

L2文法力測定の経年変化とCEFRレベルの利用可能性

A Longitudinal Study of Second Language Grammar Knowledge Assessment and the Feasibility of Utilizing the CEFR Levels

中村 優治 NAKAMURA, Yuji

● 慶應義塾大学
Keio University

Background and the Rationale of the Paper

Grammar knowledge is an important factor in measuring reading skill in EFL. Keio University (Faculty of Letters) has administered an in-house placement test to incoming freshman students and new sophomore students since 2006. Part of this test is devoted to examining students' grammar knowledge in regard to reading.

Larsen-Freeman (2009) claims that because of the preference in recent years for measuring the use of grammar holistically, through speaking and writing, some standardized examinations, e.g., the TOEFL iBT, no longer have a separate section of the test that deals explicitly with grammatical structure. As Larsen-Freeman mentions, the decision to eliminate the explicit testing of grammar was made, in at least two cases, based on research showing that a separate subtest of grammatical knowledge could not be adequately differentiated from other sections of a test. However, it might be difficult to separate out what, in the ability to read or write text, is due to the lack

of knowledge concerning grammatical structures and what might be due to other factors. She also points out that we have no way of diagnosing grammatical difficulties learners may be experiencing, or in providing them with feedback. In summary, discrete-point and integrative tests represent different approaches to grammar assessment, each of which have a contribution to make (Larsen-Freeman, 2009).

Jones (2013) also indicates that many large format proficiency tests do not directly include a section on grammatical knowledge, because that knowledge is indirectly tested through demonstration of the four main language skills. Why is it necessary to assess discrete grammatical items explicitly in a separate test when they could be tested implicitly through speaking and writing test tasks? Some tests, such as the Paper-based TOEFL or the higher-level Cambridge ESOL exams, do include sections related to the structure and usage of the language.

Whether or not we use a test to independently assess grammatical ability depends on the purpose of the assessment. There are occasions when the

knowledge gained from detailed information is beneficial for teachers and students. For instance, diagnostic and placement tests often have a grammar element, since detailed information may be required regarding a student's grammatical ability.

Furthermore, it is possible to test and score a considerably large number of items in a relatively short period of time. Similarly, in achievement and proficiency tests, the inclusion of a grammar component, such as the Use of English paper in the higher-level Cambridge ESOL examinations, might be deemed necessary simply because it can give us more information regarding students' grammatical ability.

Pupura (2013) maintains that while language teachers have always acknowledged the importance of grammar in second language (L2) teaching and testing, the notion of what it means to "know" the grammar of an L2 has evolved in many ways. Consistent with developments in L2 pedagogy, models of L2 proficiency used in assessment have shifted from a primary focus on decontextualized linguistic structures to an emphasis on measuring communicative language ability through performance (Pupura, 2013).

The Common European Framework of Reference (CEFR) has had a great impact on the learning, teaching, and assessment of foreign languages in Japan as well as Europe. The CEFR Levels are: C2 (Mastery), C1 (Effective Operational Proficiency), B2 (Vantage), B1 (Threshold), A2 (Waystage), and A1 (Breakthrough). Whenever different stages of learning and attainment are proposed, one needs some way of distinguishing them (Hawkins & Filipovic, 2012). However, these are all taken from the global scale of Common Reference Levels. Though overall guidelines are well used, each individual testing item or testing task varies from institution to institution. Therefore, it is difficult to establish correspondences between a test and the

CEFR levels.

Hawkins and Filipovic (2012), by paying special attention to syntactic (and lexical) properties of English, proposed an illustrative set of grammatical (and lexical) criterial features for the learning of English. Basically speaking, the idea of criterial features is that properties of learners' English are identified which are characteristic and indicative of L2 proficiency at each of the levels, and that distinguish higher levels from lower levels (Hawkins & Filipovic, 2012). According to Hawkins and Filipovic (2012), criterial features are defined in terms of linguistic properties of the L2 as used by native speakers, which have either been correctly or incorrectly attained by learners at a given level.

Since there are many properties that potentially serve as criteria, defining the criteria for each level will necessarily involve multiple factors, such as grammatical, lexical, phonological, and discourse features. Of course the different language functions can also be the criterial features that learners can perform (cf. Hawkins and Filipovic, 2012).

Purpose

The purpose of the present study is: 1) to conduct a longitudinal analysis of students' grammar knowledge, 2) to discuss important issues in second language grammar knowledge assessment (e.g. the significance of discrete-point tests and the necessity of the global standard scale, and 3) to discuss the feasibility of using the CEFR criterial features in the grammar sections of a university placement test.

In this paper we will focus primarily on the grammatical properties of English in order to consider the levels. The purpose of the paper is to contribute towards establishing correspondences for grammar and reading proficiency. In this study, a placement test framework and the CEFR will be compared based on the placement test results, reading and grammar.

| | |
|--|---|
| Nouns and noun phrases Pre-determiners, determiners Nouns (countability, affixation, compounding) | Pronouns & reference (cohesion) Personal, demonstrative, reciprocal Relative, indefinite, interrogative |
| Verbs, verb phrases, tense & aspect Tense-present, etc., aspect-progressive Subject-verb agreement | Questions & responses Yes/no, wh-, negative Tags |
| Modals & Phrasal Modals (be able to) Forms-present etc., Obligation | Conditionals Forms-present, past, future Factual, counterfactual conditionals |
| Phrasal verbs Form-2-word, 3-word Separability | Passive Voice Form-present, past etc. Other passive-get something done |
| Prepositions & prepositional phrases (PP) Co-occurrence-rely on, fond of Spatial or temporal-at the store at 5 | Complements & complementation V+NP+NP Want him to, believe him to |
| Adjectives and adjectival phrases Formation Adjective order-the old black Joe | Comparisons Comparatives and superlatives Equatives-as big as |
| Logical connectors Relationships of time, space, reason & purpose Subordinating & coordinating conjunctions | Adverbials and adverbial phrases Forms-adverb phrases, clause, Placement-sentence initial, medial and final |
| Relative clauses Forms: animate, inanimate, Subject NP, direct object NP, genitive NP | Reported speech Backshifting Indirect imperatives |
| Nonreferential It & There Time, distance---it's noisy in here Existence-there is / are | Focus & emphasis Emphasis do Marked word order-him I see |

Figure 1 Purpura's taxonomy of grammatical forms

Method

Subjects

Beginning in the spring semester of 2006, Keio University Faculty of Letters has administered an in-house placement exam for incoming freshmen and new sophomore students. Students are given placement tests twice a year, once at the beginning of the academic year and again in the fall. For this analysis, the results of 10 exams (approximately 800 test-takers each administration) are examined.

Material 1 (placement test)

The placement test is a 60-minute examination which consists of 50 questions in four sections: grammar, vocabulary, gap-fill (cloze), and reading.

For this study, only the vocabulary section was examined. The vocabulary section consists of 10 multiple choice questions with four options. The contents of 10 placement tests, 100 test items in total, were examined.

Material 2 (CEFR levels)

- Hawkins and Filipovic (2012). English Profile Studies: Criterial Features in L2 English (EPS)
- English Vocabulary Profile: <http://vocabularypreview.englishprofile.org/dictionary/show/uk/> (EVP)

The following grammatical (and lexical) features have been presented as criteria for levels A2-C2 based on Hawkins and Filipovic (2012).

1. Criterial features for A2

e.g. simple intransitive clauses (NP-V), transitive clauses (NP-V-NP), direct WH-questions, pronoun plus infinitive (something to eat), modals (MAY,CAN, MIGHT) in the possibility sense, modal SHOULD in the advice sense.

2. Criterial features for B1

e.g. verbs with object-controlled infinitival complements (I ordered him to gather my men to the hall), verbs with object-controlled -ing complements (I saw a girl standing behind me), relative clauses formed on a genitive position (painter whose pictures I like), indirect WH-questions in infinitival phrases (I did not know where to look for it anymore), modal MAY in the permission sense, modal MUST in the necessity sense, modal SHOULD in the probability sense.

3. Criterial features for B2

e.g. adverbial subordinate clauses with -ing (Talking about spare time, I think we could go to the Art Museum), it extraposition with infinitival phrases (it would be helpful to work in your group as well), verbs with an NP plus finite complement clauses (I told him I loved his songs)

4. Criterial features for C1

e.g. modal MIGHT in the permission sense, the bride's family's house, new subject to subject raising constructions plus passive with the verbs (assumed, discovered, felt, found, proved)

5. Criterial features for C2

e.g. new subject to subject raising constructions plus passive with the verbs (presumed) (Hawkins & Filiovic 2012).

Material 3 (cf. Purpura 2013)

An abbreviated version of Purpura's taxonomy of grammatical forms is as follows (cf. Pupura 2013).

Procedures

First, we made a list of the words that appeared in the items in the 10 placement test forms. Next,

using item response theory (IRT) we constructed an item difficulty scale using WINSTEPS and Xcalibre. We considered this scale to be a measurement of grammar skill in reading English documents. In the IRT analysis 150 test items, including common items (anchor items) were used to apply common-item nonequivalent groups design using IRT analysis. Third, these 150 items were investigated on the basis of CEFR levels and correlations are examined. Finally, correspondences between the two scales (the PT and the CEFR) were discussed.

In the following, we will first describe the CEFR (reading section) and the placement test used in the study by focusing on the underlying scale of vocabulary. Next, we will describe the participants of the test, the way the study was conducted, and the statistical analyses used, followed by the results of the study.

Results and Discussion

Results 1 (Longitudinal Results)

The results in Table 1 show that the students' grammar abilities declined during the course of each individual academic year.

Results 2 (Test results and CEFR)

Table 2-a shows the Placement Test (PT) level descriptions. Table 2-b CEFR shows the global level description. Table 1-c CEFR shows the reading level description. Table 1-d shows the CEFR grammar level description. As was mentioned above, all of these CEFR descriptions are overall general statements. Therefore, it is difficult to determine what students at each level can do in practice, or what test items they can correctly identify in reality. Each individual institution can determine this for themselves.

Table 1 Descriptive Statistics of Each Test

| Test Form | N | M/50 | SD | Max | Min | G (/15) | V (/10) | Gap (/10) | Rd (/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 | 33.2 | 6.84 | 49 | 9 | 11.62 | 7.25 | 6.41 | 7.9 |
| CT5* | 764 | 29.17 | 6.51 | 46 | 8 | 8.38 | 6.35 | 4.94 | 9.48 |

G: Grammar; V: Vocabulary; G: Gap-filling; Rd: Reading; * : Tests are not linked by anchor items.

Table 2-a: Descriptions of PT levels

| Level | Learning Objectives | % of students |
|-------------------|---|---------------|
| Advanced High | Develop multiple, advanced communication skills | 7 % |
| Advanced | Develop the ability to express oneself in English | |
| Intermediate High | Develop advanced English reading skills and an equivalent ability to express oneself in English | 30 % |
| Intermediate | Acquire by the end of the 2nd year sufficient English reading ability for major studies | 58 % |
| Basic | Review and reaffirm basic skills | 5 % |

Table 2-b: Descriptions of CEFR global levels

| Group Level | A | | B | | C | |
|------------------|--------------------------|------------------------|---------------------------|-------------------------------|---|------------------------|
| Group Level Name | Basic User | | Independent User | | Proficient User | |
| Level | A1 | A2 | B1 | B2 | C1 | C2 |
| Level name | Breakthrough or beginner | Waystage or elementary | Threshold or intermediate | Vantage or upper intermediate | Effective Operational Proficiency or advanced | Mastery or proficiency |
| Cf. | Basic | | Intermediate | Intermediate High | Advanced | Advanced High |

Table 2-b continued. *Common Reference Levels: Global scale*

C2 Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of proficient meaning even in more complex situations.

C1 Can understand a wide range of demanding, longer texts, and recognize implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.

B2 Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and independent disadvantages of various options.

B1 Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.

A2 Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate Basic need.

A1 Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

Table 2-c: Descriptions of CEFR reading levels

OVERALL READING COMPREHENSION

C2 Can understand and interpret critically virtually all forms of the written language including abstract, structurally complex, or highly colloquial literary and non-literary writings. Can understand a wide range of long and complex texts, appreciating subtle distinctions of style and implicit as well as explicit meaning.

C1 Can understand in detail lengthy, complex texts, whether or not they relate to his/her own area of speciality, provided he/she can reread difficult sections.

B2 Can read with a large degree of independence, adapting style and speed of reading to different texts and purposes, and using appropriate reference sources selectively. Has a broad active reading vocabulary,

but may experience some difficulty with low frequency idioms.

B1 Can read straightforward factual texts on subjects related to his/her field and interest with a satisfactory level of comprehension. Can understand short, simple texts on familiar matters of a concrete type which consist of high frequency everyday or job-related language.

A2 Can understand short, simple texts containing the highest frequency vocabulary, including a proportion of shared international vocabulary items.

A1 Can understand very short, simple texts a single phrase at a time, picking up familiar names, words and basic phrases and rereading as required.

Table 2-d CEFR grammar levels

GRAMMATICAL ACCURACY

C2 Maintains consistent grammatical control of complex language, even while attention is otherwise engaged (e.g. in forward planning, in monitoring others' reactions).

C1 Consistently maintains a high degree of grammatical accuracy; errors are rare and difficult to spot. Good grammatical control; occasional 'slips' or non-systematic errors and minor flaws in sentence structure may still occur, but they are rare and can often be corrected in retrospect.

B2 Shows a relatively high degree of grammatical control. Does not make mistakes which lead to misunderstanding. Communicates with reasonable accuracy in familiar contexts; generally good control though with noticeable mother tongue influence. Errors occur, but it is clear what he/she is trying to express.

B1 Uses reasonably accurately a repertoire of frequently used 'routines' and patterns associated with more predictable situations.

A2 Uses some simple structures correctly, but still systematically makes basic mistakes – for example tends to mix up tenses and forget to mark agreement; nevertheless, it is usually clear what he/she is trying to say.

A1 Shows only limited control of a few simple grammatical structures and sentence patterns in a learnt repertoire.

Table 3a CEFR levels and PT item difficulty (Unique Examples)

| Level | Item | Distinction |
|-----------|------------|---|
| C2 | 75 | quite high CEFR level but easy item |
| C1 | 74 | high CEFR level but easy item |
| B2 | 94, 92, 91 | rather high CEFR level but very easy item |
| B1 | 98, 97 | intermediate CEFR level and easy item |
| A2 | 44, 42 | rather low CEFR level but difficult item |
| A1 | 45, 24, 23 | quite low CEFR level but quite difficult item |

In order to link tests with CEFR in a valid manner, much more research evidence is needed. However, we could argue that correspondences may be

compared between the tests and the CEFR in terms of grammar items in the following way.

Table 3a and Table 3b indicate that words in each

CEFR level have various levels of difficulty. In the A1 level, there are 8 items which range in difficulty from 0.23 to 0.84 (difficult to easy). Even at the lowest levels there exist some quite difficult items. At the A2 level, one item has a difficulty of 0.42, and at the B1 level 0.16 and 0.20 items were included. However, at the C2 level, one item has a difficulty of 0.75, and at the C1 level, three items are over 0.65. This means that it is not necessarily true that higher-level items are more difficult than lower-level items. Each of the five CEFR levels contains both easy and difficult items for Japanese students. In other words, the higher-level words could be easier test items. In fact, one of the easiest items is a C2 level word (0.75). Conversely, there are some lower-level items which were very difficult, with the most difficult item being a B1 word (0.16, 0.20), or an A1 level word (0.23 or 0.24). Nevertheless, these seem to be rare cases.

More importantly, each level has a variation in the difficulty of test items (easy to difficult). Therefore, a person within each ability level (A1 to C2) should be able to answer correctly at least half of the items of his/her ability level, or more. Then, he or she can move either up or down to the next level. The CEFR grammatical criteria features levels should be adjusted to the local teaching, learning or testing context to be utilized effectively.

A correlation between PT difficulty and CEFR levels
CEFR vs PC of grammar -0.01961

The above information shows almost no correlation between the CEFR level and the item difficulty (PC).

Conclusions and Implications

1. Longitudinal conclusion

The results show that the students' grammar abilities declined during the course of each

individual academic year. Also, the discrete-point test results provided us with accurate and reliable information about the decline of grammatical ability within a year or over the years. There should be an advantage to using the discrete-point test for this purpose.

2. Significance of discrete-point tests and complementary support from other tasks or methods

The choice of appropriate test methods or test tasks is paramount due to the potential negative effects they might have on test performance. Weaknesses of testing formats such as multiple choice, however, may often result from the way in which they have been used, rather than because of some inherent defect. For instance, employing only multiple-choice questions to assess a student's "grammatical knowledge" is severely limiting and may prove detrimental to a test taker's performance, so teachers should be encouraged to use a number of different question types depending on the structures they want to test.

A helpful way to categorize test tasks is according to the types of responses they require from students: selected-response tasks, limited-production tasks, and extended-production tasks (Jones, 2013).

As Purpura (2013) mentions, a form-based approach is conspicuously narrow in scope because in language use a form can rarely be disassociated from its meaning potential.

To measure grammatical knowledge, learners are presented with tasks allowing them to demonstrate their receptive, emergent, or productive knowledge of grammar. Tasks involve a collection of characteristics that vary on several dimensions. The choice of grammar tasks should be selected according to the assessment purpose.

In designing tasks for grammar, assessment specialists find it useful to categorize them according to the type of elicited response (Figure

Table 3b CEFR Levels and Test Item Difficulty

| C2 | C1 | B2 | B1 | A2 | A1 |
|-----------|----|------|------|-------|------|
| 75 | 74 | 94 | 98 | 98 | 84 |
| | 70 | 92 | 97 | 97 | 73 |
| | 66 | 91 | 96 | 89 | 71 |
| | | 90 | 94 | 88 | 71 |
| | | 86 | 92 | 86 | 68 |
| | | 86 | 92 | 86 | 45 |
| | | 85 | 91 | 86 | 24 |
| | | 84 | 91 | 80 | 23 |
| | | 81 | 91 | 77 | |
| | | 74 | 89 | 77 | |
| | | 74 | 89 | 77 | |
| | | 72 | 89 | 72 | |
| | | 72 | 89 | 72 | |
| | | 71 | 89 | 71 | |
| | | 68 | 88 | 66 | |
| | | 67 | 85 | 62 | |
| | | 63 | 85 | 56 | |
| | | 57 | 85 | 55 | |
| | | 56 | 84 | 53 | |
| | | 56 | 83 | 51 | |
| | | 55 | 82 | 44 | |
| | | 52 | 82 | 42 | |
| | | 52 | 81 | | |
| | | 51 | 80 | | |
| | | 47 | 79 | | |
| | | 46 | 78 | | |
| | | 44 | 77 | | |
| | | 42 | 77 | | |
| | | 40 | 73 | | |
| | | 40 | 70 | | |
| | | 36 | 69 | | |
| | | 35 | 69 | | |
| | | 30 | 65 | | |
| | | | 64 | | |
| | | | 64 | | |
| | | | 61 | | |
| | | | 61 | | |
| | | | 61 | | |
| | | | 58 | | |
| | | | 56 | | |
| | | | 52 | | |
| | | | 42 | | |
| | | | 33 | | |
| | | | 25 | | |
| | | | 20 | | |
| | | | 20 | | |
| | | | 16 | | |
| <i>M</i> | 70 | 63.3 | 72.5 | 72.04 | 57.3 |
| <i>SD</i> | 4 | 19.1 | 21.7 | 16.5 | 23.5 |

3). Selected-response tasks present input in the form of an item, requiring examinees to choose the response. These tasks aim to measure recognition or recall. While selected-response tasks are typically designed to measure one area of grammatical knowledge, they may engage more than one area (form and meaning). Constructed-response tasks elicit language production. One type of CR task, limited production tasks, presents input in the form of an item, requiring examinees to produce anywhere from a word to a sentence. These tasks aim to measure emergent production, and are typically designed to measure one or more areas of knowledge. Another type of CR task, extended-production tasks, presents input in the form of a prompt, requiring examinees to produce language varying in quality and quantity. Extended-production tasks are designed to measure full production, eliciting several areas of knowledge simultaneously (Purpura, 2013).

3. Suggestions on the criteria features

From this study, we can conclude that criterial features can be put to use for teaching, learning,

and testing purposes as Hawkins and Filpovic (2012) indicate. Teaching materials and methods, and even testing instruments, can be calibrated to the criterial features of each level. The grammatical properties of English can be presented to learners in ways that are level appropriate, and learners can be encouraged to focus on the features of their target level. The criterial features can help learners prepare for their respective exams (cf. Hawkins & Filpovic, 2012).

Although practical applications of the grammatical criterial features of each CEFR level for syllabus design, preparation of teaching materials, and assessment, will require more information, it is clear that teachers, learners, and testers will benefit greatly from them. The CEFR grammatical criterial features levels should be adjusted to the local teaching or testing context so that they can be more effectively utilized. With this adjustment, the grammatical properties of English can be presented to learners in ways that are level appropriate, and learners can be encouraged to focus on the features of the target level.

| Selected Response Tasks | Constructed Response Tasks | | | |
|-------------------------|----------------------------|------------------------|----------------------------|-------------------------------------|
| | Limited Production | Extended Production | | |
| Noticing | Labeling | <u>Product focused</u> | <u>Performance focused</u> | <u>Process focused observartion</u> |
| Matching | Listing | Essay | <u>simulation</u> | Checklist |
| Same/different | Gap-filling | Report | Role-play | Rubric |
| True/false | Cloze | Project | Improvisation | <u>Reflection</u> |
| Agree/disagree | Sentence completion | Portfolio | Interview | Journal |
| Judgment tasks | Discourse completion | Interview | <u>Recasts</u> | Learning log |
| Multiple choice | Short answer | Presentation | Retelling | Think aloud |
| Ordering | | Debate | Narration | |
| Categorizing | | Play | Summary | |
| Grouping | | | <u>Exchange</u> | |
| | | | Information gap | |
| | | | Problem-solving | |
| | | | Decision-making | |
| Receptive | Emergent | Productive | | |

Figure 2 Response Tasks

We have focused on the correspondence between the tests and the CEFR in terms of reading, and we have argued that correspondences may well be established between the tests and the CEFR in grammar items. However, between the tests and the CEFR, it may be more difficult to establish correspondences in terms of reading.

In conclusion, if we combine grammar items and vocabulary items and put them all in the order of facilities from A1, A2, B1, B2, C1, and C2, we have more A1 and A2 grammar items to distinguish students while we have more C1 and C2 vocabulary items to distinguish students. Also, both grammar and vocabulary items have enough items for B1 and B2. Therefore, for lower level students we could use grammar items, for higher-level students we could use vocabulary items, while for intermediate level students we might choose from both grammar and vocabulary items. Especially for the higher-level students, we should use vocabulary items, as it may be very difficult to provide challenging grammar items compared with challenging vocabulary items at higher levels of study.

Acknowledgement

I am grateful to Adam Murray (Miyazaki International College) and Kazunari Shimada (Takasaki University of Health and Welfare) for their help and support to conduct this research.

Note

This work was supported by the Grant-in-Aid for Scientific Research (KAKENHI) (C) (24520646).

References

Assessment Systems Corporation. Xcalibre 3 [computer software] <http://www.assess.com>
Chapelle, C.A. (2013). (Ed.). *The Encyclopedia of*

- Applied Linguistics*. UK: Wiley-Blackwell.
Coombe, C., Davidson, P., O' Sullivan, B. and Stoyhoff, S. (Eds.). (2012). *The Cambridge Guide to Second Language Assessment*. Cambridge: Cambridge University Press.
Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
Council of Europe (2009). Manual for relating language examinations to the Common European Framework of Reference for Languages (CEFR), Strasbourg: Language Policy Division, retrieved from: http://www.coe.int/t/dg4/linguistic/manuael1_en.asp (February 12, 2013).
English Vocabulary Profile: retrieved from <http://vocabulary.englishprofile.org/dictionary/search/uk/> (September 6, 2013)
Hawkins, J.A. & Filipovic, L. (2012). *English Profile Studies 1: Criterial Features in L2 English*. Cambridge, UK: Cambridge University Press.
Jones, W. (2013). Assessing Students' Grammatical Ability. In Coombe, C., Davidson, P., O' Sullivan, B. and Stoyhoff, S. (Eds.), *The Cambridge Guide to Second Language Assessment* (pp. 247-256). Cambridge: Cambridge University Press.
Larsen-Freeman, D. (2009). Teaching and Testing Grammar. In M.H. Long & C.J. Doughty (Eds.), *The Handbook of Language Teaching* (pp. 519-538). West Sussex, UK: John Wiley & Sons.
Linacre, L. (2005). WINSTEPS (Version 3.55) [Computer software]. <http://www.winsteps.com/winsteps.htm>.
Purpura, J. (2004). Assessing grammar. Cambridge, UK: Cambridge University Press.
Purpura, J. (2009). The impact of large-scale and classroom-based language assessments on the individual. In L. Taylor & C. Weir (Eds.), *Investigating the wider social and educational impact of assessment: Proceedings of the ALTE Cambridge Conference, April 2008* (pp. 301-25). Cambridge, UK: Cambridge University Press.
Purpura, J. (2013). Assessment of Grammar. In C.A. Chapell (Ed.), *The Encyclopedia of Applied Linguistics* Vol. 1 (pp. 195-204). UK: Wiley-Blackwell.
Plakans, L. (2013). Assessment of Integrated Skills. In C.A. Chapell (Ed.), *The Encyclopedia of Applied Linguistics* Vol. 1 (pp. 204-212). UK: Wiley-Blackwell.