A Rasch Analysis of TOEIC Reading Items

Atsuko NISHITANI*

Abstract

This study investigated the difficulty order of 100 items in the reading section of the TOEIC test (i.e., Test 1 from TOEIC Official Test-Preparation Guide Vol. 4) using a Rasch analysis in the hope of providing implications for decision making in educational institutions that offer TOEIC classes, since knowledge of the difficulty order of such items can be used to plan more efficient curricula. The participants were 83 university students, who enrolled in the TOEIC classes offered by Faculty of Foreign Languages at Kyoto Sangyo University. After the difficulty order of the 100 items was displayed according to their Rasch difficulty estimates, the items of Parts 5 and 6 were examined to identify what grammar point it was testing, and the difficulty order within the same grammatical category was examined. The reading comprehension questions (i.e., Part 7) was grouped into seven categories based on their question types and their difficulty order was also examined. However, the participants involved and the items used in this study were limited and insufficient, and these limitations should be addressed in future research.

1. Introduction

Recently a lot of universities have introduced TOEIC classes in their English curricula to meet students' immediate needs to get high scores on the test because an increasing number of companies claim that they consider TOEIC scores when hiring new employees. In line with this trend, Kyoto Sangyo University has also started offering TOEIC classes as required courses. This study investigated the difficulty order of 100 items in the TOEIC reading section using a Rasch analysis in the hope of providing implications for decision making in educational institutions that offer TOEIC classes, since knowledge of the difficulty order of such items can be used to plan more efficient curricula.

* Institute for General Education, Kyoto Sangyo University

2. Participants

A total of 83 students, who were taking the TOEIC classes in the Special English Program of the Faculty of Foreign Languages, Kyoto Sangyo University, participated in this study. Of the 83 students, 20 were male and 63 were female. Twenty-nine were first-year students, 39 were second-year students, 14 were third-year students, and 1 was a fourth-year student when the test was administered. Fifty-three students were majoring in English or other faculties that have a curriculum stressing English. The TOEIC scores they had before taking this test ranged from 305 to 830 ($n = 32$, $M = 583.44$, $SD = 129.78$), and their TOEFL ITP scores ranged from 380 to 513 ($n = 36$, $M = 448.56$, $SD = 35.20$). Their TOEIC scores after taking the course ranged from 230 to 920 ($n = 72$, $M = 567.43$, $SD = 144.11$). The participants were asked for permission to use the data for research purposes, and all of them gave their permission to do so.

3. Instrumentation

In order to investigate the difficulty of TOEIC reading items, the reading section (i.e., Parts 5 to 7) of Test 1 from TOEIC Official Test-Preparation Guide Vol. 4 was administered. The test was in a multiple-choice format with one correct answer and three distractors.

4. Procedures

The advanced-level classes administered the whole reading section on the first day of class, but the intermediate and the elementary-level classes administered Parts 5 and 6 on the first day and Part 7 later in the semester. The time allotted to the advanced-level students was 75 minutes, which is exactly the same as the actual TOEIC test. The time allotted for Parts 5 and 6 in the intermediate and elementary-level classes was set at 30 seconds per item (i.e., a total of 26 minutes) and Part 7 was 60 seconds per item (i.e., a total of 48 minutes), which is often said to be the maximum response time for the TOEIC test. However, the primary purpose of this study is to investigate the item difficulty, and thus it was more important that students answer as many items as possible than we strictly keep to the time limit. Therefore, I asked the teachers to give additional five minutes or so if they found many of the students had too many items unanswered.

The Rasch analysis was conducted using the WINSTEPS computer software, version 3.75.0 (Linacre, 2013), and the infit mean-square statistics for each item were examined to determine whether they were in the range of the mean $\pm$ twice the standard deviation of the mean square statistic. According to McNamara (1996), infit statistics are the most informative, given that they are focused on the fit of the most typical observations, and "for n sizes of 30 or more, the [acceptable] range is the mean $\pm$ twice the standard deviation of the mean square statistic" (p. 181).

5. Results and Discussion

In order to examine the construct validity of the test, a Rasch analysis was conducted, and dimensionality and item fit to the model were assessed. The mean Rasch person ability estimate was .17 (SD = .58; SE = .26), and the person reliability estimate and the person separation index were .80 and 2.02, respectively. The mean Rasch item difficulty estimate was .00 (SD = 1.23; SE = .29), and the item reliability estimate and the item separation index were .94 and 3.89, respectively. Then the infit mean square statistic was examined, and the criterion of .73 to 1.15, the mean ± twice the standard deviation of the mean square statistic, was used to identify misfitting items. As Table 1 shows, only one item, #128, misfit the model.

Table 1. *Rasch Item Statistics*

Item	Measure	S.E.	Infit		Outfit		Pt-Measure Correlation
			MNSQ	ZSTD	MNSQ	ZSTD	
41R101	–1.53	0.30	0.99	0.00	1.11	0.50	0.18
41R102	–2.97	0.52	0.95	0.00	1.01	0.20	0.21
41R103	–0.66	0.24	1.00	0.10	1.01	0.10	0.25
41R104	–0.64	0.25	1.07	0.70	1.05	0.40	0.15
41R105	2.40	0.36	0.93	–0.20	0.95	0.00	0.27
41R106	2.17	0.33	1.12	0.60	1.37	1.20	–0.09
41R107	–2.35	0.40	1.01	0.20	0.88	–0.20	0.19
41R108	–2.52	0.43	0.97	0.00	0.84	–0.20	0.23
41R109	0.05	0.23	1.00	0.10	0.99	–0.10	0.26
41R110	0.41	0.23	0.91	–1.50	0.91	–1.20	0.40
41R111	–1.53	0.30	0.85	–0.80	0.70	–1.30	0.50
41R112	0.15	0.23	0.87	–2.50	0.85	–2.30	0.48
41R113	0.15	0.23	1.03	0.60	1.08	1.20	0.19
41R114	0.52	0.23	0.92	–1.20	0.91	–1.20	0.39
41R115	0.68	0.23	1.06	0.80	1.13	1.40	0.13
41R116	0.05	0.23	1.09	1.60	1.13	1.80	0.10
41R117	0.73	0.23	1.11	1.40	1.10	1.10	0.08
41R118	0.20	0.23	0.98	–0.30	0.97	–0.30	0.30
41R119	–0.55	0.24	0.96	–0.40	0.92	–0.70	0.34
41R120	0.68	0.23	0.97	–0.40	0.95	–0.50	0.31
41R121	–0.32	0.23	1.06	0.80	1.11	1.30	0.15
41R122	–0.16	0.23	1.02	0.30	1.01	0.20	0.23
41R123	–0.49	0.24	0.94	–0.70	0.92	–0.70	0.36

41R124	3.30	0.52	0.98	0.10	1.60	1.10	0.02
41R125	−0.38	0.24	0.95	−0.60	0.93	−0.70	0.35
41R126	−1.53	0.30	0.89	−0.50	0.89	−0.40	0.38
41R127	0.79	0.24	1.11	1.30	1.12	1.10	0.07
41R128	0.84	0.24	1.26	2.90	1.41	3.50	−0.23
41R129	1.96	0.31	1.01	0.10	0.94	−0.10	0.21
41R130	0.72	0.24	0.93	−0.90	0.91	−0.90	0.37
41R131	−0.49	0.24	0.94	−0.70	0.91	−0.90	0.37
41R132	1.14	0.25	1.01	0.10	1.14	1.00	0.19
41R133	−1.05	0.26	0.92	−0.60	0.81	−1.10	0.42
41R134	1.14	0.25	1.06	0.60	1.06	0.40	0.14
41R135	0.15	0.23	0.96	−0.80	0.94	−0.90	0.34
41R136	0.89	0.24	1.04	0.50	1.03	0.30	0.18
41R137	−0.29	0.23	1.06	0.80	1.07	0.80	0.16
41R138	0.83	0.24	1.01	0.20	1.09	0.80	0.20
41R139	0.06	0.23	1.02	0.40	1.02	0.30	0.23
41R140	1.25	0.26	1.15	1.20	1.22	1.40	−0.04
41R141	0.29	0.23	1.02	0.30	1.05	0.70	0.22
41R142	−2.31	0.40	0.98	0.10	0.91	−0.10	0.21
41R143	0.54	0.23	1.00	0.00	1.00	0.00	0.26
41R144	0.01	0.23	0.98	−0.20	0.97	−0.40	0.30
41R145	−0.61	0.25	0.92	−0.80	0.91	−0.70	0.39
41R146	−0.27	0.24	1.15	2.00	1.17	1.90	0.01
41R147	−0.87	0.26	1.01	0.10	1.02	0.20	0.22
41R148	0.89	0.25	1.00	0.00	0.99	0.00	0.24
41R149	1.92	0.31	0.96	−0.10	0.95	−0.10	0.25
41R150	0.69	0.24	0.96	−0.60	0.94	−0.60	0.32
41R151	2.31	0.36	1.12	0.50	1.79	2.10	−0.15
41R152	2.78	0.43	1.00	0.10	2.12	2.10	−0.04
41R153	−1.32	0.31	0.98	0.00	0.92	−0.30	0.28
41R154	−4.27	1.01	1.04	0.40	3.23	1.60	−0.20
41R155	−0.67	0.27	0.96	−0.30	0.93	−0.50	0.33
41R156	−1.06	0.29	1.01	0.10	1.03	0.20	0.23
41R157	−1.53	0.33	1.04	0.20	1.16	0.70	0.14
41R158	−1.23	0.30	1.01	0.10	1.08	0.40	0.21
41R159	−0.17	0.26	1.06	0.90	1.08	0.90	0.16
41R160	−1.75	0.35	0.91	−0.30	0.91	−0.20	0.33
41R161	−1.64	0.34	1.06	0.30	0.96	−0.10	0.17
41R162	0.05	0.25	0.93	−1.10	0.91	−1.20	0.39

41R163	−1.23	0.30	0.88	−0.70	0.74	−1.30	0.47
41R164	−0.40	0.26	0.87	−1.50	0.84	−1.60	0.48
41R165	−1.14	0.29	0.85	−0.90	0.80	−1.00	0.48
41R166	0.49	0.25	1.00	0.00	0.98	−0.20	0.28
41R167	0.35	0.25	1.05	0.70	1.04	0.50	0.20
41R168	−0.79	0.28	0.93	−0.50	0.95	−0.30	0.35
41R169	−0.26	0.25	1.02	0.30	1.02	0.30	0.24
41R170	0.09	0.25	1.01	0.20	1.04	0.50	0.24
41R171	0.53	0.26	0.95	−0.70	0.92	−0.90	0.36
41R172	1.41	0.30	1.04	0.30	1.30	1.50	0.11
41R173	−0.94	0.28	0.90	−0.70	0.82	−1.10	0.43
41R174	−0.23	0.26	0.97	−0.30	0.97	−0.30	0.31
41R175	−0.16	0.25	0.85	−2.10	0.84	−1.90	0.50
41R176	−0.13	0.26	0.94	−0.80	0.92	−0.90	0.38
41R177	0.06	0.25	0.94	−1.00	0.97	−0.40	0.36
41R178	−0.27	0.26	1.05	0.60	1.03	0.30	0.21
41R179	0.19	0.25	0.94	−0.90	0.93	−0.90	0.37
41R180	0.39	0.26	1.15	2.20	1.20	2.20	0.02
41R181	−1.16	0.30	1.06	0.40	1.06	0.40	0.16
41R182	−1.45	0.33	0.96	−0.10	0.85	−0.50	0.33
41R183	−1.47	0.33	0.88	−0.50	0.85	−0.50	0.41
41R184	−0.44	0.26	0.86	−1.50	0.82	−1.60	0.50
41R185	0.29	0.25	0.97	−0.50	0.95	−0.60	0.33
41R186	0.27	0.26	1.00	0.00	1.03	0.40	0.27
41R187	0.08	0.26	0.91	−1.40	0.95	−0.60	0.40
41R188	0.44	0.26	0.99	−0.20	0.97	−0.30	0.30
41R189	0.54	0.27	1.01	0.10	1.00	0.10	0.26
41R190	1.97	0.35	1.07	0.40	1.36	1.20	0.01
41R191	1.08	0.29	0.96	−0.30	0.93	−0.40	0.32
41R192	−0.60	0.28	1.08	0.70	1.08	0.60	0.16
41R193	−0.04	0.26	1.03	0.40	1.00	0.00	0.24
41R194	0.05	0.27	0.98	−0.30	1.00	0.10	0.30
41R195	0.54	0.27	1.11	1.50	1.12	1.20	0.10
41R196	2.06	0.39	0.98	0.00	1.06	0.30	0.19
41R197	1.73	0.36	1.15	0.80	1.34	1.20	−0.07
41R198	0.96	0.31	1.01	0.10	1.16	1.10	0.21
41R199	1.05	0.33	0.89	−0.90	0.88	−0.70	0.42
41R200	0.55	0.31	1.04	0.50	1.02	0.20	0.22

A Rasch PCA of item residuals indicated that the variance explained by the measures was 24.7%, and the first five residual contrasts accounted for 3.3, 3.3, 3.0, 2.8, and 2.6% (eigenvalues of 4.4, 4.3, 4.0, 3.7, and 3.5) of the variance, respectively. The eigenvalue of the five contrasts exceed 3.0. Therefore, the test is not considered unidimensional. This is likely the result of including different types of items, vocabulary, grammar, and reading comprehension, on the same test.

Figure 1 is an item-ability map, which shows the distribution of student ability estimates relative to item difficulty estimates. The mean of item difficulty is set at zero, and the greater the value the higher the ability of the students and difficulty level of an item. This item-ability map visually shows that there was no floor or ceiling effect, which means there were enough items to distinguish higher and lower-level students. It also shows that the most difficult item was Item #124, a vocabulary item from Part 5, and the easiest one was Item #154 from Part 7. The most difficult grammar item was Item #105, a parts-of-speech item, and the easiest one was Item #102, placing "both" before "and."

(1) Parts 5 and 6

Each item from Parts 5 and 6 was examined to identify what grammar point it was testing. There were nine grammar points tested by two or more items: parts of speech (13 items), prepositions (5 items), verb form (3 items), adverbs (3 items), tense (2 items), pronouns (2 items), relative pronouns (2 items), and conjunctions (2 items). These 33 items and also 15 items that were testing the knowledge of vocabulary were subjected to the next analysis.

If you simply examine the mean of the difficulty estimates of each category, the vocabulary items were the most difficult (.71), followed by conjunctions (.50), parts of speech (.40), adverbs (.15), comparisons (.01), tense (.00), verb form (-.11), relative pronouns (-.32), pronouns (-.47), and prepositions (-.66). However, note that each category had a different number of items. Some categories, such as tense, for example, had only two items, while parts of speech had 13 items. If there had been more tense items, the mean of the difficulty estimates could have been considerably different. Thus, this is only a tentative order of difficulty, and prepositions may not necessarily be the easiest grammar point, for example.

a) Noun Items

Next I compared the difficulty within each grammatical category. Of the 13 parts-of-speech items, five are noun items, five are adjective items, and three are adverb items. Table 2 shows the difficulty estimates and the S.E.s of noun items. According to Linacre (2011), "when we want to say 'Item A is definitely more difficult than Item B'... their measures need to be more than 3 SEs different. Using this 3 S.E. criterion, their difficulty levels can be expressed as item 105 > items 132 and 138 > item
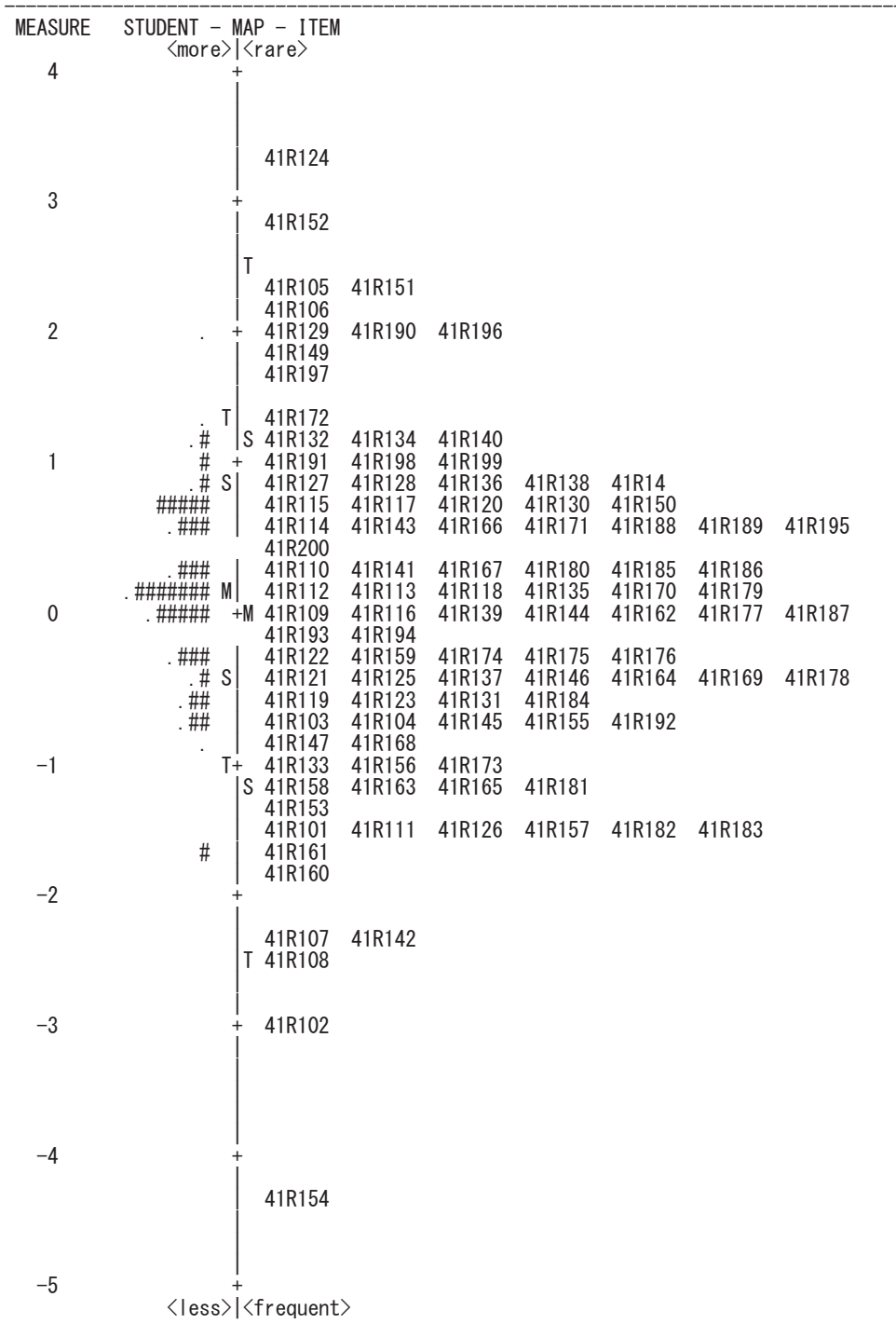
```
MEASURE   STUDENT - MAP - ITEM
            <more>|<rare>
   4          +
              |
              |
              |    41R124
              |
   3          +
              |    41R152
              |
              |T
              |    41R105   41R151
              |    41R106
   2        . +    41R129   41R190   41R196
              |    41R149
              |    41R197
              |
            . T|   41R172
           .# |S 41R132   41R134   41R140
           #  + 41R191   41R198   41R199
   1       .# S|   41R127   41R128   41R136   41R138   41R14
         ####  |   41R115   41R117   41R120   41R130   41R150
         .###  |   41R114   41R143   41R166   41R171   41R188   41R189   41R195
              |    41R200
         .### |    41R110   41R141   41R167   41R180   41R185   41R186
      .###### M|   41R112   41R113   41R118   41R135   41R170   41R179
   0  .####   +M 41R109   41R116   41R139   41R144   41R162   41R177   41R187
              |    41R193   41R194
         .### |    41R122   41R159   41R174   41R175   41R176
          .# S|   41R121   41R125   41R137   41R146   41R164   41R169   41R178
         .##  |    41R119   41R123   41R131   41R184
         .##  |    41R103   41R104   41R145   41R155   41R192
         .    |    41R147   41R168
  -1        T+    41R133   41R156   41R173
            |S 41R158   41R163   41R165   41R181
              |    41R153
              |    41R101   41R111   41R126   41R157   41R182   41R183
         #    |    41R161
              |    41R160
  -2          +
              |
              |    41R107   41R142
             |T 41R108
              |
  -3          +    41R102
              |
              |
              |
  -4          +
              |
              |    41R154
              |
              |
  -5          +
            <less>|<frequent>
```

Figure 1. Wright map for the items. Each "#" is 2 persons. Each "." is 1 person. M = Mean; S = 1 *SD*; T = 2 *SD*.

113 > item 145. Although item 105 had a blank between an article and a preposition, which seems obvious that a noun should be inserted, it was found the most difficult. This must be because the correct answer "representative" has a suffix –*ive*, which is known as a suffix of adjectives. In fact, 73% of the students chose "representation," which has a suffix –*tion*. Both items 132 and 113 had a blank after a noun, but item 132 was significantly more difficult than item 113. Forty-eight percent of the students chose "carefully" for item 132. This must be because the blank came after a verb and a noun, and the students must have thought an adverb should be placed to finish the sentence. Both items 138 and 145 had a blank in a subject position, but item 138 was found more difficult. While 35% of the students chose the correct answer, 41% of the students chose "various" and 13% chose "Varies." They may not have known which one is a noun.

Table 2. *Difficulty Estimates of Noun Items*

#	Item	Estimate	S.E.
105	A (representative) from Jensen-Corporation will be....	2.40	.36
132	... while improving patient (care).	1.14	.25
138	(Varieties) of homegrown and organic fruits....	.83	.24
113	As part of its business expansion (strategy), ...	.15	.23
145	... I am told that (verification) of the payment can....	-.61	.25

Note. Correct answers are shown in parentheses.

b) Adjective Items

Table 3 shows the difficulty estimates and the S.E.s of adjective items. Using the 3 S.E. criterion, their difficulty levels can be expressed as item 149 > items 128, 117, and 141 > item 123. In other words, item 149 was significantly more difficult than the other items, and item 123 was significantly easier than the other items. While 17% of the students chose the correct answer for item 149, 43% of the students chose "competed" and 31% chose "competitively." Those who knew the structure "keep + O + past participle" must have chosen "competed" and those who saw a verb and a noun before the blank must have chosen the adverb "competitively." As for item 123, most students seemed to have known that either an adjective or an adverb should be placed between "as" and "as," which would have increased the possibility of choosing the correct answer. In fact, 65% of the students chose the correct answer, an adjective, and 25% of the students chose an adverb.

Table 3. *Difficulty Estimates of Adjective Items*

#	Item	Estimate	S.E.
149	... keep our rates (competitive) with those of....	1.92	.31
128	... provides every amenity (necessary) for small....	.84	.24
117	... has been the nation's (leading) commercial supplier....	.73	.23
141	I am (delighted) that you have joined....	.29	.23
123	... a new material that will be twice as (durable) as ordinary concrete.	-.49	.24

Note. Correct answers are shown in parentheses.

c) Adverb Items

Table 4 shows the difficulty estimates and the S.E.s of adverb items, and their difficulty levels can be expressed as item 115 > item 121 > item 107. Item 107 must be the easiest because "speak" and "quietly" are words that students commonly see or hear together. It is interesting to note that item 115 was found significantly more difficult than item 123, an adjective item discussed above, even though they both have a blank between "as" and "as." For item 115, while 39% of the students chose the correct answer (i.e., an adverb), 36% of the students chose an adjective "accurate," and 23% of the students chose a noun "accuracy." Those who chose "accurate" might have thought that an adjective, not an adverb, should be placed between "as" and "as," and they could have chosen an adjective for item 123 for the same reason and got the correct answer luckily.

Table 4. *Difficulty Estimates of Adverb Items*

#	Item	Estimate	S.E.
115	... should answer the questions on the application form as (accurately) as possible.	.68	.23
121	... are (moderately) priced and....	- .32	.23
107	... are asked to speak (quietly) and....	-2.35	.40

Note. Correct answers are shown in parentheses.

d) Preposition Items

Table 5 shows the difficulty estimates and the S.E.s of preposition items. Item 114 was found more difficult than items 125, 133, and 108, and item 108 was found the easiest of the four. I assumed that item 114 would not be so difficult because by-agent in a passive sentence should be familiar to the students. However, they may not have noticed it because there was no be-verb before the verb "suggested" and the verb "received" immediately after "Ms. Juntasa" might have made students believe that "Ms. Juntasa received" was a subject and a verb. While 42% of the students chose the correct answer, 28% of the students chose "in" and 19% chose "of." They might have just guessed the

answer or they might have had reasons that are not clear to me. The other four items (i.e., items 135, 125, 133, and 108) were similar to vocabulary items because they require the students to understand the meaning of the sentence to choose the correct answer. It is understandable that item 108 was the easiest, since it has a simple structure and easy vocabulary, and thus the students did not have any difficulty understanding the meaning of the sentence.

Table 5. *Difficulty Estimates of Preposition Items*

#	Item	Estimate	S.E.
114	The department-wide changes suggested (by) Ms. Juntasa received unanimous approval from....	.52	.23
135	Patrons who arrive at the theater (after) the show has begun will not be seated until....	.15	.23
125	The content ... may not be reproduced (without) the company's written consent	–.38	.24
133	The afternoon's flights ... have all been delayed (due to) inclement weather in....	–1.05	.26
108	The Milltown Cinema's outdoor café will be closed (during) the winter months.	–2.52	.43

Note. Correct answers are shown in parentheses.

e) Verb Form Items

Table 6 shows the difficulty estimates and the S.E.s of verb form items. Their difficulty levels can be expressed as item 134 > item 109 > item 101. It is understandable that item 134 was the most difficult because the subjunctive is always one of the most difficult grammar points for Japanese students. While 29% of the students chose the correct answer, 33% of the students chose "to minimize," 23% chose "minimizing," and 16% chose "have minimized." Item 109 was a tricky item. Because the noun immediately before the blank is a singular noun, many students chose "begins," a third-person-singular form.

Table 6. *Difficulty Estimates of Verb Form Items*

#	Item	Estimate	S.E.
134	... Alberta Industries recommends that any associates who travel for business (minimize) the amount of luggage....	1.14	.25
109	Daily guided tours of the warehouse (begin) at 10:00 A.M. in the reception area....	.05	.23
101	Next year Khosun Industries will (send) several employees to work in....	–1.53	.30

Note. Correct answers are shown in parentheses.

f) Tense Items

Table 7 shows the difficulty estimates and the S.E.s of tense items. They both are from Part 6, and it is necessary to use the context to choose the correct answer. Using the 3 S.E. criterion, item 151 was found significantly more difficult than item 142. Although they both had keywords such as "last year" and "next few days" in the preceded sentence, item 151 was found difficult and item 142 was found easy. For item 151, while 11% of the students chose the correct answer (i.e., the past tense), 49% of the students chose "will be completed," 20 percent chose "has been completed," and 19 percent chose "will have been completed." In other words, 68 percent of the students chose the future tense, either "will be completed" or "will have been completed." This could be because the preposition "within" is often used in the future-tense sentences, even in Japanese, and the students might have automatically associated "within" with the future tense, though they should have used the context to choose the correct answer.

Table 7. *Difficulty Estimates of Tense Items*

#	Item	Estimate	S.E.
151	The work at Cardiff (was completed) within seven months.	2.31	.36
142	These (will help) you learn about our company policies and....	–2.31	.40

Note. Correct answers are shown in parentheses.

g) Pronoun Items

Table 8 shows the difficulty estimates and the S.E.s of pronoun items. Item 146 is from Part 6, which requires using the context to choose the correct answer, and item 103 is from Part 5, which is not a context-dependent item. However, item 146 was not significantly more difficult than item 103.

Table 8. *Difficulty Estimates of Pronoun Items*

#	Item	Estimate	S.E.
146	... and I will put (you) in touch with a manager at my bank....	–.27	.24
103	Dr. vargas will soon retire, ending (her) distinguished 30-year career as a....	–.66	.24

Note. Correct answers are shown in parentheses.

h) Relative Pronoun Items

Table 9 shows the difficulty estimates and the S.E.s of relative pronoun items. Item 148 is from Part 6 and item 111 is from Part 5. Although item 148 does not require using the context to choose

the correct answer, it was found significantly more difficult than item 111. For item 148, while 35% of the students chose the correct answer "whose," 55% chose "that." They could have chosen "that" simply because the antecedent is "business," which is not a person.

Table 9. *Difficulty Estimates of Relative Pronoun Items*

#	Item	Estimate	S.E.
148	Although we are a small business (whose) priority is the Asian market,....	.89	.25
111	... and make room for next season's products, (which) arrive very soon.	–1.53	.30

Note. Correct answers are shown in parentheses.

i) Conjunction Items

Table 10 shows the difficulty estimates and the S.E.s of conjunction items. They both require understanding the meaning of the sentence rather than the structure to choose the correct answer, and they were not significantly different in terms of difficulty.

Table 10. *Difficulty Estimates of Conjunction Items*

#	Item	Estimate	S.E.
127	... will be charged a service fee (whether) they pay by cash or by credit card.	.79	.24
118	... is not to be used at any time (unless) a trained lifeguard is on duty.	.20	.23

Note. Correct answers are shown in parentheses.

(2) Part 7

I grouped the reading comprehension items into seven categories based on their question types: inference questions (13 items), specific information questions with keywords (11 items), purpose questions (7 items), information questions without keywords (6 items), NOT questions (5 items), cross-reference questions (3 items), and vocabulary questions (3 items). To answer inference questions, you have to understand unwritten ideas based on the logic of the passage. For example, item 196, the most difficult reading item, asks "What does the article imply about Desparte Systems?" The word "imply" shows it is an inference question. Specific information questions with keywords are the easiest to find the answer for by skimming for the keyword of the question. For example, item 163 asks "At what time will the store close on Saturdays during the summer?" and the answer can be found easily

if you skim for "Saturday." Purpose questions asks the purpose of the passage. For example, item 153 asks "Why was the postcard send?" In order to answer this type of question, you have to identify the main idea of the passage, which often takes time. Information questions without keywords also take time to find the answers. These are the items that look like specific information questions with keywords, but actually the keyword technique does not work. You cannot find the answer even if you skim for the keyword of the question. For example, item 174 asks "Who is James Novak?" and we most likely assume that we can find the answer easily by skimming for "James Novak." However, the answer is not in the sentence "James Novak" appears. It appears a few lines above in the paragraph. In other words, you have to read and understand the whole paragraph or passage to answer this type of question correctly. NOT questions are the most time-consuming. For example, item 197 asks "What is NOT mentioned as an expected benefit of the company's changes?" and in order to answer this question correctly, you have to find three benefits mentioned in the passage and eliminate them from the choices given. Cross-reference questions are also time-consuming because you have to read two passages and connect the information to find the correct answer. An example of vocabulary questions is item 177, "The word 'finds' in paragraph 1, line 3, is closest in meaning to." Vocabulary questions could be difficult because you have to understand the meaning of the word in the context.

Table 11 shows the number of items, the mean estimates, and the mean S.E.s of each type. Looking at the mean estimates of each type, cross-reference questions were the most difficult, followed by inference questions, information questions without keywords, NOT questions, purpose questions, vocabulary questions, and specific information questions with keywords, which seems quite plausible.

Table 11. *Difficulty Order of Reading Comprehension Questions*

Type of Question	Mean Estimate	Mean S.E.
Cross-reference	.93	.30
Inference	.24	.28
Without keywords	-.04	.35
NOT	-.16	.30
Purpose	-.39	.30
Vocabulary	-.48	.28
With keywords	-.90	.27

6. Conclusion

In this study, I investigated the difficulty order of 100 TOEIC reading items (i.e., Parts 5 to 7). Before conducting the main analysis, the test was validated in terms of item and person fit statistics and dimensionality using Rasch analyses. Then the difficulty order of the items was displayed

according to their Rasch difficulty estimates. As for the items of Parts 5 and 6, I examined the difficulty order within the same grammatical category. I also grouped the items of Part 7 into seven categories based on the question type and compared the difficulty estimates.

There are limitations in this study. First, the number of participants was small and they were not considered as representatives of university students. A wider variety of participants should be involved in a future study. Second, only one test set from TOEIC Official Test-Preparation Guide Vol. 4 was used, which is insufficient because there are many more types of items that appear onthe TOEIC test. More items should be tested in the future. Third, some items were unexplainable as to why the participants chose one choice over another. Interviews should be conducted for such items in future research.

References

Educational Testing Service. (2009). TOEIC Official Test-Preparation Guide Vol. 4.

Linacre, J. M. (2011). Winsteps Help for Rasch Analysis. Retrived from http://www.winstep.com/winman

Linacre, J. M. (2013). WINSTEPS: Rasch-model computer programs (Version 3.75.0). Chicago, IL: winsteps.com.

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

TOEIC リーディング問題の難易度の検証
—ラッシュ分析を用いて

西 谷 敦 子

要 旨

　本研究は，TOEIC クラスを開講する大学等においてより効果的なカリキュラムを導入するための一歩として，TOEIC のリーディングセクション 100 問(TOEIC テスト新公式問題集 Vol. 4 の Test 1)の難易度をラッシュ分析を用いて検証するものである。参加者は京都産業大学の外国語学部開講の TOEIC クラスに登録した 83 名の大学生。ラッシュ分析により，100 問の難易度を検証した後，パート 5 と 6 の文法問題に関しては文法項目別に分け，同項目内での難易度を検証し，またパート 7 の読解問題に関しては，質問のタイプ別に分け，タイプ間の難易度差を比較した。参加者も使用問題も非常に限られた数であったため，今後の研究においてはともに数を増加し再検証することが望まれる。

キーワード：TOEIC，ラッシュ分析，項目難易度，リーディング，大学生