Title	プレイスメントテストの分析 : 実験テスト結果報告
Sub Title	Analysis of a placement test : an interim report of a pilot version
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	2006
Jtitle	慶應義塾大学日吉紀要. 言語・文化・コミュニケーション No.37 (2006. 9) ,p.81- 91
JaLC DOI	
Abstract	
Notes	
Genre	Departmental Bulletin Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN10032394-20060930-0081

Analysis of a Placement Test:

An Interim Report of a Pilot Version

Yuji Nakamura*
Project Team Members of the Faculty of Letters at Hiyoshi, Keio University**

1. Introduction

In Japan there has been mounting pressure on high schools and universities to produce graduates with the ability to communicate in English. This has led to increased reliance on the communicative approach in English language classrooms. At the same time research indicates that students' reading ability along with their grammar knowledge has been deteriorating nationwide for the past years .

The social situation at universities has changed on a large scale, prompting universities to reconsider their attitudes to English teaching. Research shows that the English proficiency of freshman students has been declining. This is bound to have an impact on students' use of English materials and reference books in their studies. This in turn has two far-reaching implications. First, it disturbs the progress of their research and secondly makes it almost impossible for students to maintain a high enough level of English proficiency to continue a sound education in both liberal arts and their choice of major. Under these circumstances the internationalization of education cannot be achieved. It is high time that we consider this situation and start the English teaching revolution.

---

* Present Author

** Takuji Oda, Fumihisa Matsumoto, Andrew Armour, Yuji Nakamura, Kenji Adachi, Yoshiko Uzawa, Hikaru Sakamato, William Snell, Yoko Hemmi, Nobuya Takahashi, Kyoko Yoshida

2. Theoretical Background and Rationale

The Faculty of Letters of Keio University primarily aims to improve students' reading ability for their further college learning. For that purpose the development of a placement test is needed in order to place students into their appropriate proficiency level to facilitate the learning process, and to offer enough activities to enhance students' multi-faceted English communication ability.

The purpose of this placement test is to "measure their English reading ability to collect information of their English proficiency to make classes according to their English reading ability." The immediate goals are as follows:

1) to make classes according to their English reading ability
2) to offer classes for those who need remedial instruction
3) to offer classes for those who already are at the required level that need to continue to further advance their study of the language.

Although there are some commercialized existing tests such as TOEFL-ITP, TOEIC-IP, G-TELP, EIKEN (STEP), and CASEC, it was agreed among the faculty members that the content, the level and the purpose of those tests are not appropriate for placement in the literature department. Furthermore, the results of the admissions test cannot be used for another purpose other than the entrance examination selection. For these reasons, we have decided to develop our own placement test.

Westrick (2005) says:

> More studies on the use of commercially-produced tests and in-house tests for placement purposes at other Japanese colleges and universities are needed. Creating an effective placement test involves developing test items related to a true curriculum with clear goals and objectives, piloting the tests items, analyzing the data, and revising the tests to ensure that the scores are reliable and sound placement decisions can be made. This requires hard work, but it must be done if fair and defensible placement decisions are to be made (p.90).

The following scholars take a similar stance about the placement test. Brown (1996) says that a placement test must be more specifically related to a given program. Hughes (2003) claims that placement tests should be developed by the users themselves so that they specifically meet their needs. Fulcher (1997) argues,

"The goal of placement testing is to reduce to an absolute minimum the number of students who may face problems or even fail their academic degrees because of poor language ability or study skills."

3. Purpose of the Study

The purpose of the present study is to examine the pilot version of the placement test developed for the Faculty of Letters in order to determine what changes need to be made in order to arrive at a final version.

McNamara (2000, p.83) states, "There are three basic critical dimensions of tests —validity, reliability, and feasibility, whose demands need to be balanced." McNamara (2000, pp.50–51) also mentions three aspects that can threaten test validity: 1) test content, 2) test method and 3) test construct. Taking these three facets of a test into consideration, the research question for this study is the following: Does the Pilot version of this particular placement test have enough validity, reliability and practicality to proceed to the real test? This question gives rise to the following presuppositions :

Presupposition 1: The test has content validity as well as face validity. The test has content validity if the questions reflect the course content or syllabus. Face validity indicates if the test takers think that the test is measuring their reading ability.
(In the discussion of content validity, the test construct and the test method are additionally discussed. The test construct will be discussed in terms of the construct of the difficulty order of the subsections. The test method discussion will focus on how the test was planned, administered and scored.)

Presupposition 2: The test has the acceptable reliability (the internal consistency from Classical Language Testing theory and the information of misfitting items from the Item Response theory).

Presupposition 3: The test is practical from the viewpoint of the testing time and the analysis time as a placement test. (The test method is discussed as well.)

The purpose of the pilot version of the placement test is to examine the above presuppositions under the research question.

4. Method

a. Subjects

The participants were 809 freshman university students in the Faculty of Letters.

b. Materials/ Instruments

A placement test was used to measure students' English reading ability as well as their grammar and vocabulary knowledge. The test has four components: grammar section (15 items), vocabulary section (10 items), reading section (3 long passages with five questions each), cloze section (10 items).

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. The levels were determined by the teachers based on their prior experience teaching courses in this department. The length of the passages are about 400–500 words. The cloze section was intended to measure students' ability to grasp meaning from context.

c. Procedures

Test Construction

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

1) From the teachers' teaching experience
2) From the reading section of other existing tests
3) From linguistic theories
4) From 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) From the textbooks that are actually used in their area of study.

The materials were searched and selected in the following way:

1) The grammar items were chosen by taking into consideration almost all the grammar items that should have been mastered at the high school level.
2) The reading passages were selected from three areas (the humanities, social sciences and natural sciences). Vocabulary level was also taken into account.

The final components of the present placement test were the aforementioned grammar, vocabulary, reading and cloze sections.

Test Method, Test Format and Test Scoring

Because of limitations of time for both students and evaluators, the test format

was multiple choice rather than the constructed response test where students are asked to produce their answers in a written form. The reason for this was that the test was administered during the busiest time of the academic year, i.e., just after the entrance ceremony. The testing time was 60 minutes and scoring was done using an optical mark reader.

Test Analysis

The test data was analysed in two ways: Classical Test analysis and IRT based analysis. For this purpose, SPSS and Winsteps statistical programs were used. The benchmark for the acceptable range for the misfitting items for this analysis can be between 0.7- 1.3 in this type of dichotomous data.

Questionnaire for face validity

"Do you think this placement test measures your reading ability effectively and appropriately?"

This questionnaire was used informally to ask about students' opinions of the test in order to check the face validity.

5. Results and Discussion

5.1. Descriptive Statistics

Table 1 Descriptive Statistics

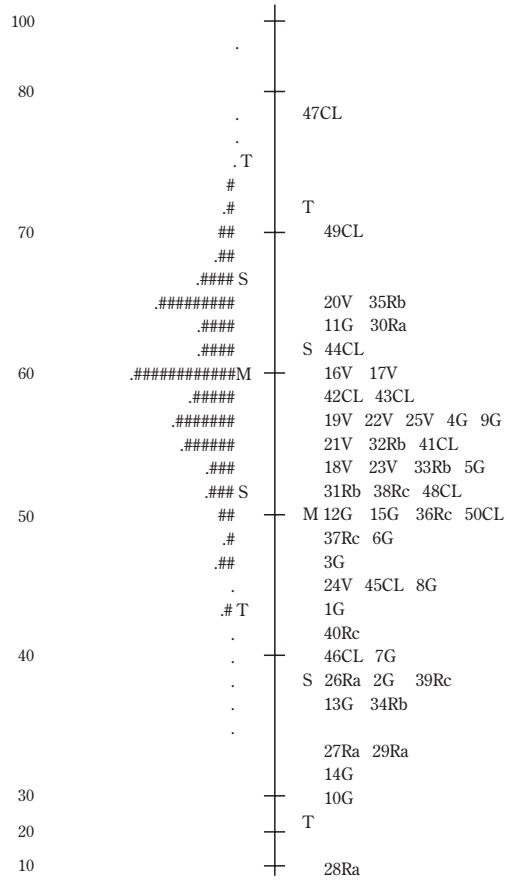
Total 50 items	Raw	Measure
MEAN	33.1	59.25
S.D.	6.6	7.82
Gram 15 items		
MEAN	11.1	64.57
S.D.	2.5	11.51
Voc 10 items		
MEAN	5.6	53.15
S.D.	1.9	10.13
Read 15 items		
MEAN	11	65.69
S.D.	2.4	12.48
Cloze 10 items		
MEAN	5.5	52.66
S.D.	1.9	11.41

Table 1 shows the mean scores and the standard deviation of the whole population for the whole test and the mean scores of the whole population for the four sub-sections (grammar, vocabulary, reading and cloze). The scores are presented in two ways (as raw scores (Raw) and as logit scores (Measure). Apparently, the Cloze section is the most difficult, followed by the Vocabulary section, then by the Grammar section. The Reading section was the easiest.

N.B. Gram=Grammar, Voc=Vocabulary, Read=Reading, Cloze=Cloze

5.2. Relative Position between Students' Ability and Item Difficulty

Figure 1 Relative Position between Students' Ability and Item Difficulty



(EACH '#' IS 10.)

N.B. G=Grammar, CL=Cloze, V=Vocabulary, Ra=Reading level 1, Rb=Reading level 2, Rc=Reading level 3

Figure 1 provides us with information about how we can divide the whole population into several ability groups by taking into consideration test item difficulty. The tables show that the top end of the students (about 60 students) seem to be already at the required level, while the bottom end of the students (about 40 students) need to have some additional or supplementary remedial instruction. The mid-level students can be divided into two groups: 1) those who are close to the top and may be able to take some advanced classes, and 2) those who are in the middle of the whole population and need to strengthen their reading ability in order to come closer to the next group.

One thing that should be pointed out is that this figure suggests we need more items with a greater degree of difficulty in order to better identify the ability of the top level students.

Another thing is that although we have observed the construct of the difficulty order of the four sub-sections, 50 items spread widely. In the table above, the cloze section was the most difficult followed by the vocabulary, then by the grammar. The reading section was the easiest. Therefore, we need to think about both the item level difficulty and the sub-section level difficulty when we increase the number of more difficult items in the future.

5.3. Information of the misfit items

Table 2 Information of the misfit items

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	REAL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEA CORR.	Item
47	126	800	77.8	1.1	1.10	1.5	1.44	4.1	A .09	47CL
11	331	806	63.3	.8	1.14	5.5	1.22	6.1	B .14	11G
24	628	808	45.3	1.0	1.14	2.9	1.20	2.9	C .12	24V
49	209	789	70.7	.9	1.03	.7	1.19	3.1	D .23	49CL
19	437	806	57.3	.8	1.09	4.0	1.15	4.9	E .20	19V
16	386	806	60.2	.8	1.10	4.3	1.14	4.5	F .20	16V
44	357	797	61.6	.8	1.05	2.3	1.10	3.2	G .25	44CL
4	446	807	56.9	.8	1.07	2.8	1.09	2.8	H .24	4G
18	498	804	53.8	.8	1.03	.9	1.07	1.9	I .29	18V
50	538	788	50.7	.8	1.06	1.8	1.07	1.5	J .25	50CL
17	388	806	60.1	.8	1.03	1.5	1.06	2.1	K .28	17V
12	556	806	50.3	.8	1.00	-.1	1.06	1.3	L .31	12G
8	628	805	45.1	.9	1.03	.6	1.06	.9	M .26	8G
31	530	806	51.9	.8	1.03	1.0	1.05	1.2	N .28	31Rb
25	443	808	57.1	.8	1.04	1.8	1.05	1.5	O .28	25V
42	406	803	59.0	.8	1.02	.9	1.04	1.5	P .30	42CL
41	483	804	54.7	.8	1.04	1.5	1.03	.8	Q .28	41CL
23	520	807	52.6	.8	1.03	1.0	1.02	.5	R .29	23V
36	567	799	49.2	.8	1.03	.7	1.02	.5	S .28	36Rc
48	533	797	51.4	.8	1.00	-.1	1.03	.6	T .32	48CL
5	509	804	53.2	.8	1.02	.7	1.00	.1	U .31	5G
22	433	802	57.5	.8	1.01	.7	1.01	.4	V .31	22V
43	427	806	57.9	.8	1.00	.2	1.01	.4	W .32	43CL
38	530	802	51.8	.8	1.01	.3	1.00	.1	X .31	38Rc
15	568	806	49.5	.8	1.00	.1	1.00	.1	Y .31	15G
2	708	808	37.5	1.1	1.00	.0	.96	-.3	y .26	2G
30	326	808	63.6	.8	.97	-1.0	.99	-.2	x .35	30Ra
33	511	808	53.2	.8	.99	-.3	.99	-.2	w .34	33Rb
27	735	808	33.8	1.3	.98	-.1	.85	-1.1	v .27	27Ra
26	705	808	37.9	1.1	.98	-.2	.95	-.5	u .28	26Ra
32	476	808	55.2	.8	.98	-.8	.97	-.8	t .35	32Rb
35	308	806	64.7	.8	.96	-1.4	.98	-.5	s .36	35Rb
9	458	806	56.2	.8	.98	-1.0	.96	-1.3	r .36	9G
20	305	807	64.8	.8	.97	-1.0	.97	-.8	q .35	20V
28	788	807	19.1	2.3	.97	-.1	.67	-1.2	p .20	28Ra
40	669	804	41.4	1.0	.96	-.7	.93	-.7	o .33	40Rc
45	634	803	44.5	.9	.94	-1.3	.96	-.6	n .37	45CL
21	479	807	55.0	.8	.95	-1.8	.93	-2.0	m .39	21V
37	579	805	48.7	.8	.95	-1.3	.90	-1.9	l .38	37Rc
1	658	808	42.7	.9	.93	-1.2	.90	-1.2	k .36	1G
6	578	807	48.9	.8	.93	-1.8	.89	-2.1	j .40	6G
39	700	803	38.0	1.1	.93	-.9	.79	-2.0	i .36	39Rc
10	757	807	29.4	1.5	.93	-.6	.71	-1.8	h .31	10G
13	710	806	37.0	1.1	.87	-1.7	.92	-.7	g .40	13G
14	748	807	31.3	1.4	.91	-.8	.73	-1.9	f .34	14G
46	682	801	39.8	1.0	.91	-1.3	.79	-2.3	e .39	46CL
34	711	807	37.1	1.1	.90	-1.2	.77	-2.1	d .38	34Rb
7	683	807	40.2	1.0	.89	-1.8	.79	-2.3	c .41	7G
29	735	808	33.8	1.3	.88	-1.3	.60	-3.4	b .41	29Ra
3	597	806	47.5	.8	.87	-3.2	.81	-3.6	a .47	3G
MEAN	534.3	804.6	50.0	.9	.99	.2	.98	.2		
S.D.	149.8	4.3	11.2	.3	.06	1.7	.15	2.1		

According to the benchmark of the acceptable range (0.7–1.3), items 47, 28 and 29 in Table 2 are misfitting, but items 28 and 29 are just overfitting because they are very easy. However, they are retained in the test because they do not cause any real harm to the data and some easy items are necessary for psychological reasons. Item 47 is underfitting and could be deleted, but this is one out of 50 items, so this does not affect the whole test. Besides, this item is in the cloze test; therefore, it cannot be replaced easily because cloze items are all interrelated with each other. For these reasons, we will keep them in the test.

Although there was no warning about item 11 which was close to the margin of the acceptable range, future research should reexamine this item.

In summary, all of the 50 items can be retained in the test as a whole. Probably for the future test, as Figure 1 suggests, we need more difficult items to measure the more able students' ability more precisely.

5.4. Reliability (Information of item internal consistency)

K-R 21 Reliability=0.801

This reliability coefficient 0.8 confirms that the test items are consistently measuring the students' reading ability with each other. Furthermore, the information in the misfitting section above also partially supports the notion that all the items in the test function in a consistent way to measure the students' reading ability.

5.5. Examination of 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 the test development agreed with the test content. Furthermore, the construct validity was also investigated along with the discussion of the test format and the content. It was found that the construct of the difficult order of the four sub-sections based on the measure scores was that the cloze test was the most difficult followed by the vocabulary test, and the grammar test. The reading test was the easiest. The eventual test format will be composed of the four subsections of English proficiency focusing on the reading ability.

The face validity was examined through the informal questionnaire and discussions 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.

5.6. Examination of the reliability

The reliability was verified by the results of the internal consistency coefficient

(Classical Test theory: 0.8) and also by the information pertaining to the very few misfitting items of the test (Item Response theory).

5.7. Examination of the practicality

The 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 by an optical mark reader.

5.8. Summary of the results and discussion

The information given through the results and discussion has confirmed that the presupposition has been partially supported. In other words, the basic three components of the test examination factors (validity, reliability, and practicality) were explained relatively convincingly.

6. Conclusions and Implications

The research question for this study "Does the Pilot version of a placement test have enough validity, reliability and practicality to proceed to the real test?" was partially supported with the examination of the three presuppositions. Also, the information obtained from the person-item relative position will help us divide the students into appropriate groups.

Considering McNamara's (2000) statement "...The right balance will depend on the test context and test purpose."(p.83), the present placement test should be acceptable judging from the statistical analysis and the test context as well as the test purpose.

For future improvement, the predictive validity should be investigated as well.

7. References and Bibliography

Alderson, J. C. (2000). *Assessing reading*. New York: Cambridge University Press.

Bachman, L. F. (1999). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Brown, J. D. (2005). *Testing in Language Programs: A Comprehensive Guide to English Language Assessment*. New Edition. New York: McGraw-Hill.

Fulcher, G. (1997). An English language placement test: issues in reliability and validity. *Language Testing* 14, 2, 113–138.

Grabe, W.(2000). Reading research and its implications for reading assessment. In A. Kunnan (Ed.), *Fairness and validation in language assessment* (pp.226–62). Cambridge: Cambridge University Press.

Hughes, A. (2003). *Testing for Language Teachers*. Cambridge: Cambridge University Press.

Linacre, M. (2004). *WINSTEPS Rasch Measurement computer program* (Version 3.51). Chicago: Winsteps. com.

McNamara, T. (2000). *Language Testing*. Oxford: Oxford University Press.

Westrick, P. (2005). Score Reliability and Placement Testing. *JALT Journal* 27, 1, 71–92.