

Title	An analysis of a placement test for establishing an item bank
Sub Title	項目銀行構築に向けたブレイスメントテスト分析
Author	中村, 優治(Nakamura, Yuji)
Publisher	慶應義塾大学日吉紀要刊行委員会
Publication year	2008
Jtitle	慶應義塾大学日吉紀要. 言語・文化・コミュニケーション (Language, culture and communication). No.40 (2008.) ,p.89- 106
JaLC DOI	
Abstract	
Notes	
Genre	Departmental Bulletin Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN10032394-20081220-0089

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

An Analysis of a Placement Test for Establishing an Item Bank

Yuji Nakamura

1. Introduction

An item bank, according to Beeston (2000), is a large collection of test items that have been classified and stored in a database so that they can be retrieved at a later time and chosen for new tests. The items are all classified according to certain characteristics such as the topic of a text, the testing point of an item or statistical information about item difficulty. It is important for the difficulty level of each item to be determined on a common scale of difficulty so that any combination of items can be put into a new test and the item difficulties added together to give a precise measure of the difficulty of that test.

Gronlund (1998) also says that item banks are files of various suitable test items and, further, that they are coded by subject area, instructional level, instructional objective, and various pertinent item characteristics (e.g., item difficulty and discriminating power). Item banks are commonly used 1) for the construction of equivalent or alternate forms of standardized tests (different combinations of homogeneous items are drawn from the bank), and 2) as the basis for computer adaptive tests (items at a suitable level of difficulty for individual candidates are retrieved from the computer bank as required).

Choppin (1979) describes an item bank as a large collection of test questions organized and catalogued like the books in a library. The idea is that the test user can select test items as required to make up a particular test. Since one would think in terms of item banks with several thousand items, the number of possible tests which could be composed from such a bank is huge. Choppin claims that the great advantage of this system is its flexibility. Tests can be long or short, hard or difficult, as the teacher desires.

According to Davies et al (1999), the requirements for an item bank are 1) an adequate pool of test items, 2) an inventory of the abilities and content that each item purports to measure, 3) statistical data indicating the characteristics of each item as evidenced in test trailing (e.g. item difficulty and item discrimination indices), and 4) a theory or construct of ability that enables the meaning of scores on any test that may be constructed from the banked items to be interpreted. Davies et al further suggest that latent trait models are particularly useful in item banking because they have the advantage of allowing item scores to be translated into estimates of ability on a common scale. Thus, all tests deriving from a logit scale item bank are automatically equated since a person's score on any combination of test items can be converted into an ability estimate on the common bank scale. This means that any group of people can be given a test made up of items particularly suitable for them, yet all the results can be compared to one another.

Among the applications of the Rasch model, one-parameter model developed by George Rasch, item banking is useful for language testing. Item banking is the process of creating a pool of items with known and invariant measurement characteristics. The Rasch model provides estimates of item difficulties that are meaningful, irrespective of ability level tested. This paper focuses mainly on how the model can contribute to the feasibility of item banking in terms of language testing.

2. Purpose of the research

The purpose of the present study is to examine a Placement Test for the purpose of establishing an Item Bank. Thus the purpose is twofold: to examine the validity, reliability and practicality of the test and to take the necessary steps to start item banking.

Research Question 1: Is the test valid?

The validity issue will be examined in terms of the following five aspects plus the content and face validity ideas.

Basically, the validity can be examined whether the results fit the model or not. The construct validity in the Rasch model is investigated through the examination of five elements: 1) Chisquare examination, 2) Fitsresidual examination, 3) Location examination, and 4) Item Characteristic Curves, and 5) Targetting information. Among these, the item analysis using the Item Characteristic Curves is the main

focus of this present research because this can make a great contribution to a better improvement of the revised test. Along with the ICC, the information of distractors will be discussed as well.

Also, the content validity and the face validity can be discussed by using the interview survey and the questionnaire analysis (cf. Appendix)

Research Question 2: Is the test reliable?

The reliability is investigated by the person separation index, which is equivalent to the cronbach alpha. The benchmark for the acceptable boundary is over 0.7.

Research Question 3: Is the test practical/feasible?

The practicality of the test can be examined mainly by the timing factor for administration and scoring.

3. Method

3.1. Subjects

853 freshman university students in the Faculty of Letters of Keio University

3.2. Materials/ Instruments

A placement test for measuring students' English reading ability as well as grammar and vocabulary knowledge was administered. It has four components: grammar section (15 items), vocabulary section (10 items), cloze section (10 items), reading section (3 long passages with five questions each),. N.B. The reading section has three reading passages which are classified as beginning level, intermediate level, and advanced level in terms of the content, the topic, and the vocabulary level out of the teachers' teaching experience. The length of the passages are about 450-500 words. The cloze section was intended to measure their grasping ability of the context. The level of the cloze passage is for the intermediate level and the length is about 450-500 words.

3.3. Procedures

Test Construction

The Construct of Reading Ability, in other words, what is reading ability, was established mainly from the following five aspects:

- 1) the teachers' teaching experience

- 2) the reading section of other existing tests
- 3) linguistic theories
- 4) the needs of the Mita campus where students are required to read the major books and references for their study areas. In other words, the required reading ability at the Mita campus.
- 5) the text books that are actually used in their study areas.

The materials were searched and selected in the following way.

- 1) The grammar items were chosen by taking into consideration almost all of the grammar items that were supposed to have been mastered at the high school level.
- 2) The reading passages were selected from the three viewpoints (humanities, social sciences and natural sciences) by taking into consideration the appropriate vocabulary level.

Test Method, Test Format and Test Scoring

By taking into consideration the limitations of the nature of a placement test, that is, administering the test at the busiest time of the academic year, just after the entrance ceremony, scoring and informing the results should be done very quickly. Therefore, the test was a multiple-choice format rather than a response construct test, the testing time was 60 minutes and the scoring was done using the optical mark reader in an objective way.

Test Analysis

The test data was analysed using the RUMM statistical program. The benchmark for the acceptable range for the Fitresiduals is between -3 and 3. The Chisquare is investigated if there is a significant gap in the neighboring scores. The location order is examined to obtain the construct of the item difficulty order. The item characteristic curves will be examined to check the discriminating power of each item. Also, the distractor information will be discussed as well. The benchmark for the person separation index of the test reliability is over 0.7.

4. Results and Discussion

4.1. ChiSquare Pobability Order

Table 1 shows that items R348, G12, G10, R346, R350 and V19 need to be examined because there is a gap in the neighboring items.

4.2. Fitresidual Order

According to the benchmark of the acceptable range (-3 to 3), R350, R346, G10,

An Analysis of a Placement Test for Establishing an Item Bank

Table 1 ChiSquare Probability Order

Seq	Item	Type	Location	SE	Residual	DF	ChiSq	DF	Prob				
45	R245	MC	-0.697	0.090	-0.666	820.16	5.471	9	0.791512
7	G7	MC	0.589	0.074	0.936	824.08	6.664	9	0.672095
26	C26	MC	-0.268	0.082	0.484	825.06	8.244	9	0.509715
44	R244	MC	0.532	0.074	0.947	821.14	8.422	9	0.492202
11	G11	MC	1.246	0.074	1.716	824.08	8.889	9	0.447597
18	V18	MC	0.669	0.073	1.701	825.06	9.061	9	0.431705
22	V22	MC	1.241	0.074	0.420	825.06	11.418	9	0.248112
17	V17	MC	-1.176	0.103	-0.025	824.08	11.661	9	0.233098
14	G14	MC	-0.733	0.091	-1.837	824.08	11.705	9	0.230427
43	R243	MC	-2.747	0.185	-1.378	825.06	12.190	9	0.202829
32	C32	MC	-0.161	0.080	-0.765	822.12	12.305	9	0.196632
47	R347	MC	2.061	0.084	0.227	809.40	12.877	9	0.168250
25	V25	MC	0.326	0.075	0.957	825.06	13.433	9	0.143965
28	C28	MC	0.585	0.074	1.949	824.08	13.618	9	0.136572
49	R349	MC	2.057	0.085	1.581	797.65	13.847	9	0.127867
21	V21	MC	0.333	0.075	0.011	826.04	14.407	9	0.108569
41	R241	MC	-0.470	0.085	-2.424	826.04	14.733	9	0.098548
35	C35	MC	0.151	0.077	-2.329	826.04	15.344	9	0.081911
3	G3	MC	-0.218	0.081	-2.118	824.08	15.576	9	0.076292
4	G4	MC	-1.192	0.103	0.757	826.04	15.913	9	0.068711
9	G9	MC	-1.471	0.113	-1.173	825.06	16.548	9	0.056287
42	R242	MC	-0.633	0.089	-1.731	826.04	16.569	9	0.055918
34	C34	MC	-1.579	0.117	-1.871	824.08	17.319	9	0.043947
24	V24	MC	1.779	0.079	-0.162	823.10	17.687	9	0.038979
8	I0008	MC	-0.973	0.097	1.350	823.10	18.458	9	0.030215
36	R136	MC	-0.417	0.084	0.384	824.08	19.275	9	0.022957
39	R139	MC	0.528	0.074	-0.199	826.04	19.324	9	0.022579
40	R140	MC	-0.077	0.079	-1.417	825.06	19.327	9	0.022555
20	V20	MC	0.337	0.075	2.284	823.10	20.629	9	0.014405
5	G5	MC	-0.790	0.092	-1.570	826.04	20.676	9	0.014172
33	C33	MC	-0.833	0.093	-2.710	823.10	21.782	9	0.009599
38	R138	MC	-0.192	0.081	-2.653	825.06	22.381	9	0.007748
16	V16	MC	-0.556	0.087	2.352	821.14	22.516	9	0.007381
37	R137	MC	-0.860	0.094	-1.923	826.04	23.179	9	0.005808
6	G6	MC	-1.197	0.104	-1.715	826.04	25.393	9	0.002567
15	G15	MC	-1.212	0.104	-2.673	826.04	25.630	9	0.002349
29	C29	MC	-0.451	0.085	-2.718	825.06	26.261	9	0.001851
2	G2	MC	0.056	0.078	-3.796	826.04	28.699	9	0.000728
13	G13	MC	0.036	0.078	-0.982	826.04	28.936	9	0.000664
31	C31	MC	-0.577	0.088	-3.387	825.06	31.611	9	0.000232
30	C30	MC	-1.275	0.106	-3.016	824.08	31.657	9	0.000228
1	G1	MC	-1.216	0.104	-2.999	827.01	38.190	9	0.000016
27	C27	MC	-0.811	0.093	-3.679	826.04	41.269	9	0.000004
23	V23	MC	1.914	0.081	2.792	825.06	44.984	9	0.000001
10	G10	MC	0.577	0.074	5.732	826.04	52.306	9	0.000000
46	R346	MC	1.503	0.077	5.537	813.31	64.831	9	0.000000
50	R350	MC	1.887	0.083	4.980	778.08	69.607	9	0.000000
19	V19	MC	1.802	0.079	5.737	823.10	76.517	9	0.000000
48	R348	MC	1.551	0.078	5.994	800.59	101.651	9	0.000000
12	G12	MC	1.022	0.073	8.839	826.04	121.726	9	0.000000

Table 2 Fitresidual Order

Seq	Item	Type	Location	SE	Residual	DF	ChiSq	DF	Prob				
2	G2	MC	0.056	0.078	-3.796	826.04	28.699	9	0.000728
27	C27	MC	-0.811	0.093	-3.679	826.04	41.269	9	0.000004
31	C31	MC	-0.577	0.088	-3.387	825.06	31.611	9	0.000232
30	C30	MC	-1.275	0.106	-3.016	824.08	31.657	9	0.000228
1	G1	MC	-1.216	0.104	-2.999	827.01	38.190	9	0.000016
29	C29	MC	-0.451	0.085	-2.718	825.06	26.261	9	0.001851
33	C33	MC	-0.833	0.093	-2.710	823.10	21.782	9	0.009599
15	G15	MC	-1.212	0.104	-2.673	826.04	25.630	9	0.002349
38	R138	MC	-0.192	0.081	-2.653	825.06	22.381	9	0.007748
41	R241	MC	-0.470	0.085	-2.424	826.04	14.733	9	0.098548
35	C35	MC	0.151	0.077	-2.329	826.04	15.344	9	0.081911
3	G3	MC	-0.218	0.081	-2.118	824.08	15.576	9	0.076292
37	R137	MC	-0.860	0.094	-1.923	826.04	23.179	9	0.005808
34	C34	MC	-1.579	0.117	-1.871	824.08	17.319	9	0.043947
14	G14	MC	-0.733	0.091	-1.837	824.08	11.705	9	0.230427
42	R242	MC	-0.633	0.089	-1.731	826.04	16.569	9	0.055918
6	G6	MC	-1.197	0.104	-1.715	826.04	25.393	9	0.002567
5	G5	MC	-0.790	0.092	-1.570	826.04	20.676	9	0.014172
40	R140	MC	-0.077	0.079	-1.417	825.06	19.327	9	0.022555
43	R243	MC	-2.747	0.185	-1.378	825.06	12.190	9	0.202829
9	G9	MC	-1.471	0.113	-1.173	825.06	16.548	9	0.056287
13	G13	MC	0.036	0.078	-0.982	826.04	28.936	9	0.000664
32	C32	MC	-0.161	0.080	-0.765	822.12	12.305	9	0.196632
45	R245	MC	-0.697	0.090	-0.666	820.16	5.471	9	0.791512
39	R139	MC	0.528	0.074	-0.199	826.04	19.324	9	0.022579
24	V24	MC	1.779	0.079	-0.162	823.10	17.687	9	0.038979
17	V17	MC	-1.176	0.103	-0.025	824.08	11.661	9	0.233098
21	V21	MC	0.333	0.075	0.011	826.04	14.407	9	0.108569
47	R347	MC	2.061	0.084	0.227	809.40	12.877	9	0.168250
36	R136	MC	-0.417	0.084	0.384	824.08	19.275	9	0.022957
22	V22	MC	1.241	0.074	0.420	825.06	11.418	9	0.248112
26	C26	MC	-0.268	0.082	0.484	825.06	8.244	9	0.509715
4	G4	MC	-1.192	0.103	0.757	826.04	15.913	9	0.068711
7	G7	MC	0.589	0.074	0.936	824.08	6.664	9	0.672095
44	R244	MC	0.532	0.074	0.947	821.14	8.422	9	0.492202
25	V25	MC	0.326	0.075	0.957	825.06	13.433	9	0.143965
8	I0008	MC	-0.973	0.097	1.350	823.10	18.458	9	0.030215
49	R349	MC	2.057	0.085	1.581	797.65	13.847	9	0.127867
18	V18	MC	0.669	0.073	1.701	825.06	9.061	9	0.431705
11	G11	MC	1.246	0.074	1.716	824.08	8.889	9	0.447597
28	C28	MC	0.585	0.074	1.949	824.08	13.618	9	0.136572
20	V20	MC	0.337	0.075	2.284	823.10	20.629	9	0.014405
16	V16	MC	-0.556	0.087	2.352	821.14	22.516	9	0.007381
23	V23	MC	1.914	0.081	2.792	825.06	44.984	9	0.000001
50	R350	MC	1.887	0.083	4.980	778.08	69.607	9	0.000000
46	R346	MC	1.503	0.077	5.537	813.31	64.831	9	0.000000
10	G10	MC	0.577	0.074	5.732	826.04	52.306	9	0.000000
19	V19	MC	1.802	0.079	5.737	823.10	76.517	9	0.000000
48	R348	MC	1.551	0.078	5.994	800.59	101.651	9	0.000000
12	G12	MC	1.022	0.073	8.839	826.04	121.726	9	0.000000

An Analysis of a Placement Test for Establishing an Item Bank

Table 3 Location Order

Seq	Item	Type	Location	SE	Residual	DF	ChiSq	DF	Prob				
43	R243	MC	-2.747	0.185	-1.378	825.06	12.190	9	0.202829
34	C34	MC	-1.579	0.117	-1.871	824.08	17.319	9	0.043947
9	G9	MC	-1.471	0.113	-1.173	825.06	16.548	9	0.056287
30	C30	MC	-1.275	0.106	-3.016	824.08	31.657	9	0.000228
1	G1	MC	-1.216	0.104	-2.999	827.01	38.190	9	0.000016
15	G15	MC	-1.212	0.104	-2.673	826.04	25.630	9	0.002349
6	G6	MC	-1.197	0.104	-1.715	826.04	25.393	9	0.002567
4	G4	MC	-1.192	0.103	0.757	826.04	15.913	9	0.068711
17	V17	MC	-1.176	0.103	-0.025	824.08	11.661	9	0.233098
8	I0008	MC	-0.973	0.097	1.350	823.10	18.458	9	0.030215
37	R137	MC	-0.860	0.094	-1.923	826.04	23.179	9	0.005808
33	C33	MC	-0.833	0.093	-2.710	823.10	21.782	9	0.009599
27	C27	MC	-0.811	0.093	-3.679	826.04	41.269	9	0.000004
5	G5	MC	-0.790	0.092	-1.570	826.04	20.676	9	0.014172
14	G14	MC	-0.733	0.091	-1.837	824.08	11.705	9	0.230427
45	R245	MC	-0.697	0.090	-0.666	820.16	5.471	9	0.791512
42	R242	MC	-0.633	0.089	-1.731	826.04	16.569	9	0.055918
31	C31	MC	-0.577	0.088	-3.387	825.06	31.611	9	0.000232
16	V16	MC	-0.556	0.087	2.352	821.14	22.516	9	0.007381
41	R241	MC	-0.470	0.085	-2.424	826.04	14.733	9	0.098548
29	C29	MC	-0.451	0.085	-2.718	825.06	26.261	9	0.001851
36	R136	MC	-0.417	0.084	0.384	824.08	19.275	9	0.022957
26	C26	MC	-0.268	0.082	0.484	825.06	8.244	9	0.509715
3	G3	MC	-0.218	0.081	-2.118	824.08	15.576	9	0.076292
38	R138	MC	-0.192	0.081	-2.653	825.06	22.381	9	0.007748
32	C32	MC	-0.161	0.080	-0.765	822.12	12.305	9	0.196632
40	R140	MC	-0.077	0.079	-1.417	825.06	19.327	9	0.022555
13	G13	MC	0.036	0.078	-0.982	826.04	28.936	9	0.000664
2	G2	MC	0.056	0.078	-3.796	826.04	28.699	9	0.000728
35	C35	MC	0.151	0.077	-2.329	826.04	15.344	9	0.081911
25	V25	MC	0.326	0.075	0.957	825.06	13.433	9	0.143965
21	V21	MC	0.333	0.075	0.011	826.04	14.407	9	0.108569
20	V20	MC	0.337	0.075	2.284	823.10	20.629	9	0.014405
39	R139	MC	0.528	0.074	-0.199	826.04	19.324	9	0.022579
44	R244	MC	0.532	0.074	0.947	821.14	8.422	9	0.492202
10	G10	MC	0.577	0.074	5.732	826.04	52.306	9	0.000000
28	C28	MC	0.585	0.074	1.949	824.08	13.618	9	0.136572
7	G7	MC	0.589	0.074	0.936	824.08	6.664	9	0.672095
18	V18	MC	0.669	0.073	1.701	825.06	9.061	9	0.431705
12	G12	MC	1.022	0.073	8.839	826.04	121.726	9	0.000000
22	V22	MC	1.241	0.074	0.420	825.06	11.418	9	0.248112
11	G11	MC	1.246	0.074	1.716	824.08	8.889	9	0.447597
46	R346	MC	1.503	0.077	5.537	813.31	64.831	9	0.000000
48	R348	MC	1.551	0.078	5.994	800.59	101.651	9	0.000000
24	V24	MC	1.779	0.079	-0.162	823.10	17.687	9	0.038979
19	V19	MC	1.802	0.079	5.737	823.10	76.517	9	0.000000
50	R350	MC	1.887	0.083	4.980	778.08	69.607	9	0.000000
23	V23	MC	1.914	0.081	2.792	825.06	44.984	9	0.000001
49	R349	MC	2.057	0.085	1.581	797.65	13.847	9	0.127867
47	R347	MC	2.061	0.084	0.227	809.40	12.877	9	0.168250

V19, G12, R348 in Table 2 are regarded as underfitting (underdiscriminating) items, and items G2, C27, and C31 are considered as overfitting (overdiscriminating) items. However, the latter four overfitting items are at the marginal demarcation of the range. So, they can be accepted as fitting the model adequately enough.

4.3. Location Order

The items V19, R350 and V23 in Table 3 are relatively difficult, which may account for the low discrimination. Items R346 and R348 and G10 are also difficult.

4.4. Item Characteristics Curves (ICC)

Item Characteristics Curve in Figure 1 indicates that Item V19 does not work properly to discriminate the lower end and the intermediate level students. And the figures 2 to 6 show more or less the same phenomenon.

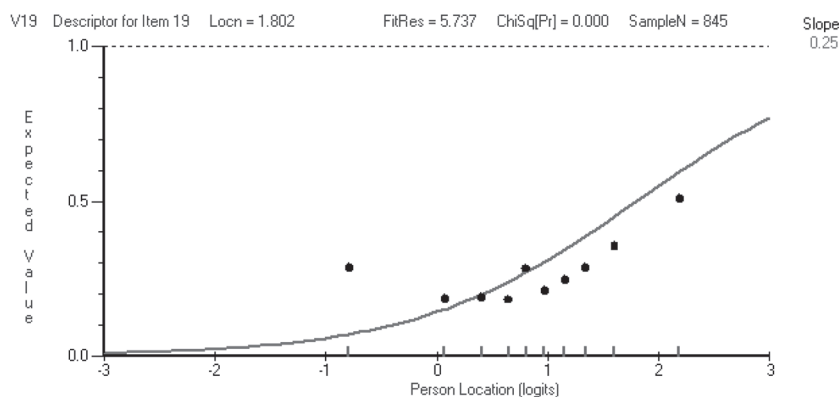


Figure 1 Item V 19 ICC

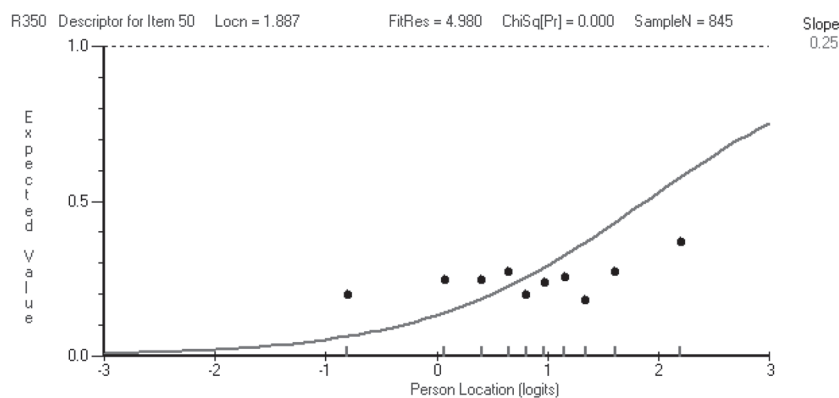


Figure 2 Item R350 ICC

An Analysis of a Placement Test for Establishing an Item Bank

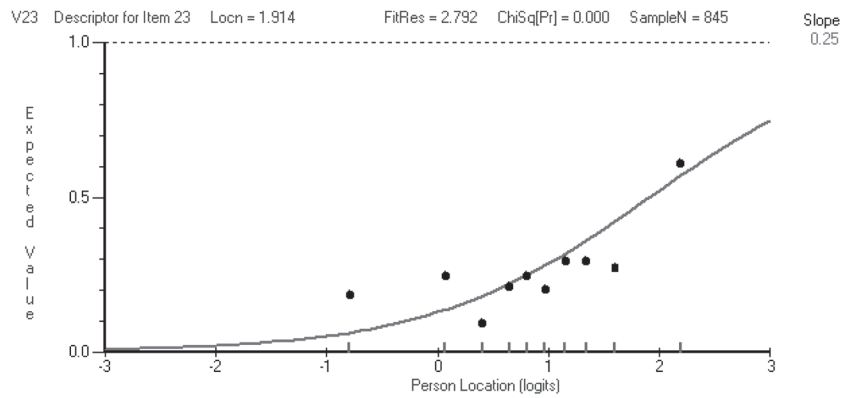


Figure 3 ItemV23 ICC

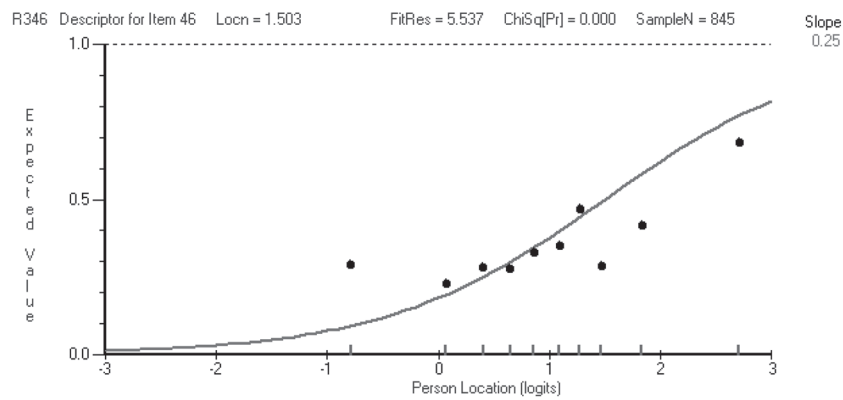


Figure 4 Item R346 ICC

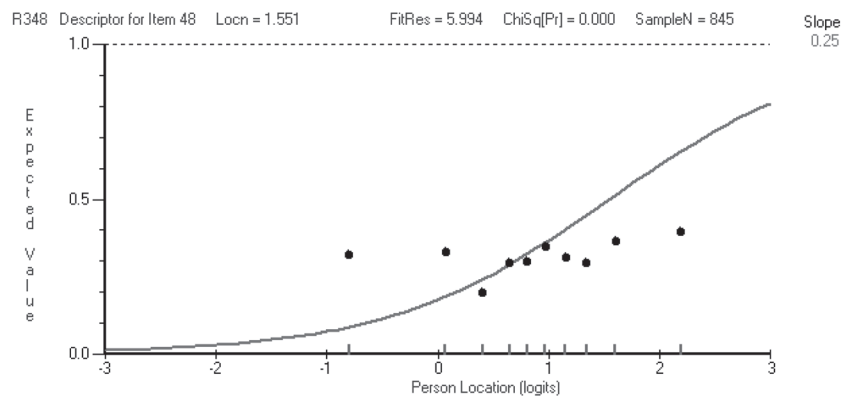


Figure 5 Item R348 ICC

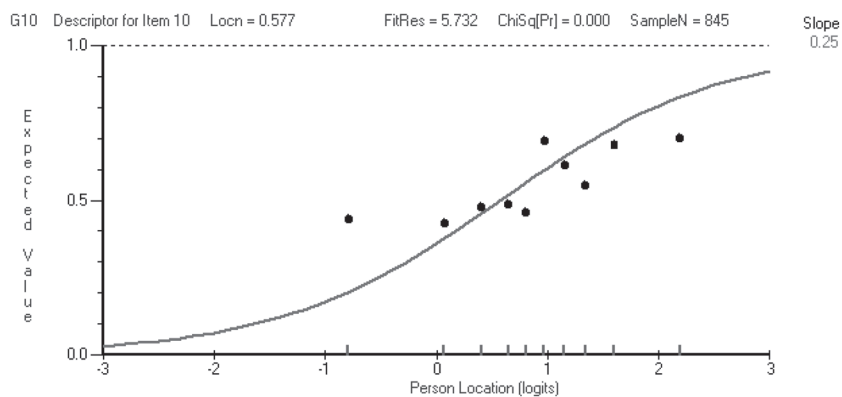


Figure 6 Item G10 ICC

4.5. Distractor Curve Information (DCI)

Figure 7 shows that Item V19 is a very difficult one and thus only the very able students tend to get it correct.

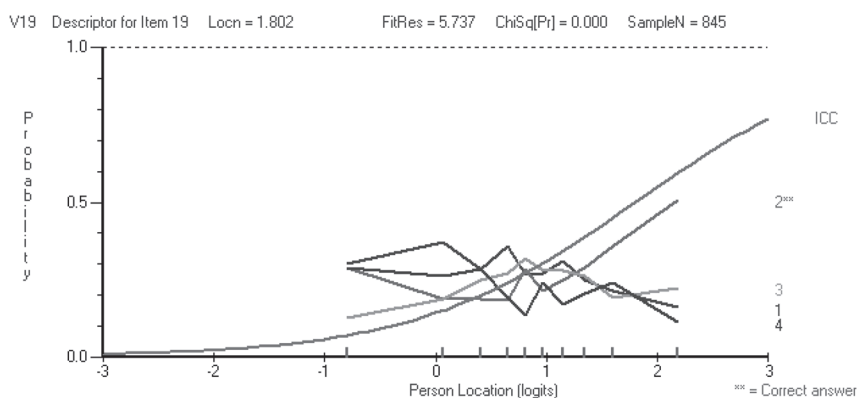


Figure 7 Item V19 DCI

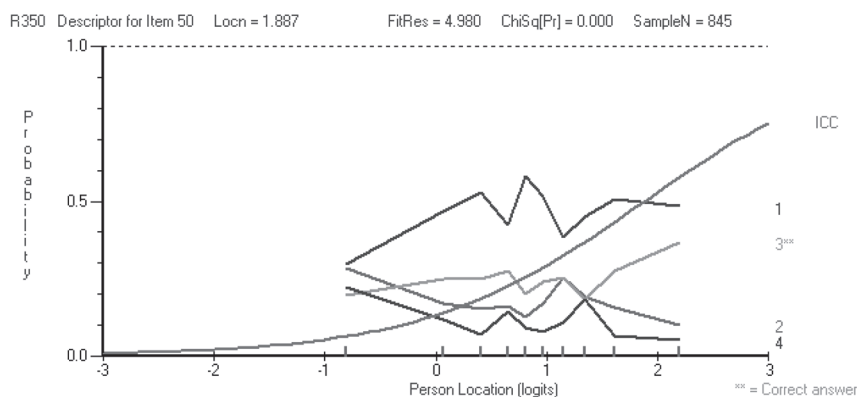


Figure 8 Item R350 DCI

Figure 8 indicates that Item R350, the correct answer, is never the one with the highest probability—even for the most able students. Another answer (1) is always more popular. It is possible that either there are two correct answers to this question, or the item difficulty level is above that of the students in this sample. Or else the correct answer may eventually be chosen but only by much more able students than in this sample.

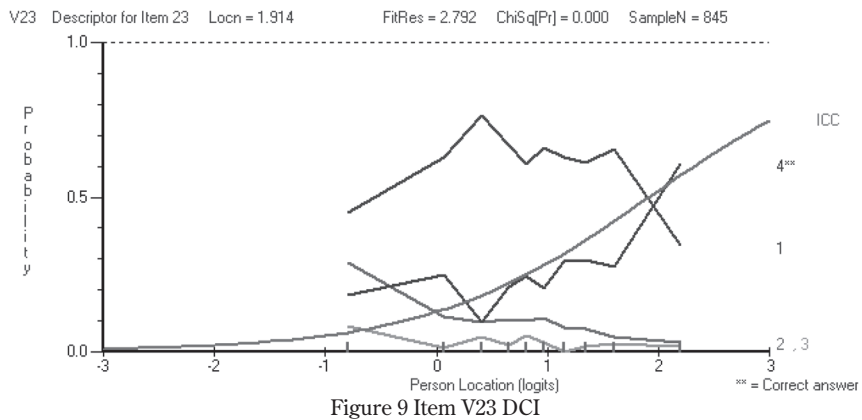


Figure 9 shows that Item V23 is so difficult that only the able student can get it correct.

4.6. Information of Targetting

Figure 10 suggest that the test is a little bit easy for the group as a whole, but it is not a bad idea to do this. Teachers do not want to frighten students unnecessarily. Overall this test is very good to measure the students' English proficiency.

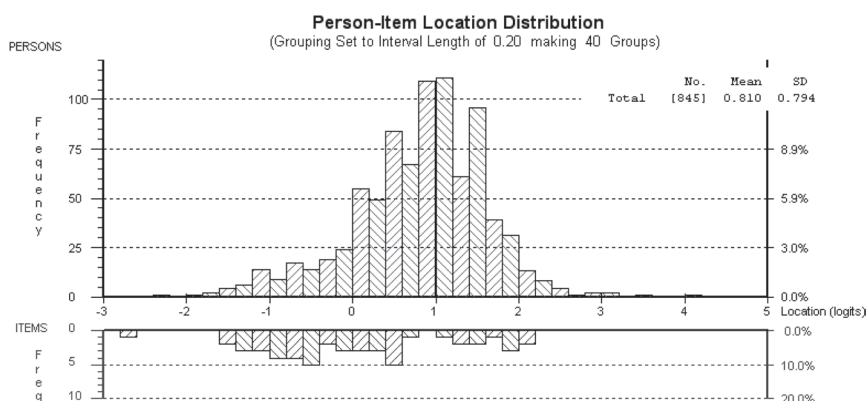


Figure 10 Person-Item Location Distribution

4.7. Examination of the reliability

The Person Separation Index (reliability)= 0.80

The reliability was verified by the acceptable score of 0.80 in the person separation index. This coefficient confirms the precision of measurement of the items, in other words, their ability to discriminate adequately among the students.

4.8. Examination of the content and face validity

The content validity was verified through the discussion of the content of the test items. All the English teachers involved in this test development agreed to this test content. Furthermore, the construct validity was also investigated in the discussion of the test format and the content. The eventual test format is composed of the four subsections of the English proficiency focusing on the reading ability.

The face validity was examined through the informal questionnaire and talk with the students by asking whether they had a feeling that they were taking a reading ability test. Most of the students agreed with the content of the test as a reading test.

In addition, one of the important aspects of a placement test quality is whether students and teachers are satisfied with the test results, i.e. whether the test results lead the students into their appropriate level so that teachers can teach more effectively and that students can learn in a more comfortable situation. This was investigated through a formal questionnaire.

An Analysis of a Placement Test for Establishing an Item Bank

Table 4 Questions to the teachers

Teacher																			
Question 1																			
The students have been placed into their appropriate levels according to their reading ability (including grammar and vocabulary knowledge) following the placement test.																			
A4 : Strongly Agree				A3 : Agree															
A2 : Disagree				A1 : Strongly Disagree															
Question 2																			
I can teach more effectively now that the students are classified on the basis of their placement test results.																			
A4 : Strongly Agree				A3 : Agree															
A2 : Disagree				A1 : Strongly Disagree															
Level 2 N=23																			
Question 1						Question 2													
A4	A3	A2	A1	A4	A3	A2	A1												
TOTAL	0	23	0	0	2	16	4	1											
						13% 70% 17% 0%													
Level 3 N=11																			
Question 1						Question 2													
A4	A3	A2	A1	A4	A3	A2	A1												
TOTAL	1	10	0	0	3	7	1	0											
						27% 64% 9% 0%													

In Table 4 the results of the questionnaire to the teachers show that most of the teachers are satisfied with the test results. This means that the test was able to separate the students into appropriate levels according to their reading ability.

Table 5 Questions to the students

Student												
Question 1												
The placement test has measured my reading ability (including grammar and vocabulary knowledge) appropriately to place me into an appropriate level class.												
A4 : Strongly Agree				A3 : Agree								
A2 : Disagree				A1 : Strongly Disagree								
Question 2												
I have been placed into an appropriate level class according to the placement test results.												
A4 : Strongly Agree				A3 : Agree								
A2 : Disagree				A1 : Strongly Disagree								
Level 1 N=25												
Question 1						Question 2						
A4	A3	A2	A1	A4	A3	A2	A1					
TOTAL	8	13	3	1	11	10	4					
32%						52%	12%	4%	0%			
Level 2 N=119												
Question 1						Question 2						
A4	A3	A2	A1	A4	A3	A2	A1					
TOTAL	7	84	24	5	10	91	20					
6%						71%	20%	3%	0%			
Level 3 N=137												
Question 1						Question 2						
A4	A3	A2	A1	A4	A3	A2	A1					
TOTAL	7	96	32	2	8	90	60					
5%						70%	23%	1%	0%			

In Table 5 the results of the questionnaire to the students also indicate that most of the students in each level are content with the test results. This suggests that they have been grouped into an appropriate level according to their reading ability.

4.9. Examination of practicality

Practicality was supported by the test method and the whole process of the test administration. It took an hour to conduct the test and the results were analysed within the same day. The test was scored objectively.

4.10. Summary of the results and discussion

Research Question 1: Is the test valid?

The validity issue was examined in terms of the following five aspects plus the content and face validity ideas and supported to a certain extent. In other words, the validity was verified by checking the tests of fit to the model.

- 1) Chisquare examination, 2) Fitresidual examination, 3) Location examination, 4) Item Characteristic Curves, and 5) Targetting information.

Although there were some misfitting items, they can be ignored due to their insignificant percentage in relation to the whole test. Furthermore, variation in test item difficulty, the main reason for the misfitting items, is necessary for the lower and upper level students.

The face validity was investigated through the questionnaire analysis, and it shows affirmative support as shown in Appendix B.

Research Question 2: Is the test reliable?

The reliability was investigated by the person separation index, which is equivalent to the Cronbach alpha. The index 0.802 cleared the benchmark 0.7. Thus, it can be said this test was reliable.

Research Question 3: Is the test practical/feasible?

The practicality of the test was examined mainly by the timing factor for administration and scoring. Also, there seems to have been no problems in administering and scoring the test itself.

5. Procedure for Item Banking

5.1. Necessary Steps for Item Banking in Theory

Item banks are collections of test questions that are stored in special computer programs where storage is structured, or organized, according to the codes assigned by users (Rudner, 1998).

One of the code sets includes item characteristics (item difficulty, for example).

The determination of item characteristics can be done in one of two ways: either by Classical Test Theory or by Item Response Theory (IRT).

Once all the items are calibrated and the difficulty of each item is determined each item can be put on the continuum of the scale according to their logit scores (difficulty level). These items along with a task can be stored as items in a bank.

The data in the present research has already been calibrated by using the Rasch statistical model which is one of the item response theories. Since the items are calibrated, we can store those items in the bank. This stage is called the deposit stage, where items are entered into special computer files. This stage is followed by the bank stage where items are stored in suitably labeled computer files. The bank stage is in turn followed by the withdrawal stage, where items are selected from the bank based on specific needs to measure test takers' ability more accurately (cf. Rudner, 1998).

In the case of the present research, since all the items have already been calibrated by the IRT based Rasch model, it can be said in theory that we can store these items in the bank stage through the deposit stage and wait for an occasion where they will be selected to match the test takers' needs or to measure their ability. However, in practice, there are some necessary procedures to make the item bank more reliable, such as increasing the number of test takers (at least 100 students, and the more the better).

Once the initial bank has been established, an advantage of calibrated item banks is in the ease of test development. Teachers withdraw from the bank those items most suitable, in terms of difficulty level, to measure the students' ability. On the basis of the test results, teachers gain greater insight into the learning process of their students. Eventually, this will be reflected in the curriculum (Rudner, 1998).

5.2. Actual Steps for Item Banking in the present research

Step 1

After checking items whether they fit the Rasch model (Chisquare examination, Fitresidual examination, and Item Characteristic Curves, Person Separation Index, and Targetting), we will consider the following problematic items mentioned above (items G12, V19, R350, R346, G10, V23, R348) as candidates of replacements. One thing we need to notice is that R350, R346 and R348 are all from the same R3 passage.

Then, we also need to find more replacements candidates to have enough anchor items for the next placement test. Generally speaking, 25% of the 50 items should be linked to equate two tests, thus, 12 items in total should be retained.

Step 2

Stage 1: We need to choose items from three sections (Grammar, Vocabulary, Reading) in a balanced way. In Grammar, out of 15 items, 11 items should be replaced. In Vocabulary, out of 10, 7 items should be replaced. In Reading out of three passages, two passages should be replaced. In Cloze, this whole passage should be replaced.

Stage 2: Simply put, stage 2 is the retention of anchors. In Grammar, we can keep 4 items. In Vocabulary, we can retain 3 items. In Reading we can leave one passage in. In Cloze we need to take the whole passage out. Thus, we can keep 12 items as anchor items which is about 25% of the 50 items for the next test. They are reasonable numbers for test equation.

Step 3

We need to choose items based on item locations, in other words, item difficulties.

So, in Grammar, I would leave items (2, 14, 13, 11) for the next test. In Vocabulary, I would keep items (17, 25, 24) for the next test. And finally, in Reading, I would retain R2 passage (R241, R242, R243, R244, R245). In total, I would leave 12 items in for the next test as anchor items for equation.

Step 4

In conclusion, we need to find 11 new items for Grammar, 7 new items for Vocabulary, Two new passages (five questions each) for Reading, and one whole passage (10 cloze questions) for Cloze. In other words, we need 38 new items in total for the next test.

6. Conclusions and Implications

Research Questions were answered relatively affirmatively. What is needed for the future development is the improvement of the validity by adding the predictive validity plus the establishment of an item bank for a wider use of the test.

Bibliography

- Baker, R. (1997). *Classical Test Theory and Item Response Theory in Test Analysis*. Language Testing Update Special Report No. 2. Lancaster: University of Lancaster.
- Beeston, S (2000). *UCLES Research Notes 2*. University of Cambridge Local Examinations Syndicate.
- Bode, R.K. and Wright, B.D. (1999). *Rasch measurement in higher education. Higher Education: Handbook of Theory and Research*, vol.XIV. (pp.287-316).
- Chapelle, A. Carol & Douglas, D. (2006). *Assessing Language through Computer Technology*. Cambridge

- UK: Cambridge University Press.
- Choppin, B. (1979). Testing the questions—the Rasch model and item banking. In M. St. J. Raggett, C. Tutt, P. Raggett (Eds.). *Assessment and Testing of Reading: Problems and Practice*. London: Ward Lock Educational.
- Crocker, L. and Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Orlando, FL: Holt, Rinehart and Winston.
- Davies, A, A. Brown, C. Elder, K. Hill, T. Lumley and T. McNamara. (1999). *Studies in Language Testing 7: Dictionary of language testing*. Cambridge UK: Cambridge University Press.
- Davidson, F. (2006). World Englishes and test construction. In Kachru, B., Kachru, Y. and Nelson, C. (eds) *The Handbook of World Englishes*. Malden, MA: Blackwell, 709-717.
- Fulcher, G. (2000). Computers in language testing. In Brett, P. and Motteram, G. (eds) *A Special Interest in Computers: Learning and Teaching with Information and Communications Technologies*. Manchester: IATEFL Publications, 93-107.
- Fulcher, G. (2007). Criteria for evaluating language quality. In Shohamy, E. (ed.) *Language Testing and Assessment. Encyclopedia of Language and Education*, vol. 7. Amsterdam: Springer.
- Fulcher, G. and Davidson, F. (2007). *Language Testing and Assessment: an advanced resource book*. London: Routledge.
- Gronlund, N. (1998). *Assessment of Student Achievement*. 6th Edition. Needham Heights, MA: Allyn and Bacon.
- Henning, G. (1987). *A Guide to Language Testing*. New York, Newbury House Publishers.
- Hozayin, R. (2000). Item Banks: Definition and Development. Paper presented at the Sixth EFL Skills Conference at the American University in Cairo, January 25th-27th, 2000.
- Linacre, John M. (1989,1994). *Many-Facet Rasch Measurement*. Chicago: MESA Press.
- Linacre, John M. and Wright, B. (1998, 2001). *A User's Guide to Bigsteps/ Winsteps: Rasch-Model Computer Program*. Chicago, IL: MESA Press.
- Nakamura, Y. (2001) Rasch Measurement and Item Banking: Theory and Practice. *Communication Studies* 15, 3-15.
- Rudner, L. (1998). Item banking. ERIC/AE Digest Series EDO-TM-98-05.
- Wright, B. and Bell, S. (1984). Item banks. What, why, how. *Journal of Educational Measurement*. 21(4), 331-345 Winter.
- Wright, B. and Stone, M. (1979). *Best Test Design: Rasch Measurement*. Chicago, IL: MESA Press.

Appendix

Questionnaire for the face validity

Questions to the students

Question 1

“The placement test has measured your reading ability including grammar and vocabulary knowledge appropriately to place you into an appropriate level class.”

Strongly Disagree	Disagree	Agree	Strongly Agree
1	2	3	4

Question 2

“You were placed into an appropriate level class according to the placement test results”

Strongly Disagree	Disagree	Agree	Strongly Agree
1	2	3	4

Questions to the teachers

Question 1

“The students are placed into their appropriate levels according to their reading ability (including grammar and vocabulary knowledge) after the placement test.”

Strongly Disagree	Disagree	Agree	Strongly Agree
1	2	3	4

Question 2

“You can teach more effectively than you used to after students are classified on the basis of the placement test results.”

Strongly Disagree	Disagree	Agree	Strongly Agree
1	2	3	4