f	Y31	-0.65	g	Y33	-1.38	h	Y01	-1.58
e	Y45	-0.47	f	Y17	-0.50	g	Y15	-1.31	h	Y44	-1.45
e	Y35	-0.12	f	Y48	-0.49	g	Y05	-1.30	h	Y21	0.18
e	Y41	0.35	f	Y39	-0.22	g	Y40	-0.88	h	Y11	0.46
	Ave.	**-0.49**		**Ave.**	**-0.64**		**Ave.**	**-1.51**		**Ave.**	**-0.95**
	Range	**1.46**		**Range**	**0.96**		**Range**	**1.25**		**Range**	**2.2**

Cat: Category, *b*-value: *b*-parameter value, Ave: Average

Table 3: *b*-parameter values of the test items in Unit Y

Cat	**Item**	***b*-value**	**Cat**	**Item**	***b*-value**	**Cat**	**Item**	***b*-value**	**Cat**	**Item**	***b*-value**
a	K27	-0.91	b	M27	-1.15	c	K09	-0.49	d	K10	-0.71
a	M31	-0.67	b	K33	-0.91	c	M24	-0.38	d	M23	-0.63
a	K14	-0.56	b	K11	-0.88	c	K19	-0.13	d	K02	-0.53
a	M19	-0.48	b	K24	-0.86	c	M28	-0.07	d	M10	-0.50
a	K01	-0.44	b	M09	-0.83	c	K26	0.26	d	K31	-0.43
a	M06	-0.33	b	M16	-0.73	c	M05	0.36	d	M30	-0.23
	Ave.	**-0.57**		**Ave.**	**-0.89**		**Ave.**	**-0.07**		**Ave.**	**-0.50**
	Range	**0.58**		**Range**	**0.42**		**Range**	**0.85**		**Range**	**0.48**
Cat	**Item**	***b*-value**	**Cat**	**Item**	***b*-value**	**Cat**	**Item**	***b*-value**	**Cat**	**Item**	***b*-value**
e	M15	-0.77	f	K18	-0.43	g	M01	-0.97	h	K25	-1.50
e	K35	-0.53	f	M20	-0.36	g	M25	-0.64	h	K15	-1.37
e	K22	-0.34	f	M33	-0.35	g	K30	-0.25	h	M17	-1.17
e	K03	-0.32	f	K16	-0.22	g	K20	-0.08	h	K29	-0.59
e	M02	-0.27	f	K28	-0.08	g	M14	0.14	h	M29	0.17
e	M34	0.40	f	M11	0.34	g	K04	*	h	M08	0.27
	Ave.	**-0.31**		**Ave.**	**-0.18**		**Ave.**	**-0.36**		**Ave.**	**-0.70**
	Range	**1.17**		**Range**	**0.77**		**Range**	**1.11**		**Range**	**1.77**
Cat	**Item**	***b*-value**	**Cat**	**Item**	***b*-value**	**Cat**	**Item**	***b*-value**	**Cat**	**Item**	***b*-value**
i	M22	-1.83	j	M36	-0.99	k	M26	-1.89	l	M35	-1.20
i	M12	-1.64	j	K06	-0.65	k	M13	-1.26	l	K08	-1.01
i	K23	-0.99	j	M07	-0.54	k	K21	-1.19	l	M18	-0.98
i	K17	-0.96	j	K32	-0.34	k	M03	-1.17	l	M04	-0.81
i	K05	-0.92	j	M21	-0.22	k	K07	-1.03	l	K34	-0.77
i	M32	-0.84	j	K13	-0.05	k	K36	-0.70	l	K12	-0.68
	Ave.	**-1.20**		**Ave.**	**-0.47**		**Ave.**	**-1.21**		**Ave.**	**-0.91**
	Range	**0.99**		**Range**	**0.94**		**Range**	**1.19**		**Range**	**0.52**

Cat: Category, *b*-value: *b*-parameter value, Ave: Average,*: too difficult to be assigned *b*-parameter value

Table 4: *b*-parameter values of the test items in Unit KM

Next, the order of difficulty in Unit Y from easiest to most difficult is Category B>A>F>E>D>C (-1.48>-1.12>-0.64>-0.49>-0.25>0.26), while that of Units KM is Category B>A>D>E>F>C (-0.89>-0.57>-0.5>-0.31>-0.18>-0.07). These results show that the learners first achieved success in Categories A and B by accepting grammatical sentences with unaccusatives and unergatives, and that they had the greatest difficulty in rejecting ungrammatical passives with unaccusatives (Category C). The order of difficulty of the categories in Units Y and KM is shown in Table 5.

Categories	Examples
B: NP+V (unergative)	We all laughed when we saw his face.
A: NP+V (unaccusative)	A funny thing happened in the office today.
F: *NP+V (unergative)+NP	*Bob is so funny. He always laughs me.
E: *NP+V (unaccusative)+NP	*Jimmy happens a lot of trouble to me.
D: *NP+be+p.p. (unergative)	*They were laughed when she told a funny joke.
C: *NP+be+p.p. (unaccusative)	*I was not there when the accident was happened.

Table 5: The hierarchy of difficulty of the categories in Units Y and KM

With regard to the NP-movement vs. causativization issue, the comparison of *b*-parameter values was made between Category C and Category E. The average *b*-parameter values of the items in Categories C and E were 0.26 and -0.49 respectively in Unit Y, and -0.07 and -0.31 in Units KM. It seems that the items in Category C were more difficult for the learners than those in Category E. Thus, it could be maintained that the lexical causativization account would be validated if learners experienced the same relatively high degree of difficulty for both Categories C and E. If this account was correct, learners would be expected to create a nonce causer of an event in their representation of unaccusatives and to accept ungrammatical sentences from both Categories C and E as grammatical. Conversely, if learners experienced higher degrees of difficulty for Category C than for Category E, it would indicate that learners did not add a nonce causer but instead moved the postverbal NP and used the nontarget be + p.p., which was analogous to the standard passive derivation. Comparison of the *b*-parameter values of the items in Categories C and E imply that the NP movement account offered a more plausible explanation for the cause of passivized unaccusatives than the lexical causativization account. This provides support to Oshita's (2000, 2001) assumption that NP movement analogous to passive formation is involved in the acquisition of unaccusatives.

As Tables 3 and 4 clearly display, the ranges of *b*-parameter values in each category in Unit Y and Units KM are noticeably different; the minimum 0.42 in Category B in Units KM and the maximum 2.2 in Category H in Unit Y. This implies that test items belonging, by definition, to the same grammatical category show such wide dissimilarities in range that some items in the same category may exhibit wider differences in range than comparisons of different categories.

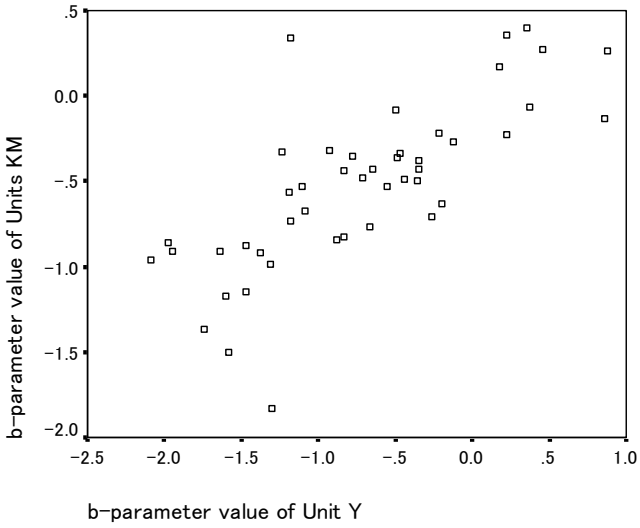


Figure 1: Scatterplot of 46 test items

Next, the 46 items that commonly appear both in Unit Y and in Units KM were extracted, and the Pearson r correlation coefficient was calculated to see if there was any relationship between the two test units in terms of difficulty values. Since the 46 items are exactly the same, it was expected that there should be strong positive relationship between the two test units (Figure 1). The results showed a relatively strong positive correlation ($r = .749$, $N=46$, $p<0.01$). This indicates that Unit Y and Units KM are ordering test items in much the same way.

However, when we look at difficulty values on a category basis, not on an item basis, we can observe some discrepancy with the 46 items between the two test units. As depicted in Figure 3 in Section 4, the distributions of the average *b*-parameter values of the categories in Unit Y and Units KM are substantially different because the categories in Unit Y are scattered more sparsely than those in Units KM. Moreover, the average *b*-parameter values of Categories A, B and C in Unit Y are vastly different from those in Units KM: Categories A and B in Unit Y are much easier than in Units KM, whereas Category C in Unit Y is much more difficult than in Units KM. As we applied IRT, which supposedly guarantees a test-free and sample-free analysis, to the same test items, the 46 items taken from Unit Y and Units KM, in theory, should display highly similar *b*-parameter values, which, in fact, was not the case here.

3.3 Unit N

Unit N intends to test the grammatical knowledge of dative alteration. It contains six kinds of category (Categories A to F). We took the animacy effect into consideration when creating the test items. Below are the characterizations of all the test items in the six categories, and the *b*-parameter values of the test items in each category are displayed in Tables 6 to 11. Unit N was originally used in one of our previous studies (Nakano et al., 2005) and the test items in Unit N were constructed on the basis of Lexical Functional Grammar (Bresnan, 1982, 1998, 2001). See Nakano (2000) and O'Grady (2001) for a further understanding of the acquisition of dative alternations in L2.

Category A consists of eight items (A01-A08 below) which are all well-formed ditransitives. The items A01 to A04 are to-datives, and the items A05 to A08 are for-datives. A01, A03, A05 and A07 bear an animate subject, and A02, A04, A06 and A08 bear an inanimate subject

Item #	Test Sentence	*b*-parameter value
A01	Mr. Jones gave me some money.	-1.32
A02	The company gave him a new job.	-1.07
A03	My wife sent me three golf clubs.	-1.28
A04	The family sent him ten apples.	-1.12
	mean	-1.20
A05	John found me a new dress.	0.20
A06	The company found him a new office.	0.47
A07	Simon made us a new dinner table.	-0.69
A08	The company made the secretary a new uniform.	2.08
	mean	0.52

Table 6: Test sentences and their *b*-parameter values in Category A

In Category A, the sentences with animate subjects were found relatively easier than those with inanimate subjects. It is interesting to note that the for-datives (the average *b*-parameter value of 0.52) were much harder than to-datives (-1.20).

Category B contains eight items (B01 - B08 below) which are al ill-formed ditransitives. The items B01 to B04 are ill-formed to-datives, and the items B05 to B08 are ill-formed for-datives. B01, B03, B05 and B07 bear an animate subject, and B02, B04, B06 and B08 bear an inanimate subject.

Item #	Test Sentence	*b*-parameter value
B01	*Mr. Jones reported me the accident.	2.03
B02	*The police reported Bill the fire.	1.13
B03	*My wife moved me three golf clubs.	-0.91
B04	*The family moved him a box of apples.	-0.52
	mean	0.43
B05	*King Arthur fought the queen the monster.	-1.07
B06	*The company burned me a lot of boxes.	-0.41
B07	*Simon discovered us a nice dinner table.	-0.63
B08	*The company discovered them a new house.	-0.82
	mean	-0.73

Table 7: Test sentences and their *b*-parameter values in Category B

The items B01 to B04 are generated by applying the dative shift rule to the sentences with the preposition “to”. The application of this rule yields ill-formed sentences. Likewise, the items B05 to B08 are generated by applying the beneficiary shift rule to the sentences with the preposition “for”. The application of this rule also yields ill-formed sentences. When we compare the average *b*-parameter values in this category, we can say that the items B01 to B04 (0.43) were far more difficult than the items B04 to B08 (-0.73).

Category C consists of eight items (C01 - C08 below) which are all grammatical prepositional datives. The items C01 to C04 are to-datives, and the items C05 to C08 are for-datives. C01, C03, C05 and C07 bear an animate subject, and C02, C04, C06 and C08 bear an inanimate subject.

Item #	Test Sentence	*b*-parameter value
C01	Mr. Jones gave some money to me.	-1.33
C02	The company gave a new job to him.	-1.29
C03	My wife sent three golf clubs to me.	-1.66
C04	The family sent ten apples to me.	-1.17
	mean	-1.36
C05	John found a new dress for me.	-1.49
C06	The company found a new office for him.	-1.02
C07	Simon made a new dinner table for us.	-1.03
C08	The company made a new uniform for the secretary.	-1.17
	mean	-1.18

Table 8: Test sentences and their *b*-parameter values in Category C

On average, the prepositional to-datives and the prepositional for-datives were of almost the same difficulty (-1.36 and -1.18, respectively), and both of them can be regarded as items with lower difficulty.

Category D consists of eight items (D01 - D08 below) which are all distracters. They are all grammatical and look like prepositional to-datives and prepositional for-datives. The items D01 to D04 look like prepositional to-datives, and the items D05 to D08 look like prepositional for-datives. D01, D03, D05 and D07 bear an animate subject, and D02, D04, D06 and D08 bear an inanimate subject.

Item #	Test Sentence	*b*-parameter value
D01	Mr. Jones reported the accident to me.	-1.27
D02	The police reported the fire to Bill.	-0.65
D03	My wife moved three golf clubs to me.	-0.9
D04	The family moved a box of apples to me.	-1.2
	mean	-1.01
D05	King Arthur fought the monster for the queen.	-1.01
D06	The company burned a lot of boxes for me.	-0.56
D07	Simon discovered a nice dinner table for us.	-1.89
D08	The company discovered a new house for them.	-1.34
	mean	-1.20

Table 9: Test sentences and their *b*-parameter values in Category D

The items with the preposition “to” and the items with the preposition “for” were of almost the same difficulty (-1.01 and -1.20, respectively), and they can be considered relatively easier than the items in other categories.

Category E contains well-formed passive sentences. The items E01 to E04 are passivized to-datives, and the items E05 to E08 are passivized for-datives. The objects of the by-phrases (i.e. the agent of the action which the verb denotes) in E01, E03, E05 and E07 are animate, and those in E02, E04, E06 and E08 are inanimate.

Item #	Test Sentence	*b*-parameter value
E01	I was given some money by Mr. Jones.	-1.1
E02	He was given a new job by the company.	-1.02
E03	I was sent three golf clubs by my wife.	-0.86
E04	He was sent ten apples by the family.	-0.45
	mean	-0.86
E05	I was found a new dress by John.	*
E06	He was found a new office by the company.	*
E07	We were made a new dinner table by Simon.	2.83
E08	The secretary was made a new uniform by the company.	1.57
	mean	Not applicable

*An asterisk (*) for the *b*-parameter value means that the value was abnormally high and was not assigned.

Table 10: Test sentences and their *b*-parameter values in Category E

On the whole, the passivized to-datives were much easier than passivized for-datives as shown in Table 10. The passivized to-datives were of relatively lower difficulty, whereas passivized for-datives could be regarded as extremely difficult.

Category F consists of ill-formed passive sentences with the "Theme" role in the subject position. The items F01 to F04 are passivized to-datives, and the items F05 to F08 are passivized for-datives. The objects of the by-phrases (i.e. the agent of the action which the verb denotes) in F01, F03, F05 and F07 are animate, and those in F02, F04, F06 and F08 are inanimate.

Item #	Test Sentence	*b*-parameter value
F01	*Some money was given me by Mr. Jones.	0.84
F02	*The new job was given him by the company.	1.03
F03	*Three golf clubs were sent me by my wife.	0.93
F04	*Ten apples were sent him by the family.	0.94
	mean	0.94
F05	*The new dress was found me by John.	-0.82
F06	*The new office was found him by the company.	-0.9
F07	*The new dinner table was made us by Simon.	-0.53
F08	*The new house was made them by the company.	-0.74
	mean	-0.75

Table 11: Test sentences and their *b*-parameter values in Category F

We can observe huge discrepancy between the average *b*-parameter values of to-datives and that of for-datives (0.94 and -0.75, respectively), and the passivized to-datives in this category were much more difficult.

So far we have reported the results of the IRT analysis on the six categories in Unit N. It is obvious that the learners find the to-datives easier than for-datives when they are well-formed (i.e. Categories A and C). We could argue that the learners' familiarity with to-datives may be related with much higher difficulty of the ill-formed to-datives in Categories B and F. In other words, the learners may have created the rule, through overgeneralization of the dative alternation rule, which stipulates that verbs which can take a prepositional phrase with "to" as in "move" or "report" could also be ditransitive verbs.

In addition, the to-datives and for-datives in Categories A, B, E and F were categorized in the same groups; however, their average *b*-parameter values were so vastly different that the to-datives and for-datives in those categories should be regarded as belonging to separate categories.

3.4 Unit G

English psych verbs are generally categorized into two classes. One class of psych verbs (ES verbs) takes an Experiencer NP as the subject and a Theme NP as the object (e.g. We admired the view). This is in accordance with the thematic hierarchy and uniform assignment of specific theta-roles. The other class of psych verbs (EO verbs), on the other hand, takes an Experiencer NP as its object, a Theme NP filling the subject position (e.g. The earthquake frightened everyone). Two different lines of accounts, viz. the lexico-semantic account (Grimshaw, 1990; Jackendoff, 1990) and the configurational account (Belletti & Rizzi, 1988), have been proposed to explain this apparent arbitrariness and some other peculiar behaviors of the psych verbs and to show how the thematic structure is mapped onto the syntactic structure.

In the context of SLA research, the EO verbs are recognized as more problematic for ESL/EFL learners. White et al. (1999) demonstrated that, regardless of the learners' native language, the EO verbs were more problematic than the ES verbs, and the difficulty with the EO verbs was more apparent among the Japanese EFL learners. Sato's (2002) analysis of a learners' corpus revealed the Japanese EFL learners used EO verbs less frequently in the transitive construction than in the passive construction. Furthermore, when the EO verbs were used as transitive verbs, all erroneous sentences had an Experiencer subject. Shomura-Isse's (2005) study with psych verbs embedded in adnominal clauses also revealed that the Japanese EFL learners had difficulty in distinguishing Experiencer and Theme (and their respective realization in the surface structure), but comparison among the proficiency levels yielded rather mixed results.

One of the reasons why the EO verbs are more problematic for the Japanese EFL learners is attributed to the fact that the Japanese language does not have the EO verb class. Compare the following sentences:

(6) a. Taro-ga jishin-wo kowagar-u
Taro-NOM earthquake-ACC fears
"Taro fears earthquakes."

b. Jishin-ga Taro-wo kowagar-ase-ru
Earthquakes-NOM Taro-ACC frightens
Earthquakes frighten Taro.

Thus, in Japanese, a Theme NP causing the changes in the mental state is explicitly marked by -(s)ase at the end of the ES verbs. Due to the lack of such explicit causative markers in English, the Japanese EFL learners need to learn the distinction between the two different types of verbs in different argument structures (Shomura-Isse, 2005; Sato, 2003).

A test set, Unit G, was prepared with the following sentence patterns (Table 12). Each category consists of four sentences with different verbs.

Cat.	Target Sentence Pattern
ES1	Human Experiencer / ES verbs / Human Theme e.g. *I admired her when I first met her and I still think she's marvelous.*
ES2	Human Experiencer / ES verbs / Inanimate Theme e.g. *They enjoy anything that breaks the dullness of their routine life.*
ES3	Inanimate Theme / ES verbs-passive / PP with Human Experiencer e.g. *His beautiful house is envied by the neighbors.*
ES4	*Inanimate Theme / ES verbs / Human Experiencer e.g. **Your skills have fully appreciated our boss.*
ES5	*Human Experiencer / ES verbs-passive / PP with Inanimate Theme e.g. **She will not be enjoyed by parties or dining out.*
EO1	Human Theme / EO verbs / Human Experiencer e.g. *A handsome man bored me with stories of the Navy.*
EO2	Inanimate Theme / EO verbs / Human Experiencer e.g. *The sharp tone of her voice amazed her boss.*
EO3	Human Experiencer / EO verbs-passive / PP with Inanimate Theme e.g. *We all were bored with his lecture.*
EO4	*Human Experiencer / EO verbs / Inanimate Theme e.g. **After the competition, I disappointed the results..*
EO5	*Inanimate Theme / EO verbs-passive / PP with Human Experiencer e.g. **Their performance was impressed by the spectators.*

Table 12: Target sentence patterns in Unit G

Figure 2 shows the distribution of the Unit G items and categories on the *b*-parameter scale; two items, namely G25 in category ES3 and G23 in ES5, were excluded because they turned out to be too easy or too difficult (the *b*-parameter value was -1.84 and 2.36, respectively).

As can be seen from Figure 2, contrary to the findings from the previous studies, there was no clear difference in difficulty between the ES verbs and the EO verbs. It should be noted, however, that four out of the five categories in the ES verb class turned out to be relatively easier with the *b*-parameter values below zero. Another interesting tendency is that the ungrammatical sentences were more difficult than the grammatical sentences, indicating that the participants in this study accepted these ungrammatical sentences as well-formed sentences.

When both NPs are human (ES1 and EO1), there was not much difference in difficulty between the ES and EO verb classes; however, when the animacy was differentiated (ES2, ES4, EO2 and EO4), the difficulty with EO4 increases, indicating that the participants accepted both grammatical and ungrammatical sentences with the EO verbs. A different tendency was observed with the passivized sentences (ES3, ES5, EO3 and EO5). The participants displayed equal ease in accepting the grammatical sentences for both verb classes (ES3 and EO3), the difficulty with the ungrammatical sentences increased not only with the EO verbs (EO5) but also with the ES verbs (ES5).

Taken together, the results indicate that the participants in this study displayed their inclination to accept sentences when the "Experiencer" role is assigned to the human NP, regardless of the verb classes or the human NP's position in the sentence. This is perfectly in line with what was discussed in Section 3.1, where it was argued that the Japanese EFL learners resort to their L1 cue dependency.

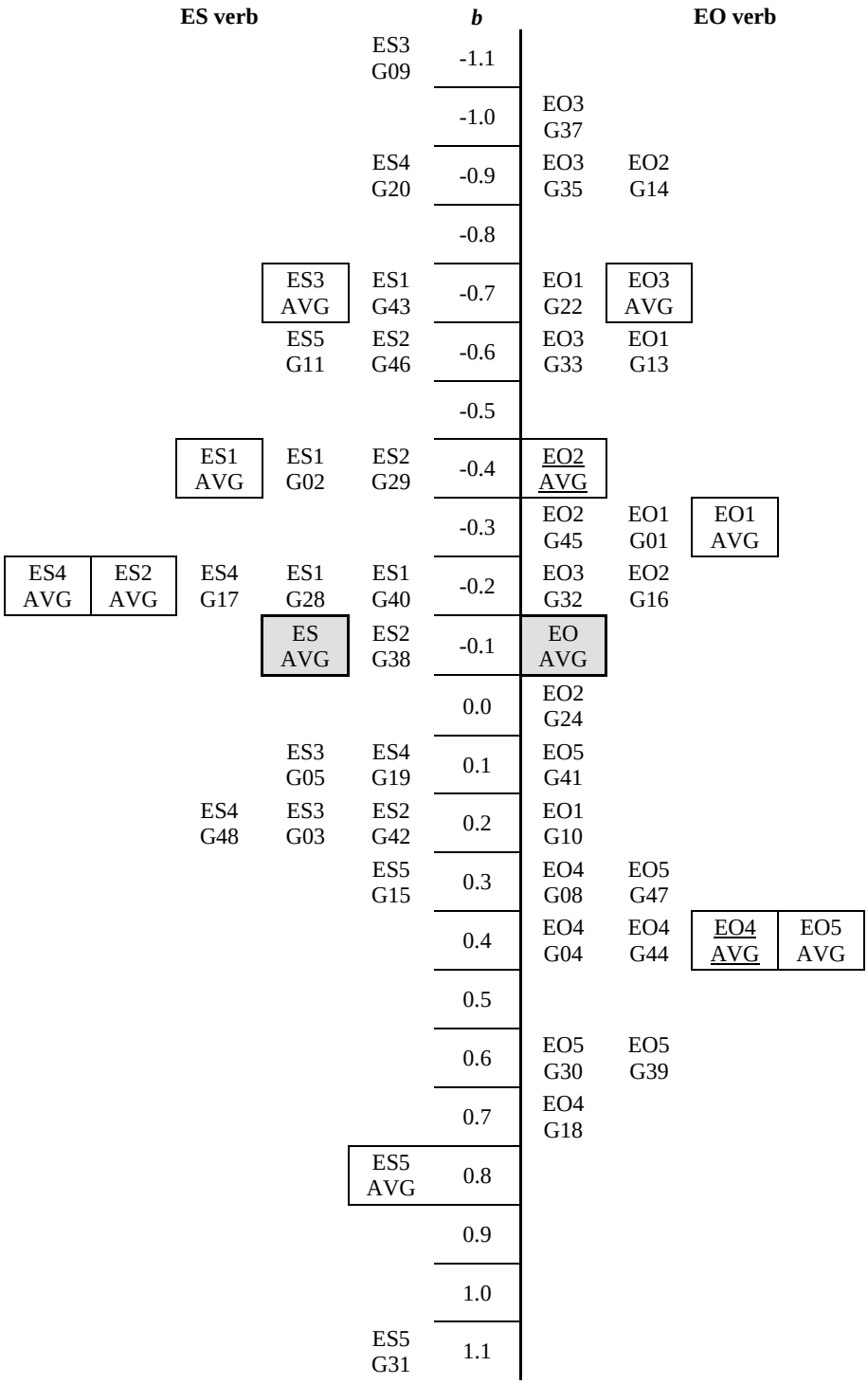


Figure 2: Distribution of the Unit G items on the *b*-parameter scale

3.5 Unit O

As Fukui and Takano (2000) mention, English and Japanese relative clause constructions crucially differ in (a) the position of the relative head and (b) whether or not the relative operator is present. For example, in English, relative clauses are formed by a relative operator who(m) being extracted from the relative clause domain and being moved to the position after the relative head modified (i.e. man), as in (7a). On the other hand, in Japanese, such a relative operator movement is not involved due to the lack of relative operators and the relative head modified (i.e. dansei) is put after the relative clause, as in (7b).

(7) a. The man [who(m)i [I saw t i yesterday]] is John.
b. [[*Watasi-ga kinoo atta*] *dansei*]-*wa John desu*.
I yesterday met man-Top John is

Within the Minimalist Program (Chomsky, 1995, 1998), overt movement is only allowed when it is motivated by the presence of a strong formal feature. English relative clauses have the feature [+R] in Complementizer (C), which drives relative-operator movement, as in (7a). In Japanese relative clauses, however, there is adjunct/predication type relation with no operator, and no feature-driven movement is required due to the lack of the operator and the feature [+R], as in (7b) (Takeda, 1999).

Since in the case of the operator-oriented relative clause as in English relative clauses, the relation between the relative head and its associated relative clause is established by binding of the relative pronoun by the relative head, they are subject to the Subjacency condition, a constraint on *wh*-movement, as in (8a). On the other hand, the Japanese relative clause is licensed by a semantic relation with the relative head such as "aboutness." Therefore, in Japanese relative clauses, the relation between gap and the relative head can be unbounded, indicating no Subjacency condition, as in (8b) (for details, see Kuno, 1973).

(8) a. *a gentlemen [whoi the suit [that ti is wearing] is dirty]
b. [[[*proi kiteiru*] *yoohuku*]-*ga yogoreteiru*] *sinsii*

Therefore, it is expected that if Japanese EFL learners make grammaticality judgments utilizing the same underlying operations as native speakers of English do, they will judge the relative clause constructions violating the Subjacency condition as well as the surface morphological properties of relative clause constructions at the same level of difficulty.

In order to investigate the acquisition of relative clause constructions, especially on whether Japanese EFL learners can acquire *wh*-movement or not, a written grammaticality judgment test, Unit O, was designed. This test includes 40 target and 5 filler sentences. The 40 target sentences fall into the following 4 groups (with 14 categories).

(9) The sentence involving the following 3 types of grammatical relative clauses
a. Category 1: Relative clauses with a *wh*-operater (7 items)
e.g. *The job which I wanted to apply for was very popular.*
b. Category 2: Relative clauses with a complementizer that (4 items)
e.g. *The picture that he is looking at was painted by Picasso.*
c. Category 3: Relative clauses with a null operator and a null complementizer (3 items)
e.g. *The house you can see on the corner was built ten years ago.*

(10) The sentences involving the following 2 types of ungrammatical relative clauses
a. Category 4: Relative clauses with a doubly-filled complementizer (who(m)/that/which that) (4 items)
e.g. *The dogs which that I gave the milk to were very small.*
b. Category 5: Relative clauses with a resumptive pronoun (4 items)
e.g. *The classmate that you don't like him is very unkind.*

(11) The sentences violating Subjacency conditions in the following 5 relative clause construction types
  a. Category 6: Relative clauses with an extraction from a relative clause (2 items)
     e.g. *This is the soup which Mari visited a restaurant which served.*
  b. Category 7: Relative clauses with an extraction from a sentential subject (2 items)
     e.g. *This is the meeting which that Taro attended shocked his parents.*
  c. Category 8: Relative clauses with an extraction from an adjunct island (2 items)
     e.g. *This is the homework which Ann went to school before she did.*
  d. Category 9: Relative clauses with an extraction from an embedded question (2 items)
     e.g. *This is the house which Peter knows when Tom bought.*
  e. Category 10: Relative clauses with an extraction from a complex NP (2 items)
     e.g. *This is the boy who(m) Jack described the way that Bill hit.*

(12) The sentences involving the following 4 grammatical island constructions
  a. Category 11: The sentences with a sentential subject (2 items)
     e.g. *That my sister went out with Johnny made Nick angry*.
  b. Category 12: The sentences with an adjunct island (2 items)
     e.g. *Many houses were damaged by the storm while I visited Korea*.
  c. Category 13: The sentences with an embedded question (2 items)
     e.g. *Steve asked me who caused the car accident yesterday.*
  d. Category 14: The sentences with a complex NP (2 items)
     e.g. *The teacher believed the claim that Jim stole the money.*

Table 13 indicates the average *b*-parameter values of the 14 categories. The surface morphological properties of relative clauses (Categories 1, 3, 4 and 5) are easier than violation of the Subjacency conditions (Categories 6, 8, 9 and 10). This means that Japanese EFL learners understand the relative clause constructions without using *wh*-movement.

Category	Average *b*-parameter value
Category 12 (adjunct island)	-1.66
Category 14 (complex NP)	-0.99
Category 13 (embedded question)	-0.96
Category 2 (that)	-0.73
Category 1 (*wh*-operator)	-0.57
Category 7 (*sentential subject)	-0.55
Category 4 (*doubly-filled comp)	-0.54
Category 5 (*resumptive pronoun)	-0.37
Category 8 (*adjunct island)	-0.30
Category 9 (*embedded question)	-0.29
Category 3 (null)	-0.17
Category 6 (*relative clause)	-0.07
Category 10 (*complex NP)	0.03
Category 11 (sentential subject)	0.29

* An asterisk (*) means that the sentence is ungrammatical

Table 13: Average *b*-parameter value of each category in Unit O

3.6 Unit H

Within the framework of the Minimalist Program (Chomsky, 1995, 1998), overt movement is only allowed when it is motivated by the presence of a strong formal feature. In *wh*-questions, it is assumed that English and Japanese vary in the feature specifications of functional Category C (Complementizer) determining how their properties are realized. In other words, English has the features [+wh, +Q] in C, and they are both strong features which force *wh*-operator movement and subject-auxiliary inversion, as in (13). However, a [wh] feature in Japanese is not strong so that it does not need *wh*-operator movement, as in (14), although a [Q] feature has the same property as in English.

(13) What$_i$ are$_j$ you t_j reading t_i?

(14) Anata-wa nani-o yonde imasu ka?
You-Nom what-Acc reading are Q
"What are you reading?"

(15) *Which book did she ask John when he read?

In English, *wh*-questions are subject to the Subjacency condition, a constraint on *wh*-movement, as in (15). On the other hand, in Japanese, there is no constraint such as the Subjacency condition because *wh*-movement is not involved in *wh*-question formation (i.e. *wh*-in situ), as mentioned above.

Therefore, it is expected that if Japanese EFL learners make grammaticality judgments utilizing the same underlying operations as native speakers of English do, they will judge the *wh*-question constructions violating the Subjacency condition as well as the surface morphological properties of *wh*-question constructions at the same level of difficulty.

In the same line with Unit O, the aim of Unit H is to investigate the acquisition of the *wh*-question constructions, especially on whether Japanese EFL learners can acquire *wh*-movement or not, a written grammaticality judgment test was designed. This test contains 36 target sentences, which fall into the following three groups (with 12 categories in total).

(16) Category 1: Grammatical *wh*-questions (8 items)
e.g. *What did your mother want to talk about?*

(17) Category 2: Ungrammatical *wh*-questions without subject-auxiliary inversion (8 items)
e.g. *Who your favorite movie stars are?*

(18) Ungrammatical *wh*-questions violating the Subjacency condition
a. Category 3: *Wh*-question with an extraction out of a relative clause (2 items)
e.g. *What did they visit a shop which sold?*
b. Category 4: *Wh*-question with an extraction out of an embedded question (2 items)
e.g. *What did you wonder who would believe?*
c. Category 5: *Wh*-question with an extraction out of a complex NP (2 items)
e.g. *Which car did he believe the claim that John stole?*
d. Category 6: *Wh*-question with an extraction out of a sentential subject (2 items)
e.g. *Who(m) did that she went out with make him sad?*
e. Category 7: *Wh*-question with an extraction out of an adjunct island (2 items)
e.g. *Which car did they cross the street when John stopped?*

(19) Grammatical sentences from which an element is extracted
a. Category 8: Grammatical relative clauses (2 items)
e.g. *The boy who(m) I met yesterday broke the car.*
b. Category 9: Grammatical embedded questions (2 items)
e.g. *He asked me who had caused the car accident.*
c. Category 10: Grammatical complex NPs (2 items)
e.g. *She heard the news that her friend would get married.*
d. Category 11: Grammatical sentential subjects (2 items)
e.g. *To know that he is poor was no surprise to everyone.*
e. Category 12: Grammatical adjunct islands (2 items)
e.g. *You have to study for the exam before you play baseball.*

Table 14 indicates the average *b*-parameter values of the 12 categories. The surface morphological properties of *wh*-questions (Categories 1 and 2) are easier than violation of the Subjacency conditions (Categories 4, 5, 6 and 7). This means that Japanese EFL learners understand the *wh*-question constructions without using *wh*-movement.

Category	Average *b*-parameter value
Category 12 (adjunct island)	-1.66
Category 3 (*relative clause)	-1.36
Category 10 (complex NP)	-1.06
Category 1 (*wh*-question)	-0.95
Category 2 (*subject-auxiliary inversion)	-0.77
Category 9 (embedded question)	-0.69
Category 6 (*sentential subject)	-0.68
Category 4 (*embedded question)	-0.36
Category 8 (relative clause)	0.00
Category 5 (*complex NP)	0.12
Category 6 (sentential subject)	0.36
Category 7 (*adjunct island)	0.89

* An asterisk (*) means that the sentence is ungrammatical

Table 14: Average *b*-parameter value of each category in Unit H

4 Discussion

4.1 Implications for research methodology in SLA

When SLA researchers design a grammaticality judgment task, they presume several categories based on a linguistic theory, and prepare some target sentences in each of the categories. Behind this process they have the assumption, which is rarely questioned, that the items (target sentences) in one category are of similar difficulty. However, what turned out to be true, from our IRT-based estimation of difficulties, is that the items within a category in most of the test units discussed above had notably different *b*-parameter values. In some cases, such as in Unit S, the *b*-parameter values of most items in one category formed a band, and the bands were ordered in an expected way. Even in such cases, however, there were a couple of exceptions, and the width of each band, or the range of the *b*-parameter values, poses serious questions as to the source of variance in one category, or to the significance of a "category" derived from a linguistic theory in understanding learners' performance and/or interpreting items' relative difficulties. Therefore, in creating a grammaticality judgment task, an equating process such as the one we used in this study would be a necessary first step so that one can accumulate a large enough number of items with approximate *b*-parameter values to form a category with high reliability. We hope that our present study strikes a note of warning against preparing a grammaticality judgment task without carefully considering and controlling the difficulty and the discriminating power of each item.

Moreover, as mentioned in Section 3.2, the distributions of the average *b*-parameter values of the categories in Unit Y and Units KM were substantially different. The categories in Unit Y were scattered more sparsely on the *b*-parameter scale than those in Units KM. Furthermore, the average *b*-parameter values of Categories A, B and C in Unit Y were widely different from those in Units KM although the items in Categories A to F were identical in both test units. These were unexpected results because IRT, which was supposed to be a sample-free and test-free method of analysis, should assign approximate *b*-parameter values to each of the identical items. However, the problem can be understood if we consider the relatively small numbers of participants who took both tests (see Section 2.2) and different environments in which the identical items were distributed in Unit Y and Units KM: Unit Y was a single test unit with 48 items whereas Units KM were separate test units, each consisting of 36 test items. These factors may have exerted an unfavorable influence on the process of data analysis utilizing IRT in the present study, thus implying that learners may give different grammaticality judgments to identical test items placed in different contexts. Therefore, when conducting grammaticality judgment tasks, it is important to ensure large enough numbers of items for any given linguistic categories and large enough numbers of participants who take the tasks. In addition, it is preferable to prepare such a research design, when possible, as to duplicate the same task with different participants.

The present study has been focusing only on the *b*-parameter values so far because our main interests are in the learners' developmental aspects, which could presumably be inferred from examining the difficulty orders of the test items on the *b*-parameter scale. It is worthwhile, however, to take a glance at the *a*-parameter (item discrimination parameter) values we obtained by utilizing the two-parameter logistic model. The usual range of *a*-parameter values is from 0 to 2 (Hambleton et al., 1991). The lower the *a*-parameter value of a test item is, the less likely it is that the item can successfully discriminate good from poor learners. Generally speaking, the *a*-parameter value of a test item should be above 0.3. As with *b*-parameter values, we can observe wide ranges of the *a*-parameter values, as displayed in Table 15 below.

	Number of items	Minimum *a*-parameter	Maximum *a*-parameter	Number of items lower than 0.3
Unit S	48	0.31	1.98	0
Unit Y	48	0.24	1.13	4
Units KM	72	0.29	1.54	1
Unit N	48	0.16	1.61	11
Unit G	48	0.23	1.42	2
Unit O	45	0.2	1.28	6
Unit H	36	0.22	0.94	4

Table 15: Ranges of the *a*-parameter values in each test unit

Considering the number of the items with *a*-parameter values lower than 0.3 in each test unit, we can safely say that each test unit, on the whole, has enough discriminatory power. On the other hand, we also observed two items in Unit N and one item in Units KM which were not assigned *a*-parameter values in our IRT-based analysis because they had abnormally high *b*-parameter values. By checking the appropriateness of *a*-parameter values as well as *b*-parameter values in this manner, we can eliminate or replace inappropriate test items, and improve the overall quality of a given grammaticality judgment task.

4.2 Overall development of the learners' grammatical competence

By analyzing 1,185 participants' data using IRT, we obtained Figure 3, which shows the distribution of the average values of all the categories on the *b*-parameter scale. As you go down to the bottom of Figure 3, categories have higher average *b*-parameter values, which means the items belonging to those categories are considered more difficult.

Being the first attempt to apply the IRT over a range of grammatical features, the present study was more hypothesis-generating than hypothesis-testing. Of particular interest for further investigation is the influence of the L1. The difficulty order with the Unit S items indicates that the items with the conflicting cues were indeed difficult for the Japanese EFL learners, who demonstrate their L1 cue dependency on the semantic/pragmatic cues in agent identification. This cue dependency appears to influence their grammaticality judgment of the sentences with the psych verbs (Unit G) when the animacy of the NPs is controlled. Similar influence of the animacy is observed with other verb-related items, such as those in Units Y, KM and N. Thus, it could be hypothesized that the L1 cue dependency, which is a performance factor, may be playing a crucial role in the grammaticality judgment task with other types of verbs.

b	Unit S	Unit Y		Unit KM		Unit N		Unit G		Unit O		Unit H	
-1.7										OC12	adjunct island	HC12	adjunct island
-1.6													
-1.5		YC-b YC-g	NP+UNERG NP+ERGpsv										
-1.4	CONV.					NC-Cto						HC3	*relative clause
-1.3						NC-C	ppDat						
-1.2				KMC-i KMC-k	NP+ERGpsv NP+VTRNS+NP	NC-Ato NC-Cfor NC-Dfor							
-1.1		YC-a	NP+UNACC			NC-D	NonDative-pp					HC10	complex NPs
-1		YC-h	NP+ERG+NP			NC-Dto				OC14 OC13	complex NP embed-ded q.	HC1	*wh*-question
-0.9				KMC-b KMC-l	NP+UNERG NP+VTRNSpsv+NP	NC-Eto							
-0.8												HC2	*subj-aux inv.
-0.7	NEUT.			KMC-h	NP+ERG+NP	NC-Bfor NC-Ffor		EO3 ES3	HEX-EOpsv-ppITH ITH-ESpsv-ppHEX	OC2	that	HC6 HC9	*sententialembedded q
-0.6		YC-f	*NP+UNERG+NP	KMC-a	NP+UNACC					OC1 OC7	*wh*-operator *sententialsubject		
-0.5		YC-e	*NP+UNACC+NP	KMC-d KMC-j	*NP+UNERGpsv *NP+VTRNS					OC4	*doubly-filled		
-0.4				KMC-g	NP+ERG			EO2 ES1	ITH-EO-HEX HEX-ES-HTH	OC5	*resumptive pron.	HC4	*embedded q.
-0.3		YC-d	*NP+UNERGpsv	KMC-e	*NP+UNACC+NP	NC-A	DITRANS-ppDat	EO1	HTH-EO-HEX	OC8 OC9	*adjunct island *embedded q.		
-0.2				KMC-f	*NP+UNERG+NP			ES2 ES4	HEX-ES-ITH *ITH-ES-HEX	OC3	null		
-0.1				KMC-c	*NP+UNACCpsv	NC-B	*DITRANS-ppDat	**EO** **ES**		OC6	*relative		
0										OC10	*complex NP	HC8	relative clauses
0.1						NC-F	*Psv-ppDat					HC5	*complex NP
0.2						NC-E	Psv-ppDat						
0.3	CONF.	YC-c	*NP+UNACCpsv							OC11	sententialsubject		
0.4						NC-Bto		EO4 EO5	*HEX-EO-ITH *ITH-EOpsv-ppHEX			HC11	sententialsubjects
0.5						NC-Afor							
0.6													
0.7													
0.8								ES5	*HEX-ESpsv-ppITH				
0.9						NC-Fto						HC7	*adjunct island
1													

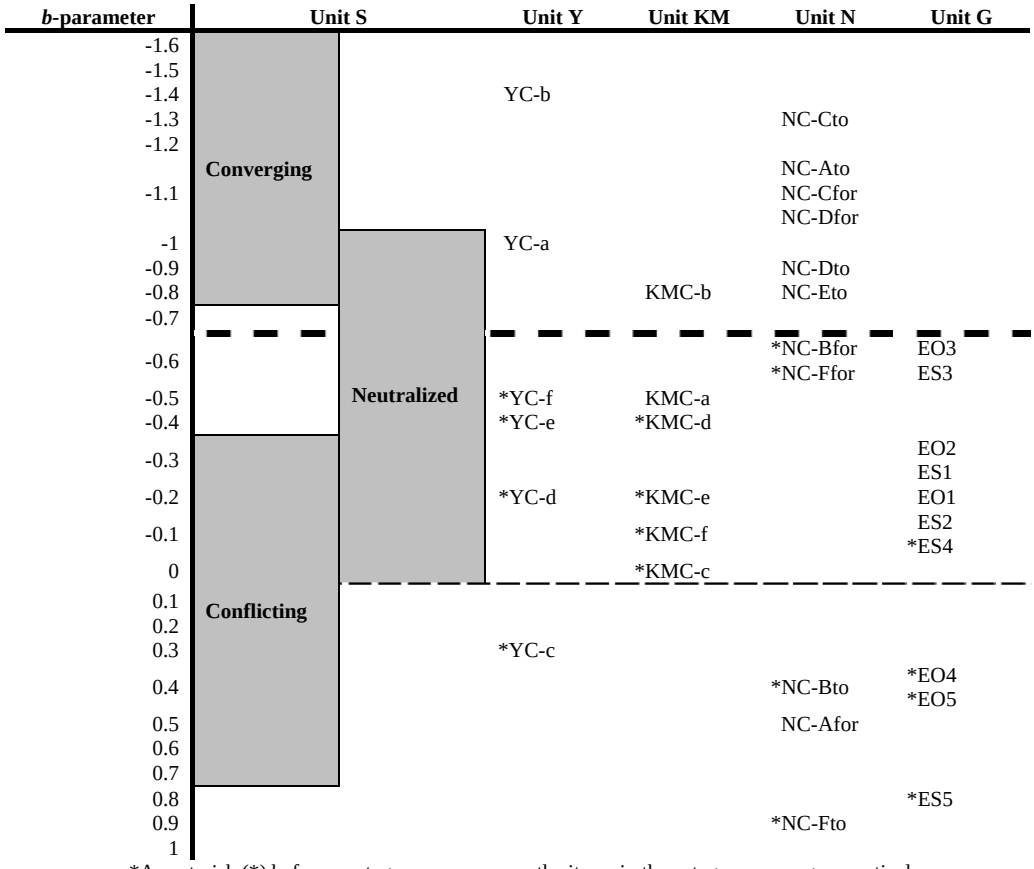
* "HC12," for example, stands for "Unit H, Category 12." For other abbreviations, see Section 3.

Figure 3: Distribution of the average values of all the categories on the *b*-parameter scale

We can observe in Figure 3 that Units N and H display wider ranges of the *b*-parameter values than the other units. As reported in Section 4.1, two items in Unit N were too difficult to assign *b*-parameter values to. The items both belong to the same category, Category E (for-datives), which is not shown in Figure 3. The wide ranges of the *b*-parameter values in Units N and H may suggest that different grammatical features require a different amount of time or effort before learners acquire them or, at least, are able to correctly judge the test items pertinent to the features.

The test units used in the present study can roughly be divided into three groups: a performance-based test (Unit S), verb-related tests (Units Y, KM, N and G) and *wh*-movement related tests (Units O and H). Regarding the verb-related tests, the categories consisting of grammatical items bear lower *b*-parameter values than ungrammatical items. In other words, learners tend to judge grammatical items as such before being able to reject ungrammatical items. This may imply that learners tend to formulate a rough and broad rule on a given grammatical feature first and then narrow down the extent to which the rule can be properly applied. For instance, we observed in the results of Unit G that the learners were inclined to accept sentences when the "Experiencer" role is assigned to the human NP, regardless of the verb classes or the human NP's position in the sentence. The learners will, at a later stage, have to adjust the overgeneralized rule and start to take the appropriate verb classes or NP's positions into account.

In addition, an interesting observation can be made with regard to the verb-related tests in Figure 3. It is possible to roughly divide all the categories in the four units into three groups according to their average *b*-parameter values, as shown in Figure 4.



*An asterisk (*) before a category name means the items in the category are ungrammatical.
**Only Categories A to F in Units U and KM are included in the figure because they are relevant to the current discussion.

Figure 4: Distribution of the categories in the verb-related units on the *b*-parameter scale

The first group of categories bear average *b*-parameter values of -0.7 or lower. The categories belonging to this group consist of "grammatical" items with relatively low difficulty. The second group of categories have average *b*-parameter values of -0.7 to 0. They consist of "ungrammatical" items except the categories belonging to Unit G. It is within this range that categories in Unit G first appear on the *b*-parameter scale, and all categories in Unit G in this group except one (ES4) are "grammatical". This may suggest that the acquisition of psych verbs, the target grammatical feature in Unit G, would start later than the grammatical features in the other verb-related test units. The third group of categories bear average *b*-parameter values of 0 or higher. The categories belonging to this group are all "ungrammatical" except one in Unit N (NC-Afor), and they can be regarded as relatively difficult.

Interestingly, the transitions from one group to the next among the three groups roughly correspond to the difficulty order of the categories in Unit S, as also shown in Figure 4. In other words, it can be argued that when the learners can correctly respond to the items in Unit S with the converging cue, they can also make correct judgments about the items in the first group. Next, when the learners start to correctly process the items in Unit S with the neutralized cue, they can also start to properly judge the items in the second group. Finally, it is shortly after the learners can correctly respond to most of the items in Unit S with the conflicting cue that they can start to reject ungrammatical items in the third group. Although this is only a rough observation, it can be maintained that, as the learners move to the next stage in terms of their cue dependency, they are more likely to make correct judgments about verb-related items. In other words, the learners' sensitivity to the word order in a sentence may be associated with the overall acquisition of verb-related constructions.

5 Conclusion

The present study has attempted to apply IRT to grammaticality judgment tasks in order to gain insights to both research methodology in SLA and the acquisition processes of Japanese EFL learners. First, it can be maintained that a close examination of *a*-parameter and *b*-parameter values of test items strengthens the validity and reliability of a given grammaticality judgment task used in SLA research. This requires researchers to prepare large enough numbers of test items allotted to each grammatical category in a test so that inappropriate test items can be eliminated or replaced at a later stage. Secondly, our IRT-based analysis produced the difficulty orders of several grammatical features, which enables us to assume the acquisition processes of different grammatical features on the same scale. Although the present study is still tentative and descriptive, some findings were reported in Section 4.

In future research, we will need to make reference to the mechanism of the acquisition processes by also considering such factors as the animacy effect of the subject of a test sentence, and the participant role which the noun phrase in a test sentence bears. Furthermore, many items in the verb-related test units (Units Y, KM, N and G) involve NP-movement because these units contain grammatical and ungrammatical passive sentences; whereas Units O and H are concerned with *wh*-movement. It would be worthwhile to investigate the way in which the acquisitions of NP-movement and *wh*-movement are related in the development of the grammatical competence. Moreover, further research on additional grammatical features, utilizing IRT again, will clarify an even wider range of developmental aspects of learners' grammatical competence.

Notes

[1] The present study is supported by Grants-in-Aid for Scientific Research 1348064 (2001–2003) and 16320078 (2004–2006) from Japan Society for Promotion of Science. The original version of this paper was presented at the Second CLS International Conference (CLaSIC 2006), December 7–9, 2006, Holiday Inn Atrium, Singapore.

References

Bachman, L., & Cohen, A. (1998). *Interfaces between second language acquisition and language testing research.* Cambridge: Cambridge University Press.

Baker, R. (1997). *Classical test theory and Item Response Theory in test analysis* (Language Testing Update, Special Report No. 2). Lancaster: Centre for Research in Language Education.

Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L.R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 173–218). Cambridge: Cambridge University Press.

Belletti, A., & Rizzi, L. (1988). Psych-verbs and θ-theory. *Natural Language and Linguistic Theory*, *6*, 291–352.

Bresnan, J. (1982). *The mental representation of grammatical relations*. Cambridge, MA: MIT Press.

Bresnan, J. (1998). *Pidgin genesis in optimality theory*. Paper Presented at the LFG98 Conference. The University of Queensland, Brisbane, Australia.

Bresnan, J. (2001). *Lexical-functional syntax*. Malden, MA: Blackwell.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, N. (1998). *Minimalist inquiries: The framework* (MIT Occasional Papers in Linguistics, No. 15).

Crocker, L. & Algia, J. (1986). *Introduction to classical & modern test theory.* Belmont: Wadsworth.

Fukui, N., & Takano, Y. (2000). Nominal structure: An extension of the symmetry principle. In P. Svenoius (Ed.), *The Derivation of VO and OV* (pp. 219–254). Amsterdam: John Benjamins.

Gass, S. (1984). The empirical basis for the universal hypothesis in interlanguage studies. In A. Davis, C. Criper & A.P.R. Howatt (Eds.), *Interlanguage: Papers in honor of S. Pit Corder* (pp. 3–24). Edinburgh: Edinburgh University Press.

Gass, S. (1986). An interactionist approach to L2 sentence interpretation. *Studies in Second Language Acquisition*, *8*, 19–37.

Gass, S. (1987). The resolution of conflicts among competing systems: A bidirectional perspective. *Applied Psycholinguistics*, *8*, 329–350.

Gass, S. (1989). How do learners resolve linguistic conflicts? In S.M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 183–199). Cambridge: Cambridge University Press.

Grimshaw, J. (1990). *Argument structure*. Cambridge, MA: MIT Press.

Hambleton, R., Swaimnathan, H., & Rogers, H.J. (1991). *Fundamentals of Item Response Theory.* Newbury Park: SAGE.

Harrington, M. (1987). Processing transfer: Language-specific processing strategies as a source of interlanguage variation. *Applied Psycholinguistics*, *8*, 351–377.

Heilenman, L.K., & McDonald, J.L. (1993). Processing strategies in L2 learners of French: The role of transfer. *Language Learning*, *43*, 507–557.

Hirakawa, M. (2003). *Unaccusativity in second language Japanese and English*. Tokyo: Hituzi Syobo.

Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.

Kilborn, K., & Cooreman, A. (1987). Sentence interpretation strategies in adult Dutch-English bilinguals. *Applied Psycholinguistics*, *8*, 415–431.

Koda, K. (1993). Transferred L2 strategies and L2 syntactic structure in L2 sentence comprehension. *The Modern Language Journal*, *77*, 490–500.

Kuno, S. (1973). *The structure of the Japanese language*. Cambridge, MA: MIT Press.

MacWhinney, B. (1987a). Applying the Competition Model to bilingualism. *Applied Psycholinguistics*, *8*, 315–327.

MacWhinney, B. (1987b). The Competition Model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Hillsdale, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. (1992). Transfer and competition in second language learning. In R.J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 371–390). Amsterdam: Elsevier.

MacWhinney, B, Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German & Italian. *Journal of Verbal Learning and Verbal Behavior*, *23*, 127–150.

McDonald, J.L. (1987). Sentence interpretation in bilingual speakers of English and Dutch. *Applied Psycholinguistics*, *8*, 379–413.

McDonald, J.L., & Heilenman, L.K. (1992). Changes in sentence processing as second language proficiency increases. In R.J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 325–336). Amsterdam: Elsevier.

Nakano, M. (2000). An experimental study of a relationship between Lexical Functional Grammar and learner-language data. In AILA '99 Tokyo Organizing Committee (Ed.), *Selected Papers from AILA'99 TOKYO* (pp. 344–361). Tokyo: Waseda University Press.

Nakano, M., Sugino, N., Ohba, H., Yamakawa, K., & Shimizu, Y. (2005). An analysis of grammatical judgment test: Dative constructions, their passive forms, unaccusative and unergative constructions. In *Proceedings of the 9th Conference of Pan-Pacific Association of Applied Linguistics* (pp. 386–394). Seoul: Namseoul University.

O'Grady, W. (2001). A linguistic approach to the study of language acquisition. *Journal of Pan-Pacific Association of Applied Linguistic*s, *5*, 57–71.

Ohba, H., Sugino, N., Nakano, M., Yamakawa, K., Shimizu, Y., & Kimura, S. (2006). The Development of Grammatical Competence of Japanese EFL Learners: Focusing on Relative Clause Constructions. *Journal of Pan-Pacific Association of Applied Linguistics*, *10*, 73–87.

Ohba, H., Yamakawa, K., Sugino, N., Shimizu, Y., & Nakano, M. (2005). *The acquisition of* wh-*questions by adult Japanese EFL learners*. Poster presented at the 10th conference of Pan-Pacific Association of Applied Linguistics, The University of Edinburgh, Scotland, UK.

Oshita, H. (1997). *The unaccusative trap: L2 acquisition of English intransitive verbs*. Unpublished Ph.D. dissertation, University of Southern California, California.

Oshita, H. (2000). *What is happened* may not be what appears to be happening: A corpus study of "passive" unaccusatives in L2 English. *Second Language Research*, *16*, 293–324.

Oshita, H. (2001). The unaccusative trap in second language acquisition. *Studies in Second Language Acquisition, 23*, 279–304.

Perlmutter, D.M. (1978). Impersonal passives and the unaccusative hypothesis. *Proceedings of the Berkeley Linguistics Society, 4*, 157–189.

Sasaki, Y. (1991). English and Japanese interlanguage comprehension strategies: An analysis based on the Competition Model. *Applied Psycholinguistics*, *12*, 47–73.

Sasaki, Y. (1994). Paths of processing strategy transfers in learning Japanese and English as foreign languages: A Competition Model approach. *Studies in Second Language Acquisition*, *16*, 43–72.

Sato, Y. (2002). Eigo Shinri Doushi no Shuutokujou no Mondaiten. [Learning problems with English psych verbs.] *Language Education and Technology*, *39*, 73–91.

Sato, Y. (2003). Japanese learners' linking problems with English psych verbs. *Reading Working Papers in Linguistics*, *7*, 125–144.

Shimizu, Y., Kimura, S., Sugino, N., Yamakawa, K., Ohba, H., & Nakano, M. (2003). Eibunpou Nouryoku Tesuto no Datousei, Shinraisei no Kenshou to Shin Eigo Bunpou Nouryoku Test Measure of English Grammar (MEG) [The validity and reliability of standardized tests and compilation of "Measure of English Grammar (MEG)"]. *Policy Science*, *10*, 59–68.

Shimizu, Y., Yamakawa, K., Sugino, N., Ohba, H., & Nakano, M. (2006). Developing a sample-free grammatical proficiency test for SLA research. In *Proceedings of the 10th Conference of Pan-Pacific Association of Applied Linguistics* (pp. 227–235). Edinburgh: Edinburgh University.

Shomura, Y. (1996). Perspectives on pedagogic grammar in English language teaching: A study of the acquisition of ergatives by Japanese learners. *Annual Review of English Learning and Teaching*, *1*, 17–35.

Shomura-Isse, Y. (2005). Argument selection with psych predicates of adnominal clauses in L2 English by Japanese-speaking learners. *JACET Bulletin*, *40*, 29–39.

Sorace, A. (1997). Acquiring linking rules and argument structures in a second language. In L. Eubank, L. Selinker & M. Sharwood Smith (Eds.), *The current state of interlanguage* (pp. 153–175). Amsterdam: John Benjamins.

Sugino, N., Nakano, M., Ohba, H., Kimura, S., Yamakawa, K., & Shimizu, Y. (2005). The development of grammatical competence of Japanese EFL learners: Focusing on dative alternation. In *Proceedings of the 9th Conference of Pan-Pacific Association of Applied Linguistics* (pp. 322–331). Seoul: Namseoul University.

Sugino, N., Yamakawa, K., Ohba, H., Nakano, M., & Shimizu, Y. (2006). *The acquisition of psych verbs by Japanese EFL learners: From the perspective of cue dependency*. Paper presented at the 11th Conference of Pan-Pacific Association of Applied Linguistics, Kangwon National University, Chuncheon, Korea.

Takeda, K. (1999). *Multiple headed structures*. Unpublished doctoral dissertation, University of California, Irvine.

White, L., Brown, C., Bruhn-Garavito, J., Chen, D., Hirakawa, M., & Montrul, S. (1999). Psych verbs in second language acquisition. In E. Klein & G. Martohardjono (Eds.), *Second language acquisition and generative grammar* (pp. 171–196). Amsterdam: John Benjamins.

Yamakawa, K., Sugino, N., Kimura, S., Nakano, M., Ohba, H., & Shimizu, Y. (2003). The development of grammatical competence of Japanese EFL learners: Focusing on unaccusative/unergative verb. *Annual Review of English Language Education in Japan*, *14*, 1–10.

Yamakawa, K., Sugino, N., Kimura, S., Nakano, M., Ohba, H., & Shimizu, Y. (2005). Nihonjin Eigo Gakushusha ni yoru Hitaikaku Doushi to Hinoukaku Doushi no Shuutoku [The acquisition processes of unaccusative/unergative verbs by Japanese EFL learners]. *JACET Chugoku-Shikoku Chapter Research Bulletin*, *2*, 91–110.

Yip, V. (1995). *Interlanguage and learnability*. Amsterdam: John Benjamins.

Zobl, H. (1989). Canonical typological structures and ergativity in English L2 acquisition. In S.M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 203–221). Cambridge: Cambridge University Press.