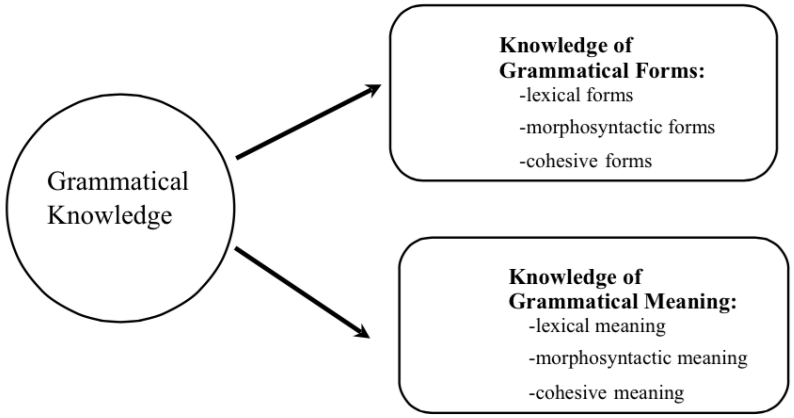
Form, Meaning, and Gender Bias in Dichotomous Grammar Items: A Many-Facet Rasch Analysis of a Grammar Placement Test

by Qie Han

Introduction

The purpose of the study is to look at how examinees perform on different types of grammar test items, i.e., Form (F) vs. Meaning (M) items, and if gender, as a background variable, has moderated examinees' performance on different types of items in this administration of the CEP grammar placement test. The conceptualization of grammar in this study is built upon Pupura's (2004) comprehensive model of grammatical knowledge, which is categorized as grammatical form and grammatical meaning, as shown in Figure 1 below.

Figure 1 Theoretical Model of Grammatical Knowledge



Research Questions

This study addresses three research questions:

- First, to what extent does the examinee ability facet contribute to score variance?
- Second, to what extent do all the grammar items differ in difficulty, especially regarding the difference between the items that focus on form and those that focus on meaning?
- Third, has gender caused bias that interacts with the other two facets (examinee and item) in determining test-takers' performance in this administration of the grammar placement test?

Method

Table 1

Item Coding for the Grammar Placement Test

Focus	Item No.
Form (F)	13, 14, 16, 17, 19, 20, 22, 24, 26, 27, 28, 29, 30, 31, 32, 34, 35, 36, 37, 38
Meaning (M)	15,18, 21, 23, 25, 33, 39, 40, 41, 42, 43

Participants include a non-randomized, intact group (n =107) of students of the Communicative English Program (CEP) grammar placement test for the 2012 Summer A semester. The CEP is an English language school and language laboratory run by TC's TESOL and Applied Linguistics program. The grammar placement test (see Appendix A) was composed of 31 items, and each item was graded dichotomously as 0/1. Each grammar test item was coded according to its focus on what it was designed to measure, meaning (M) or form (F), based on Purpura's (2004) model of grammatical knowledge (Table 1).

Analyses and Results

This study uses SPSS to generate and report descriptive statistics, which indicate the central tendency, variability, distributional characteristics, and internal-consistency reliability (r) for the entire grammar test (see Table 2).

The Facets program calibrates the examinees and test items so that both facets are positioned on the same scale, creating a single frame of reference for interpreting the results from the analysis (see Figure 2). That scale is in log-odds units, or "logits," which, under the model,

Table 2

Descriptive Statistics for the Grammar Test

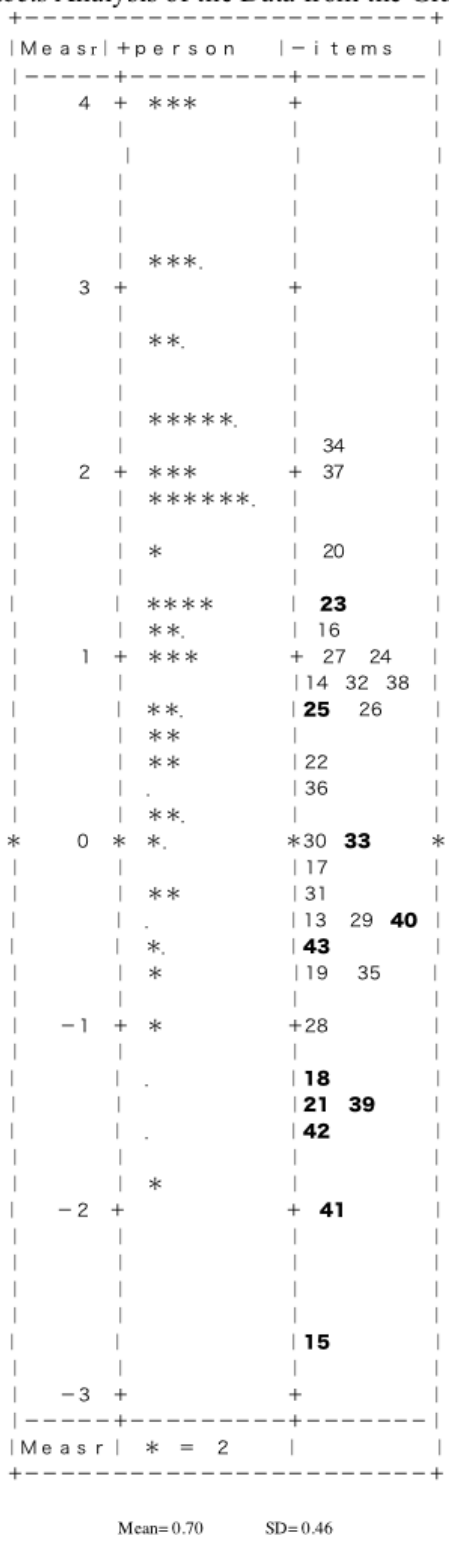
	n	k	Mean	Median	SD	Skewness	Kurtosis	r
Grammar	107	31	21.82	23	6.16	-0.72	-0.25	0.88

constitute an equal-interval scale with respect to appropriately transformed probabilities of responding in particular test items. The first column in the map displays the logit scale. The second column displays estimates of examinee proficiency on the grammar test—the higher an examinee appears on the scale, the more proficient he or she is. Each star represents 2 examinees, and a dot represents 1 examinee. The third column compares the 31 items that appeared on the grammar placement test in terms of their relative difficulties. Items appearing higher in the column were more difficult for examinees to respond correctly to than items appearing lower in the column.

Apart from the general overview of all the items, item difficulty measures for form-focused and meaning-focused items are compared. In Figure 2, the grammar items that test meaning are

emboldened. It can be seen that the Meaning (M) items were generally easier to answer than the Form (F) items in this grammar placement test. Besides, the juxtaposition of both the examinee and the item spreads shows that the item spread is narrower than the examinee one, and that there are no items that are difficult enough to match the examinees on the higher end of the scale. This means that the items in this test are generally too easy for the more advanced examinees.

Figure 2
Map from the *Facets* Analysis of the Data from the Grammar Place Test



To examine the assumption for how different groups of items function differently with both genders of examinees on this grammar placement test, interaction/bias analysis is conducted between gender and item facets for both types of items (form or meaning). Bias size is measured in log-odds units, or logits, relative to overall measures. Bias direction that is positive (+) means the item is easy for a certain gender group, while negative (–) bias direction means the item is difficult, or is biased against, a certain gender group. If the bias probability is less than 0.05, the bias is considered significant, which means a gender bias

truly exists. According to Table 3 (M) and Table 4 (F), none of the items show bias probability less than 0.05. That indicate that the gender bias for these items are not statistically significant, and that the results generated from this test can be considered as unbiased against either gender group.

Item No.
13
14
16
17
19
20
22
24
26
27
28
29
30
31
32
34
35
36
37
38

* Only bia

Table 3

Measure, bias size and probability for the Meaning (M) items

Item No.	Measure	Bias Size		Target Measure		Bias Probability
		1. male	2. female	1. male	2. female	
15	-2.69	1.00	0.00	-3.69	-2.32	
18	-1.23	-0.14	-0.13	-1.09	-1.31	
21	-1.46	0.04	-0.11	-1.50	-1.46	
23	1.32	0.25		1.07	1.42	
25	0.65	-0.78	0.33	1.43	0.33	
33	-0.04	0.15		-0.19	0.00	
39	-1.46	-0.37	0.05	-1.09	-1.64	
40	-0.38	-0.43	0.12	0.05	-0.57	
41	-2.03	-0.54	0.12	-1.50	-2.32	
42	-1.59	-0.50	0.32	-1.09	-1.83	
43	-0.45	-0.07	-0.19	-0.46	-0.57	

* Only bias probability that is less than 0.05 is shown.

Discussion and Conclusion

Examine

e proficiency measures show a wide logit spread and an appropriate separation of examinees in terms of levels of proficiency. Item difficulty measures show a narrower logit spread than the examinee proficiency measures, and a lack of suitable items for the more advanced examinees in this test. Besides, Meaning items are found to be generally easier to answer than the Form (F) items in this test. To achieve a better balance in item difficulty between form and meaning, it is suggested that more advanced knowledge of grammatical meaning, i.e., morphosyntactic, cohesive meaning, and vocabulary, should be included in the test.

Gender is not found to interact with the items in determining examinees' performance in grammar. That indicate that the results generated from this grammar test can generally be considered as unbiased against either gender group, which meets the assumption that gender should not interacts with items across all examinees to cause variation in test scores as a bias.

References

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford, UK: Oxford University Press.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied linguistics*, *1*(1), 1–47.

Grotjahn, R. (1987). On the methodological basis of introspective methods. In C. Færch & G. Kasper (Eds.), *Introspection in second language research* (pp. 54–81). Clevedon, UK: Multilingual Matters.

McNamara, T. F. (1996). *Measuring second language performance.* London and New York: Longman.

Oller, J. (1979). *Language tests in schools: A pragmatic approach*. London: Longman.

Purpura, J. (2004). *Assessing grammar*. Cambridge, UK: Cambridge University Press.

Qie Han is an Ed.D. student at Teachers College, Columbia University, where she studies applied linguistics and has a special interest in second language assessment. This article is based on the research and course project she did for an internship class at the Community English Program (CEP) at TC. <(qh2130@tc.columbia.edu) >