

Title	Five year (consecutive) analysis of Japanese students' English proficiency using a large-scale in-house placement test
Sub Title	プレイスメントテストによる大学生の英語読解力測定：5年間の経年比較
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	2011
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.43 (2011.),p.37- 49
JaLC DOI	
Abstract	
Notes	
Genre	Departmental Bulletin Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN10032394-20111231-0037

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.

Five Year (Consecutive) Analysis of Japanese Students' English Proficiency Using a Large-Scale In-House Placement Test

Yuji Nakamura

I. Rationale

Beginning in the spring semester of 2006 the Keio University Faculty of Letters has administered an in-house placement exam for incoming freshmen and new sophomore students. Students were given placement tests (hereafter PT) twice a year, once at the beginning of the academic year and again in the fall (confirmation test hereafter CT). The placement test aimed to measure students' reading ability and overall English proficiency and to provide administrators with a quantifiable measure by which to give appropriate student instruction. Further, the test sought to optimize the student learning experience by better determining which areas of the multi-faceted English communicative skills needed to be strengthened.

Reading ability is thought to consist of grammar knowledge, vocabulary knowledge, long passage reading comprehension with full context (e.g. the text material does not have any deleted words or blanks intended for other questions), and passage understanding without sufficient context or information. In other words, a test of reading ability should be composed of grammar, vocabulary, reading comprehension and gap-filling items (cloze).

II. Purpose of the Study

The purpose of the present study is to report the findings from the five-year (consecutive) investigation of students' English proficiency (i.e. grammar, vocabulary, cloze and reading comprehension).

III. Research Design and Test Method

a. Materials (Test Design and Content)

The test material, which was used for PT and CT, was based on the MELAB format. The test contained 15 grammar MC questions, 10 vocabulary MC questions, 10 gap-filling MC questions, and 3 long reading passages with 5 MC questions each. Applicants had 60 minutes to complete this test, which was scored by optical readers. The reading section consisted of one beginning level, one intermediate level and one advanced level passage; each passage consisting of ~400-500 words. Teachers rated test difficulty impressionistically in terms of content, topic, and vocabulary level.

The test format (for PT and CT) in Table 1 and the test content in Table 2 are as follows:

Table 1 Test Format

Category	Grammar	Vocabulary	Gap-filling	Reading Comprehension
No. of Items	15	10	10	15
Test format	Discrete point	Discrete point	One passage with 10 blanks	3 passages with five questions each
Anchor items	4 items	3 items	None	5 items (one passage with five items)

Table 2 Test Content

Grammar	15 discrete point items; multiple choice questions (MCQ)
Vocabulary	10 discrete point items;MCQ
Gap-filling	One passage with 10 blanks; Gap-filling questions (MCQ)
Reading Comprehension	3 passages with 5 comprehension questions each; MCQ

Except PT1 (2006), each test was anchored by 12 items so that the five year (consecutive) information could be obtained. However, for the gap-filling section, because of its uniqueness, which requires the total integrative ability, it was difficult to utilize anchor items. Therefore, this section was not included in the five- year (consecutive) study. Also, the 2008 CT was not used for test equating and was excluded from the longitudinal analysis. Eventually, four pairs of test takers- (one: PT 1 and CT1, another: PT2 and CT2, another: PT4 and CT4, and the last PT5 and CT5)- were compared for the five- year (consecutive) analysis.

b. Procedures (Test construction, Administration, Timing)

Test Construction

The “reading ability” construct for this test was established mainly from specifications from the faculty of letters, which are as follows:

- 1) instructor teaching experience
- 2) reading sections of other existing tests
- 3) linguistic theories (e.g. Alderson, 2000; Grabe, 2000; Hughes, 2003)
- 4) academic and referencing needs of students' areas of specialization when transferring to Mita campus (e.g. required reading ability at the Mita campus).
- 5) textbooks that were used in students' university curricula.

The materials were selected using the following criteria:

- 1) grammar items were chosen from those grammar items that were introduced at the high school level.

Ministry of Education authorized books, which were obtained from a bookstore were used to determine the appropriateness of the high school level classification. As items were not pre-tested to empirically determine difficulty, linguistic theory was relied on to create the varied ability levels. For example, vocabulary items were based on word frequency counts using the benchmark of English-Japanese dictionaries available at bookstores; the grammar items were based on developmental sequences and on written structures on textbook analysis. The textbooks that were used were authorized by the Ministry of Education, Sports and Science and were available at bookstores.

- 2) The reading passages were selected from three disciplines (i.e. humanities, social sciences and natural sciences), and appropriate vocabulary levels were taken into consideration. The text passages were analyzed using L1 Flesch Reading Ease (Readability Formula) together with the judgments of experienced instructors.

c. Subjects (Test takers)

Subjects were the entering freshmen from the years 2006 to 2011. Table 3 shows information about test takers and corresponding test forms used.

Table 3 Information about the test takers and the corresponding test form

Test taker	Test form	Test Date	N
2006 entering students	PT1	2006 April	853
	CT1	2007 February	790
2007 entering students	PT2	2007 April	856
	CT2	2008 February	830
2008 entering students	PT3	2008 April	841
	CT3*	2009 February	794
2009 entering students	PT4	2009 April	830
	CT4	2010 February	768
2010 entering students	PT5	2010 April	816
	CT5	2011 February	764
2011 entering students	RPT1	2011 April	820

*This test data was not included in the test equation design because no anchor items were provided.

According to Table 3, for example, the test taker group of 2006 took both PT1 and CT1. For the intents of this study, in order to examine the change of the students' reading ability between PT and CT, each group that took different tests were regarded as different test taker groups. Different test population data were provided accordingly.

The 2006 PT1 test taker group was operationally defined as the norm group in this study to investigate the students' change in ability across four consecutive years.

d. Analyses

Test Analysis

The test data was analyzed using the Winsteps statistical program, the Xcalibre statistical program and the Bilog MG calibration program. The fit-misfit information was investigated to determine if the test results fit the model or not in the Rasch measurement analysis. The information about item difficulty and item discrimination was obtained in order to check each item in terms of classical test theory. The benchmark for the Cronbach alpha index of test reliability was set at 0.75 or over. In addition, the content validity is discussed by using questionnaire analysis. The Bilog MG was used to confirm that each item was functioning properly to obtain item information as well as test information.

IV. Results and Discussion

1. Descriptive statistics of each test

Table 4 Descriptive Statistics of Each Test

Test Form	N	Mean/50	SD	Max	Mini	Grammar/15	Vocabulary/10	Gap-filling/10	Reading/15
PT1	853	32.42	6.94	49	6	10.82	5.24	7.58	8.78
CT1	790	31.39	6.42	48	4	9.53	6.08	5.86	9.93
PT2	856	29.89	6.43	48	9	9.98	6.03	5.81	8.06
CT2	830	27.81	6.29	45	1	9.72	4.85	4.53	8.71
PT3	841	31.28	5.97	49	8	10.47	5.56	5.62	9.62
CT3*	794	31.72	6.14	47	11	11.07	5.61	6.22	8.81
PT4	830	32.19	6.82	47	9	10.24	6.61	6.03	9.31
CT4	768	28.71	6.57	47	10	10.29	5.13	4.61	8.67
PT5	816	33.20	6.84	49	9	11.62	7.25	6.41	7.90
CT5	764	29.1/49	6.51	46	8	8.38/14	6.35	4.94	9.48
RPT1	820	31.9	6.41	48	6	10.91	5.70	5.38	9.45

N.B. In CT5 one grammar item was invalid, so the subtotal number for grammar is 14, and the total number of the whole test was 49.

2. Examination of reliability

The reliability was investigated by the Cronbach Alpha. The benchmark for the acceptable boundary is over 0.75. The reliability of placement tests had scores over 0.75 in terms of the Cronbach Alpha index. This suggests the items in each test were internally consistent. Table 5 shows the Cronbach Alpha index in each test form.

Table 5 The Cronbach Alpha index in each test form

Test form	PT1	CT1	PT2	CT2	PT3	PT4	CT4	PT5	CT5	RPT
Cronbach α	0.83	0.78	0.77	0.75	0.76	0.81	0.78	0.82	0.78	0.78

3. Comparison of each subtest (PT and CT)

Figure A shows the 5 consecutive year analysis of the estimated mean and standard deviation (sd) of theta in three subcategories (i.e. grammar, vocabulary, and reading). Since the project started in April 2006, PT1 (2006/04) is always the benchmark for the data analysis.

One interesting finding was that when grammar and vocabulary knowledge did

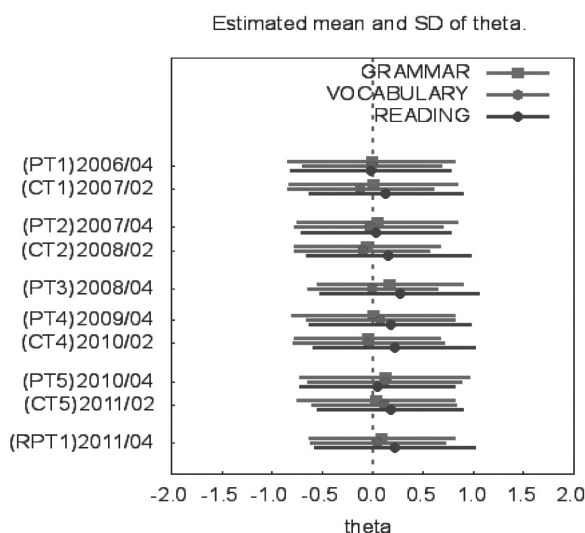


Figure A: Three sections (Grammar, Vocabulary , Reading) in comparison at a glance

not improve within the new curriculum, students' reading ability did improve. There was a continual improvement between PT and CT. Improvements between PT2 and CT2 should also be noted. The current findings may also suggest that there has been a teaching effect between pre-teaching (at the PT time) and the post-teaching (at the CT time).

Let us now examine the longitudinal change in PT test results of entering students' proficiency (grammar, vocabulary and reading) for 6 years (2006 through 2011)—(PT1, PT2, PT3, PT4, PT5, RPT 1). The grammar section indicates that with the exception of PT 4 students, all other entering students rated higher than the benchmark.

The vocabulary section is slightly different from the grammar section. PT4, PT5 and RPT1 show an improvement over the benchmark PT1; however, PT grammar remained relatively close to the benchmark, and interestingly, PT4 shows a negative improvement.

The reading part always showed a positive (i.e. improved) result when compared to the benchmark.

Secondly, we will examine paired results between PT and CT (PT1 and CT1, PT2 and CT2, PT4 and CT4, PT5 and CT5) to explore the teaching and learning effect of the same groups of students within a year.

In reference to PT1 and CT1, although there was no quantitative difference between the grammar sections, the vocabulary score decreased while the reading

score increased. In other words, reading ability showed quantifiable improvement.

In the comparison between PT2 and CT2, while vocabulary and grammar scores decreased, reading scores increased. Here again, students' reading ability improved.

Between PT4 and CT4, grammar and vocabulary scores decreased whereas reading score increased. Once again the reading score showed an improvement.

In the comparison of PT5 and CT5, while the vocabulary score shows no difference, the grammar score decreased; however, the reading score increased. Reading ability exhibited a quantifiable improvement.

In summary, the general patterns of the paired analyses show that while the grammar and vocabulary scores decreased in most cases, reading scores consistently increased. One possible explanation for this phenomenon could be attributed to the new curricula's' reading focus.

Students were expected to perform English reading activities in all English classes that were taken, so opportunities for reading English were secured.

Another potential explanation could be that typically teachers expose students to reading in English even if the eventual goal is oral communication.

Still another possibility is that the reading-focused project (i.e. curriculum) has been successful in enhancing students' reading ability.

Still another possibility is that CT tests played an important role in motivating students because the CT test results were used for the placement purposes for the following year.

As for the explanation of the deterioration of grammar and vocabulary skills, compared with their high school classes where students had vocabulary building classes or grammar-centered composition classes on a regular basis to prepare for entrance exams, there were no such classes in college. Therefore, their vocabulary storage and grammatical knowledge may have potentially decreased since the entrance exam, when their grammatical knowledge was at its peak.

4. Comparison of each subtest (consecutive four year analysis)

Note: When we compare the test results on a year basis, the benchmark basis is always the results of 2006 placement test results.

Figures G1, G2 and G3 show two things: 1. the difference between the placement test of grammar and the confirmation test of grammar in the 2006, 2007 and 2009 academic years; and 2. the change of the five year (consecutive) comparison (i.e. 2006-2011). This comparison is based on the scores from each year's placement tests.

Figures G1, G2, G3: Grammar comparison

Figure G1

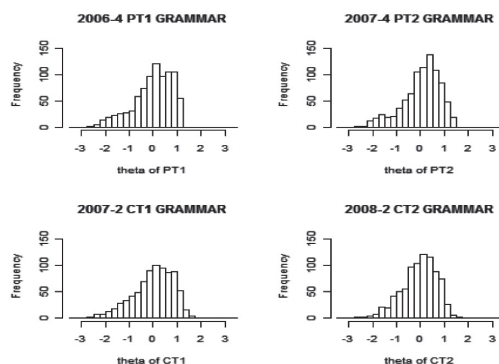


Figure G2

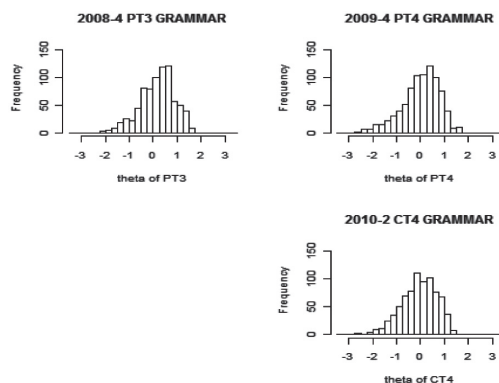
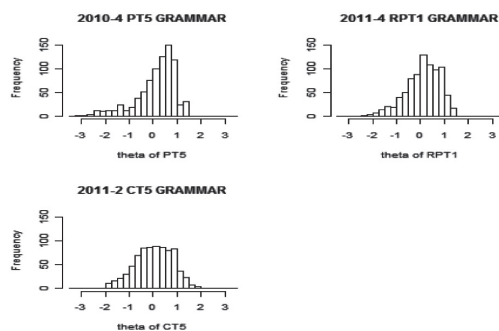


Figure G3



Figures G1,G2 and G3 show that there was little change in the students' grammar ability between the PT and the CT in each academic year. Unlike high school education, there were no English grammar courses or classes in university education

and as a result it is not surprising that students' grammatical abilities would not improve significantly. It should be noted that freshman students' grammatical abilities, peaked during university entrance examination season. University education makes a contribution to sustain basic grammatical abilities in student courses.

Although the findings did not positively suggest that placement test results for the last four years showed an increase for entering students' grammatical abilities, the pre-2010 test results indicated a greater standard deviation of the students' grammatical abilities. In other words, students' grammatical ability level showed a wider range than in previous years.

Figures V1, V2 and V3 suggest that there was no distinct increase in the results of the vocabulary section when results were compared. Moreover, students'

Figures V1, V2, V3: Vocabulary section

Figure V1

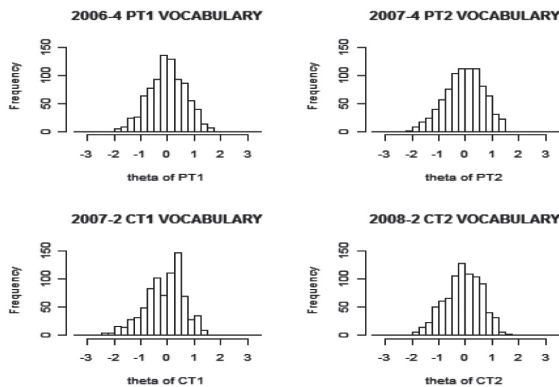


Figure V2

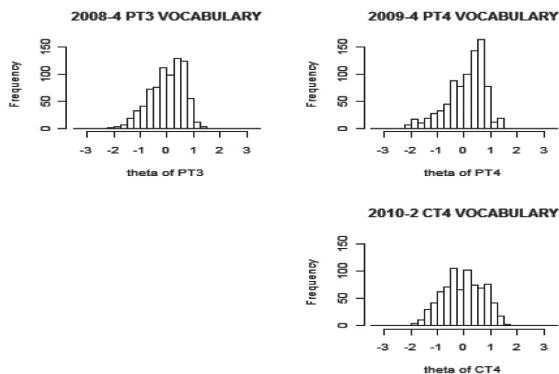
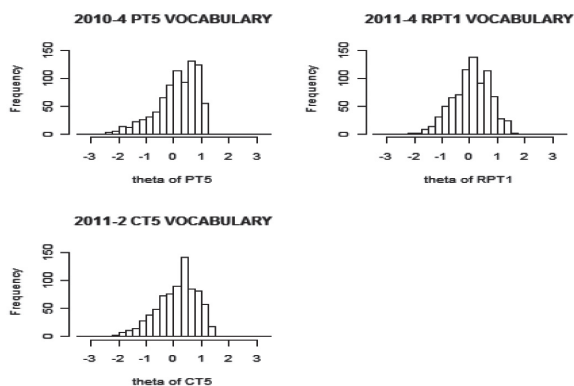


Figure V3



vocabulary abilities actually exhibited annual declines for each individual academic year. This led the researchers to believe that because there were no specific vocabulary building courses during university, students' crammed vocabulary knowledge peaked for the purposes of taking the entrance examination, and after entering university the retention of the learned vocabulary gradually waned. Student vocabulary that focused on textbook reading improved even if the width of the vocabulary did not show a significant increase.

The data suggests that entering students' vocabulary knowledge improved for the last four years.

Figures R1, R2 and R3 show that there was a visible change not only in the teaching or learning effect of reading, but also in the entering students' reading

Figures R1, R2, R3: Reading section

Figure R1

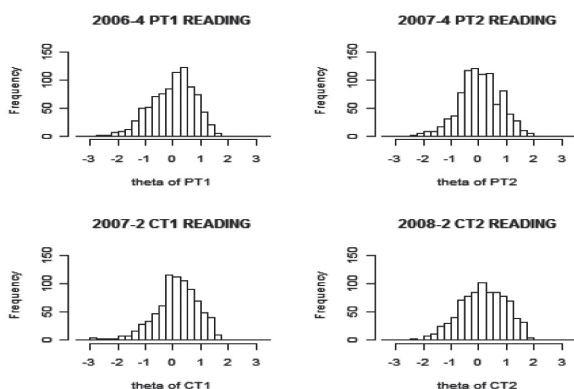


Figure R2

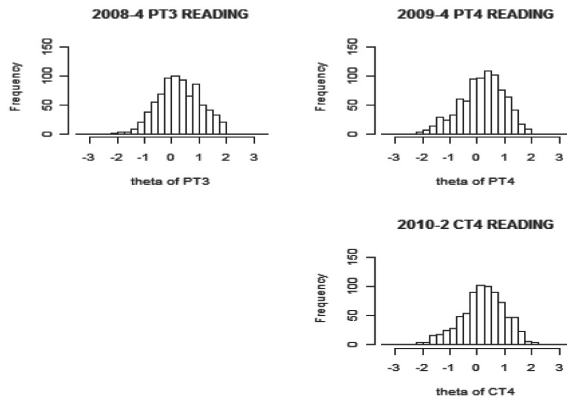
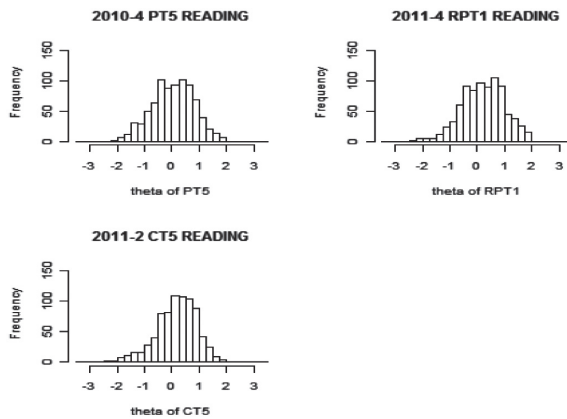


Figure R3



ability. In 2006 there was an improvement in students' reading ability between their placement test and confirmation test results. Also, in 2007 there was an increase in scores between the placement test and confirmation test. Furthermore, in 2009 there was another improvement in students' reading ability between their placement test and confirmation test. It can be said that the curriculum change, which focused primarily on enhancing students' reading ability, has been successful in this regard. One hypothesis is that freshmen students, after taking general education courses, increased their background knowledge (schemata) of subject matters in each course. As a result of students' newfound knowledge the subject matter (in other words, test topics) in the reading tests could begin to resemble those in the general education courses. In short, the increase of the background knowledge in the education courses

can affect the students' improved reading ability.

Still another possible explanation could be that unlike high school education where test taking strategies are likely utilized, university English education stresses a more thorough understanding of the text, which focuses on student analytical reading ability.

Figures R1, R2 and R3 also indicate that entering freshmen students' reading ability improved. One possible explanation for this was that placement test information, which students have access to, could have a positive impact on those who wish to enter Keio University (Faculty of Letters).

V. Conclusions and Implications

Overall the general pattern of the paired analyses show that while the grammar and vocabulary scores decreased in most cases, the reading scores always increased. One possible explanation was that there was a learning and/or teaching effect on reading because the new curriculum focused on fostering students' reading ability. The reading-focused project (curriculum) has been successful in enhancing students' reading ability.

Considering McNamara's (2000, p.83) statement "The right balance of three basic critical dimensions of tests—validity, reliability and practicality—will depend on the test context and test purpose," the present placement test should be regarded as acceptable, judging from the statistical analyses and the test context as well as the test purpose. For future improvement, predictive and concurrent validity should be measured. Further, multi-trait multi-method (MTMM) analysis could also offer valuable insights into the potential merits of test validation. Future studies should also explore the issue of face validity and practicality more systematically. For this purpose, an external benchmark such as the Common European Framework of Reference for languages teaching and learning assessment (CEFR) could be a guideline.

Acknowledgement

I would like to thank all the members of the Keio Placement Test Research Group for their cooperative work editing, as well as administering the tests. The members are as follows: Andrew Armour, Yuji Nakamura (head author of this paper), William Snell, Yoshiko Uzawa, Kenji Adachi, Hikaru Sakamoto, Yoko Hemmi, Nobuya Takahashi, Kyoko Yoshida, Satoko Tokunaga, and Yuichi Akae. I am also grateful to Mr. Haruhiko Mitsunaga for his statistical help.

References

- Alderson, J. C. (2000). *Assessing Reading*. Cambridge: Cambridge University Press.
- Alderson, J.C., Clapham, C. and Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L.F. (2004). *Statistical Analyses for Language Assessment*. Cambridge: Cambridge University Press.
- Bachman, L.F. and Palmer, A. (2010). *Language Assessment in Practice*. Oxford: Oxford University Press.
- Brown, H.D. (2010). *Language Assessment: Principles and Classroom Practices*. Second Edition. New York: Pearson Education.
- Brown, H.D. (2007). *Teaching by Principles: An Interactive Approach to Language Pedagogy*. Third Edition. New York: Pearson Education.
- Clapham, C. and Corson, D. (Eds.). (1997). Encyclopedia of language and education, vol.7, *Language testing and assessment*. Kluwer Academic Press.
- Coombe, C., Folse, K. and Hubley, N. (2007). *A Practical Guide to Assessing English Language Learners*. Ann Arbor: The University of Michigan Press.
- Douglas, D. (2010). *Understanding Language Testing*. UK: Hodder Education.
- Fulcher, G. (2010). *Practical Language Testing*. UK: Hodder Education.
- 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*. Second Edition. Cambridge: Cambridge University Press.
- Linacre, M. (2009). *WINSTEPS Rasch Measurement computer program (Version 3.64)*. Chicago, Winsteps.com.
- McNamara, T.F. (1996). *Measuring second language performance*. London: Longman.
- Stoynoff, S. and Chapelle, C.A. (2005). *ESOL Tests and Testing: A Resource for Teachers and Administrators*. Teachers of English to Speakers of Other Languages.
- Westrick, P. (2005). Score Reliability and Placement Testing. *JALT Journal*, 27, 1, 71-92.