A Rasch Analysis of TOEIC Item Difficulty

Atsuko NISHITANI*

Abstract

This study investigated the difficulty order of 100 items in the reading section of the TOEIC test using a Rasch analysis. The participants were 83 university students, whose TOEIC scores ranged from 235 to 920. The difficulty order of the 100 items was displayed according to their Rasch difficulty estimates : The most difficult item was a vocabulary item from Part 5, and the easiest one was a reading comprehension item from Part 7. The most difficult grammar item was a parts-of-speech item, and placing "both" before "and" was the easiest. This is only a partial report on a preliminary analysis of a work-in-progress project. The more detailed report will be provided in the near future.

Keywords : TOEIC, a Rasch analysis, item difficulty, reading, university students

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. This study investigated the difficulty order of 100 items in the TOEIC reading section. The dichotomous Rasch model was used to determine the difficulty of each item and to construct an empirical hierarchy of difficulty. This study should provide 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.

Participants

A total of 83 students, who were taking the TOEIC classes in the Special English Program of the Faculty of Foreign Languages, participated in this study. Of the 83 students, 29 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. Their TOEIC scores ($n = 79$) ranged from 235 to 920, which suggests that the participants' English proficiency was mixed, ranging from false beginners to advanced proficiency learners. At the end of the test, the participants were asked for permission to use the data

---

* Faculty of Foreign Languages, Kyoto Sangyo University

for research purposes. All of the participants gave their permission to do so.

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.

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 ± 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 ± twice the standard deviation of the mean square statistic" (p. 181).

Results

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. 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.

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.

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

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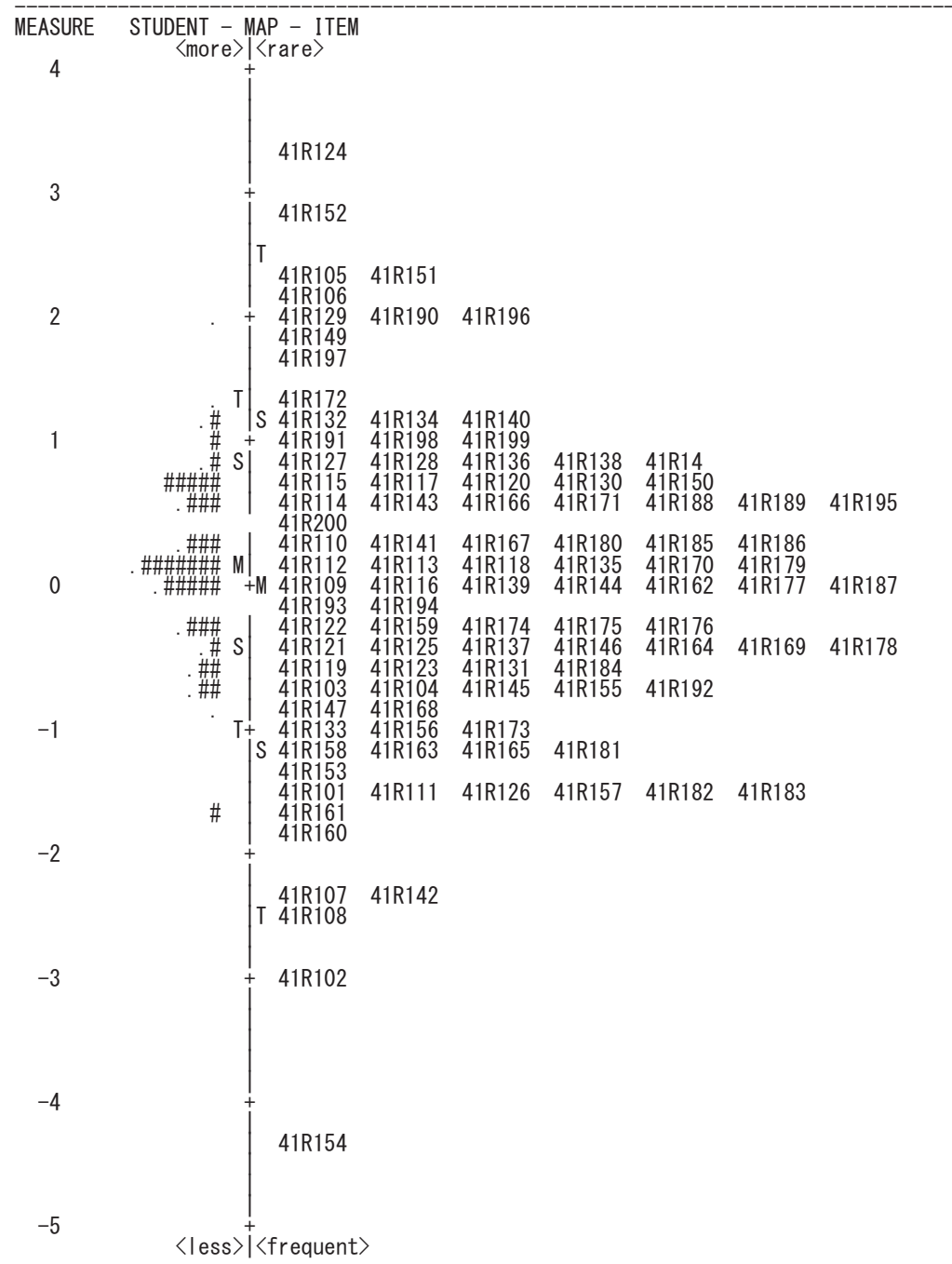
Figure 1. Wright map for the items. Each "#" is 2 persons. Each "." is 1 person. M = Mean ; S = 1 *SD* ; T = 2 *SD*.

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".

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. Because this is only a partial report on a preliminary analysis of a work-in-progress project, the analyses and results are far from complete. The more detailed report will be provided in the near future.

References

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 テスト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 のリーディングセクション 100 問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 TOEIC スコア 235 から 920 に渡る 83 名の大学生。ラッシュ分析の結果，最も難易度が高かった問題はパート 5 の語彙問題の 1 問であり，最も難易度が低かった問題はパート 7 の読解問題の 1 問であった。また，文法問題の中で難易度が最も高かったのは品詞問題で，最も低かったのは"and"の前に"both"を入れる問題であった。本稿は進行中のプロジェクトの予備分析結果を一部報告したものであり，完全なものではない。今後，さらなる分析を行い発表する予定である。

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