大学生の英語文法力・語彙力の経年比較 — IRT 手法によるロジット得点を用いて —

Consecutive Four-Year Analysis of University Students' Grammar and Vocabulary Knowledge Using IRT Program

中村 優治 NAKAMURA, Yuji

● 慶應義塾大学
  Keio University

ABSTRACT

　本稿では大学生の英語文法力及び語彙力の経年的変化を測定し，さらに項目応答理論の手法によりアイテムバンキングを構築し，等化されたテストを作成することで素点による能力の比較が容易にできるような平行テストを作成する基礎づくりを目指した。経年変化は学生のテスト結果による基礎統計（素点をもとにした資料）と潜在能力（ロジット得点による資料）をもとに比較検討する。文法及び語彙のテストは全体で15問のうち7問（文法4問，語彙3問）をアンカー項目として用いて等化を行った。現時点では149項目はアイテムバンクとして構築されており，今後項目を増やすことで平行テストの数も充実させることができると思われる。文法力と語彙力は教授効果，学習効果は確認されないようである。一要因としては高校までのように文法，語彙力を意識した，さらにこれらの習得に特化したプログラム，科目は大学で設けていないことが考えられる様である。今後の継続分析が待たれるところである。

　This paper reports on the consecutive four-year investigation of students' grammar and vocabulary knowledge, using the IRT-based statistical program and suggests the implementation of an item bank which will eventually have 200 linked items by means of equation. The consecutive four-year investigation is examined

through the comparison of students' test results (descriptive statistics and latent traits). The 15 items in each test are linked by 7 anchor items (4 items in the grammar section and 3 items in the vocabulary section) for the whole item bank. 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 corresponding task can be stored as items in a bank.

1. Purpose of the Study

The purpose of the present study is to report on the consecutive four-year investigation of students' grammar and vocabulary knowledge, using the Rasch-based statistical program and to suggest the implementation of an item bank. The consecutive four-year investigation is examined through the comparison of students' test results. The 15 items in each test are linked by anchor items for the whole item bank. 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. These items can be stored as items in a bank.

2. Definition of Grammar Knowledge and Vocabulary Knowledge in the Present Study

Grammar has traditionally been considered as a syntactic system that decides how words are arranged in sentences. This view of grammar as form has been questioned by several researchers who posited that grammar involves not only formal patterns of the language in terms of morphology and syntax, but also meanings expressed through the use of forms (Bolinger, 1977).

According to Purpura's (2004) model of grammatical knowledge, knowledge of words and structures involves two dimensions: form and meaning. Thus, the two terms *grammatical knowledge* and *lexico-grammatical knowledge* are interchangeable.

Lexico-grammatical knowledge can be simply defined as knowledge of the words and grammatical structures (Celce-Murcia & Larsen-Freeman, 1999). In other words, lexico-grammatical knowledge refers to lexical and grammatical knowledge.

In the present study, we separate the lexico-grammatical knowledge into grammar knowledge and vocabulary knowledge so that the test results of each can be more effectively utilized for diagnostic purposes.

Grammar knowledge is measured in the grammar test which is based on the traditional grammatical items (e.g. eight parts of speech, or subjunctive mood etc.). The items covered in the test are supposed to have been mastered at the high school level according to the course outline by the Ministry of Education, Culture, Sports, Science and Technology (MEXT). The basic stance towards the test is to measure the students' knowledge of grammatical rules which refers to morphosyntactic and semantic knowledge which concerns linguistic structures, word formation, and meaning conveyed by those linguistic forms (cf. Rea Dickns, 1997).

Lexical knowledge is measured in the vocabulary test which should cover the basic and necessary words that students will encounter in their major study reading in English. This test aims to measure students' knowledge of words which refers to the ability to comprehend the meaning of words in an appropriate context for academic purposes.

Keio University (the faculty of letters) has administered our in-house placement test to incoming freshman students and new sophomore students since the spring of 2006 (placement tests (hereafter PT) are administered twice a year,

once at the beginning of the academic year and a confirmation test (hereafter CT) at the end of the academic year. This placement test measures students' reading ability in English and overall proficiency in order to provide streamed instruction appropriate to their proficiency levels so as to optimize their learning experience, and to provide multi-faceted English communicative skills.

The goals of this project are as follows:

1) To offer EFL at four levels of classes for students according to their English reading ability as ascertained by the method below.

2) To offer classes for those who, according to the method below, need remedial instruction.

3) To offer classes for those who have already reached the required level and desire further study.

These grammar tests and vocabulary tests are part of the whole placement tests.

3. Research Design and Test Method

a. Materials

The test contained 15 grammar MC questions, 10 vocabulary MC questions.

Table 1 Test Format

Category	Grammar	Vocabulary
No. of Items	15	10
Test format	Discrete point	Discrete point
Anchor items	4 items	3 items

Table 2 Test Content

Grammar	15 discrete point items; multiple choice questions (MCQ)
Vocabulary	10 discrete point items; MCQ

Table 3 indicates that except PT1 (2006), each test is anchored by 7 items so that the four-consecutive year information can be obtained. Also, the 2008 CT

which was not used for test equating was excluded from the longitudinal analysis. Eventually, three pairs of test takers (one: PT 1 and CT1, another: PT2 and CT2, and the other: PT4 and CT4) were compared for the four-consecutive-year analysis.

N.B. PT stands for Placement Test and CT stands for Confirmation Test.

Table 3 Test information

Test form	Test Date
PT1 CT1	2006 April 2007 February
PT2 CT2	2007 April 2008 February
PT3 CT3*	2008 April 2009 February
PT4 CT4	2009 April 2010 February
PT5	2010 April

b. Procedures (Test construction, Administration, Timing)

Test Construction

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. To obtain this high school level information, there are textbooks authorized by the Ministry of Education that are available at bookstores. Since we did not pretest items in order to determine their difficulty empirically, we relied on theory to create items and sections at different ability levels.

2) The 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 the written

structures on textbook analysis. The textbooks authorized by the Ministry of Education, Sports and Science are available at bookstores.

c. Subjects (Test takers)

The subjects were the entering students (2006 to 2010). Table 4 shows the information about the test takers and the corresponding test form.

Table 4   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

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

According to Table 4, for example, the test taker group of 2006 took both PT1 and CT1. In this study, in order to examine the change in the students' reading ability between PT and CT, each group who took a different set of tests is regarded as a different test-taker group. Different test population was provided accordingly.

The 2006 PT1 test-taker group was operationally defined as the norm group in this study to investigate the students' ability change 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 to check each item in terms of the classical test theory. The Bilog MG was used to confirm that each item was functioning properly to obtain item information as well as test information.

4. Results and Discussion

4. 1 Mean scores of each test (based on raw scores)

Table 5 and Figures 1-2 show the change of the mean scores of each test (grammar and vocabulary) for four years (PT1 through PT5). The test results are based on the raw scores.

Table 5   Mean scores of grammar and vocabulary tests

Test Form	N	Grammar/15	Vocabulary/10
PT1	853	10.82	5.24
CT1	790	9.53	6.08
PT2	856	9.98	6.03
CT2	830	9.72	4.85
PT3	841	10.47	5.56
CT3*	794	11.07	5.61
PT4	830	10.24	6.61
CT4	768	10.29	5.13
PT5	816	11.62	7.25

4. 2 Possibilities of Item Bank using the Rasch model

The 25 items in each test are linked by 7 anchor items for the whole item bank. 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. Up to now we have been able to store 149 items linked by anchor items.
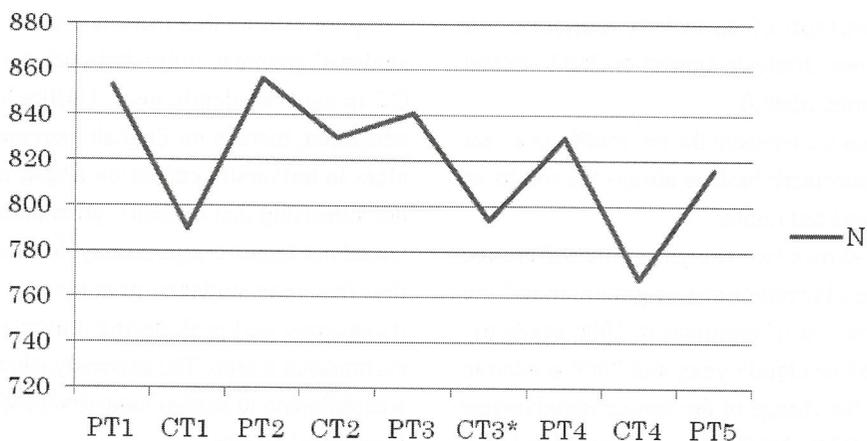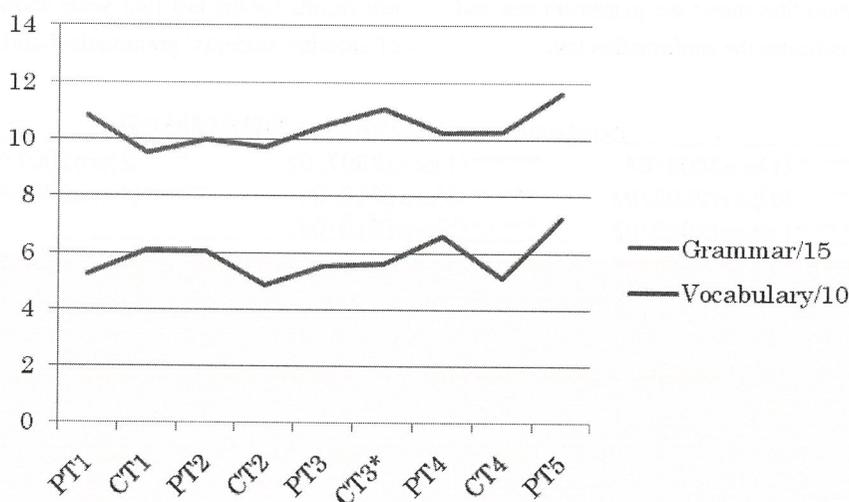
Figure 1

Figure 2

As Table 6 shows, it is possible to make four testforms. One testform consists of four sub-sections (grammar, vocabulary). In one test, grammar has fifteen items, vocabulary has ten items, Four parallel tests can be used to check students learning practice or teaching effect within an academic year or for the longitudinal purpose by using the raw scores because all the items were linked and equated by the anchor items.

Table 6   Information about misfitting items in each sub-section (grammar, vocabulary)

	grammar	vocabulary
Underfitting items	11	7
Overfitting items	11	2
Misfitting items (sub-total)	22	9
Total items of each section	90	59
Candidates (total-misfit)	68	50

Possible testforms (4)
Necessary items
Grammar: 15 items x 4 (testforms)=60 items/68
Vocabuary: 10 items x4 (testforms)= 40 items/50

4. 3 Comparison of each test: consecutive four year analysis based on logit socres (students' ability)

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

Figure 3 shows two things: 1. the difference between the placement test of grammar and the confirmation test of grammar in 2006 academic year, in 2007 academic year, and 2009 academic year; and 2. the change of the four-consecutive-year comparison (2006-2010). This comparison is based on the scores from the placement tests from each year. The dotted line shows the placement test and the solid line indicates the confirmation test.

Figure3 shows that there is little change in the students' grammar ability between the PT and the CT in each academic year. Unlike high school education, there is no English grammar course or class 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 in its separate unit peak during university entrance examination season. The university education makes a contribution to sustain basic grammatical abilities in students' courses.

Although it is difficult to prove that the placement test results for the last four years show an increase of entering students' grammatical abilities, the pre
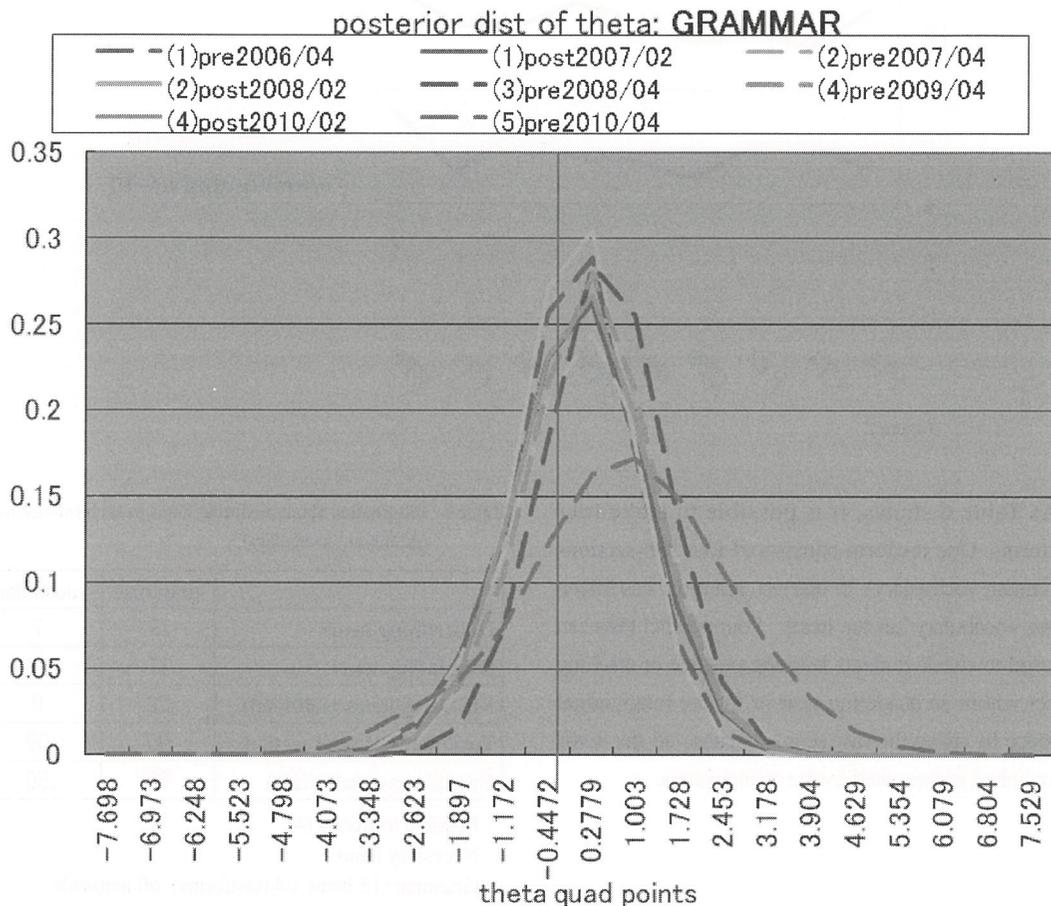

Figure 3 Grammar comparison

2010 test result indicate bigger standard deviation of the students' grammatical abilities. In other words, students' grammatical abilities differ more widely than the previous years.

Figure 4 suggests that there was no noticeable impact in vocabulary section when we compare the dotted and solid lines. Moreover, students' vocabulary knowledge actually exhibited annual declines for each individual academic year. This lead the researchers to believe that because there were no specific vocabulary building courses in university, students' crammed vocabulary knowledge peaked for the purposes of taking the entrance examination, but after entering university the retention of the learned vocabulary gradually faded.

Instead, student vocabulary that focused on textbook reading improved even if the width of the vocabulary did not show a significant increase.

The solid lines suggest that entering students' vocabulary knowledge has been improving for the last four years.

5. Conclusions and Implications

We can draw five conclusions as follows:

1. Using logit scores, we were able to conduct the four-consecutive-year analysis to examine the students' improving within a year or to investigate the entering students' grammar and vocabulary knowledge.
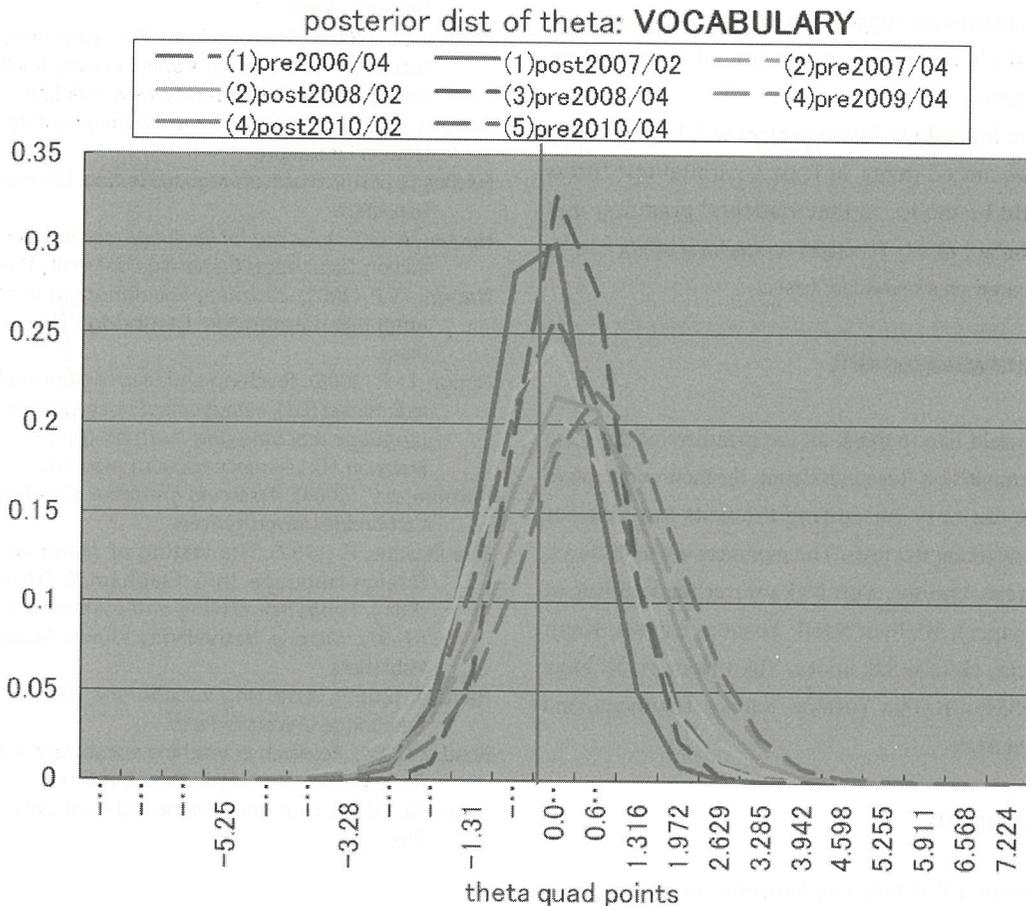

Figure 4   Vocabulary section

2. The analysis shows that there is no remarkable teaching or learning effect in students' grammar and vocabulary knowledge. This is probably because in university, unlike high school, there are no specific grammar courses or vocabulary building courses as such.

3. Through test equation linked by anchor items, from now it will be possible to compare students' ability change using raw scores.

4. Vocabulary knowledge could be more precisely measured by using an appropriate corpus to establish an ideal or practical objective to master.

5. Communicative grammar ability and vocabulary knowledge should be investigated in performance tests (e.g. speaking and writing skills) where students are supposed to express their productive skills which were not measured in the present tests.

One immediate future project will be that using 149 anchored items, at least 4 parallel test forms should be made, so that students' grammar and vocabulary ability (change) or teaching effect can be compared on a raw-score basis.

Acknowledgement

References

Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L.F. and Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University Press.

Bolinger, D. (1977). *Meaning and form*. New York: Longman.

Borg, S. (1998). Teachers' pedagogical systems and grammar teaching: A qualitative study. *TESOL Quarterly* 32,1: 9-38.

Borg, S. (1999). Teachers' theories in grammar teaching. *ELT Journal*, 53,3 : 157-167.

Brown, H.D. (2007). *Teaching by principles: an interactive approach to language pedagogy*. Third Edition. New York: Pearson Education.

Brown, H.D. (2010). *Language assessment: principles and classroom practices*. Second Edition. New York: Pearson Education.

Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The grammar book*. Boston: Heinle & Heinle.

Coombe, C., Folse, K. and Hubley, N. (2007). *A practical guide to assessing English language learners*. Ann Arbor: The University of Michigan Press.

Davies, A., Brown, A., et al. (Eds.). (1999). *Dictionary of language testing*. Cambridge: Cambridge University Press.

Doughty, C. (2002). Focus on form: the latest word for teachers . Third Annual Apple Lecture, Teachers College, Columbia University, New York City.

Douglas, D. (2010). *Understanding language testing*. UK: Hodder Education.

Fulcher, G. (2010). *Practical language testing*. UK: Hodder Education.

Hughes, A. (2003). *Testing for language teachers*. Second Edition. Cambridge: Cambridge University Press.

Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I.S.P. (2005). Teaching and learning vocabulary. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp.581-595). Mahwah, NJ: Lawrence Erlbaum Associates.

Purpura, J. E. (2004). *Assessing grammar*. Cambridge: Cambridge University Press.

Rea-Dickins, P. (1997). The testing of grammar in a foreign language. In C. Clapham, & D.Corson (Eds.), *Language testing and assessment* (pp. 87–97). Victoria, Netherlands: Kluwer Academic Publishers.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics, 24*, 146-161.

Swan, M. (2005). *Grammar*. Oxford: Oxford University Press.