慶應義塾大学学術情報リポジトリ
Keio Associated Repository of Academic resouces

Title	A longitudinal analysis of students' vocabulary knowledge and the feasibility of using CEFR levels for a placement test
Sub Title	大学生の語彙力の経年的分析とプレイスメントテストにおけるCEFRレベルの実施可能性
Author	中村, 優治(Nakamura, Yuji) Murray, Adam
Publisher	慶應義塾大学日吉紀要刊行委員会
Publication year	2014
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.46 (2014.) ,p.31- 42
JaLC DOI	
Abstract	
Notes	挿表
Genre	Departmental Bulletin Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN10032394-20141231-0031

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.

A Longitudinal Analysis of Students' Vocabulary Knowledge and the Feasibility of Using CEFR Levels for a Placement Test

Yuji Nakamura
Adam Murray

Introduction

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. As Swender, Tschirner, and Barenfanger (2012) summarize, the CEFR Reading Scales include *Overall Reading Comprehension, Reading Correspondence, Reading for Orientation, Reading for Information and Argument and Reading Instructions* (Council of Europe, 2001). Furthermore, *the Manual for Relating Language Examinations to the Common European Framework of Reference for Languages* (Council of Europe, 2009) contains a chart that lists the salient characteristic: reception (listening and reading). This chart includes a hierarchy of content areas, called 'setting', of text genres, called 'source', and indicates what readers can do and understand at each CEFR level, including the conditions, called 'restrictions', that need to apply for readers to be able to understand (Swender et al., 2012).

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

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

Hawkins and Filpovic (2012) stress the importance of defining grammatical and lexical features of the CEFR levels as fully and accurately as possible. Khalifa and Weir (2009) also claim that it is the lexical and grammatical resources, rather than function, which are more significant in determining level. We can start with or return to an emphasis on grammar and lexis, which can lead to better and more comprehensive illustrative descriptors of different levels. In this paper, however, we will primarily focus on the lexical (and grammatical) properties of English to consider the levels.

The purpose of the paper is to make a small contribution towards establishing correspondences for grammar, vocabulary 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, and vocabulary.

First, 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. In the discussion section, we will focus on the correspondence between the tests and the CEFR in terms of reading, and we will argue that correspondences may well be established between the tests and the CEFR in vocabulary (and grammar) items, but that, between the tests and the CEFR, it may be difficult to establish correspondences in terms of reading.

Purpose of the Study

The purpose of the present study is twofold: 1) to describe the placement test and the CEFR level in terms of reading, and 2) to analyze the vocabulary section of a university placement test and the CEFR levels.

In general, there are two approaches to define 'difficulty' or how difficult it is for EFL learners to acquire vocabulary. One method is based on the 'word level', which is defined as how necessary a word is for reading English. The other is based on the testing approach, in which vocabulary difficulty is explained by test results. In this approach, measurement of reliability and validity are necessary. To do this, a large-scale test must be administered to construct an item difficulty scale for the words. We estimated a standardized item difficulty parameter for each item to compare the level of difficulty in a common scale through placement tests (PTs) and confirmation tests (CTs) from 2006 to 2010.

A scale based on word level can be considered to reflect English reading skills, so our placement tests also measure English reading skills. However, the construct validity for the vocabulary section has not been fully examined yet. In this study, our placement tests do not measure vocabulary skill directly, but assess vocabulary skill by reading English documents that can be seen in college or university classrooms. If the word level approach does not match with the test scale, there might be a need for different scales to measure different factors of vocabulary skill.

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.

Materials/Instruments

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.

Analysis 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. We considered this scale to be a measurement of vocabulary skill in reading English documents. In the IRT analysis, 92 test items were examined. The total number of test items was not 100 because each test form contains common items (anchor items) to apply common-item nonequivalent groups design using IRT analysis. Third, these 92 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.

Results and Discussion

Table 1–a shows the Placement Test level descriptions. The majority of the students can be described as being intermediate level users of English. Table 1–b gives the global descriptions for each of the six CEFR levels. In terms of specific language skills, Tables 1–c and 1–d describe reading comprehension and vocabulary knowledge for the levels. 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. All of these decisions can be made by each individual institution.

Table 2 and Figure a describe the mean score and standard deviation of each level both numerically and graphically. Figures b-f further illustrate that words in each CEFR level have various levels of difficulty. As shown in Figure b, there are 8 items which range in difficulty from 0.26 (difficult) to 0.97 (easy), at the A2 level. At

Table 1–a
Descriptions of Placement Test levels

English 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 1–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

Table 1–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.

User 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.

User 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.

User 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.

(cf. The Council of Europe, 2009)

Table 1–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 specialty, 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.

(cf. The Council of Europe, 2009)

Table 1–d

Description of CEFR lexical levels

VOCABULARY RANGE

C2

Has a good command of a very broad lexical repertoire including idiomatic expressions and colloquialisms; shows awareness of connotative levels of meaning.

C1

Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions; little obvious searching for expressions or avoidance strategies. Good command of idiomatic expressions and colloquialisms.

B2

Has a good range of vocabulary for matters connected to his/her field and most general topics. Can vary formulation to avoid frequent repetition, but lexical gaps can still cause hesitation and circumlocution.

B1

Has a sufficient vocabulary to express him/herself with some circumlocutions on most topics pertinent to his/her everyday life such as family, hobbies and interests, work, travel, and current events. Has sufficient vocabulary to conduct routine, everyday transactions involving familiar situations and topics.

A2

Has a sufficient vocabulary for the expression of basic communicative needs. Has a sufficient vocabulary for coping with simple survival needs.

A1

Has a basic vocabulary repertoire of isolated words and phrases related to particular concrete situations.

(cf. The Council of Europe, 2009)

Table 2
Distribution of Items by Difficulty in Each CEFR Level

Average of each level	0.38	0.53	0.50	0.69	0.75	0.76
Level	off	C2	C1	B2	B1	A2
	0.69	0.83	0.90	0.93	0.90	0.97
Mean =0.59	0.57	0.68	0.86	0.9	0.87	0.86
SD=0.23	0.52	0.62	0.85	0.86	0.84	0.86
	0.51	0.60	0.84	0.84	0.82	0.83
	0.50	0.59	0.78	0.84	0.78	0.8
	0.49	0.58	0.70	0.84	0.78	0.77
	0.45	0.24	0.69	0.83	0.77	0.76
	0.43	0.15	0.62	0.83	0.77	0.26
	0.41		0.57	0.82	0.77	
	0.33		0.54	0.82	0.74	
	0.30		0.54	0.8	0.73	
	0.29		0.54	0.65	0.72	
	0.25		0.54	0.65	0.71	
	0.21		0.49	0.63	0.61	
	0.21		0.49	0.62	0.43	
	0.17		0.48	0.62		
	0.07		0.47	0.61		
			0.46	0.61		
			0.41	0.55		
			0.36	0.5		
			0.35	0.49		
			0.3	0.48		
			0.28	0.48		
			0.23	0.42		
			0.23			
			0.2			
			0.2			
			0.1			
Cf. PT levels		A-H	Advanced	I-H	Intermediate	Basic

the B1 level, one item has a difficulty of 0.43 (Figure c). However, at the C2 level, one item has a difficulty of 0.83 (Figure f), and at the C1 level, four items are over 0.80 (Figure e). This means that it is not necessarily true that higher-level items are more difficult than lower-level items. Generally speaking, each of the five CEFR levels (Figures b – f) contains both easy and difficult items for Japanese students. Similarly, items that are not included in the CEFR levels range in item difficulty (Figure g). In

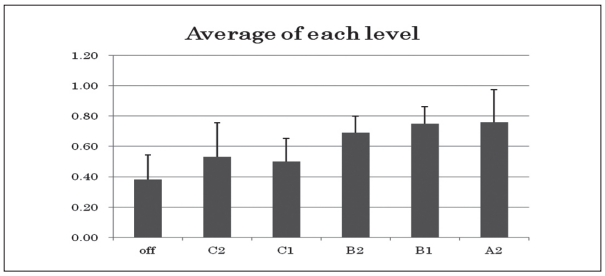

Figure a - Mean and standard deviation of each level

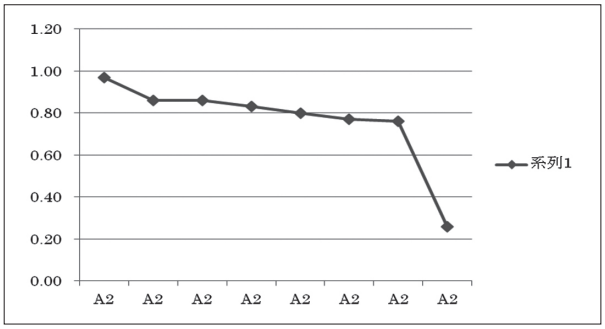

Figure b - Distribution of A2 level items

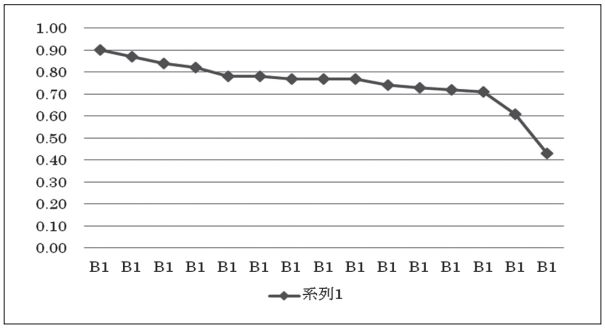

Figure c - Distribution of B1 level items

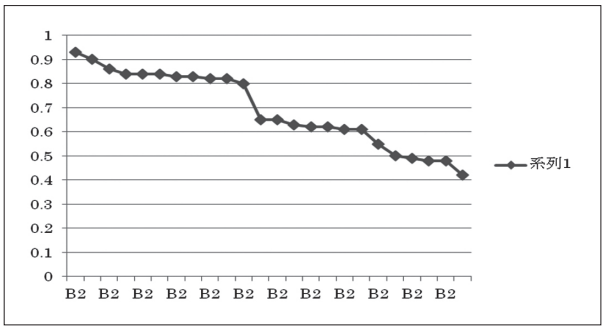

Figure d - Distribution of B2 level items

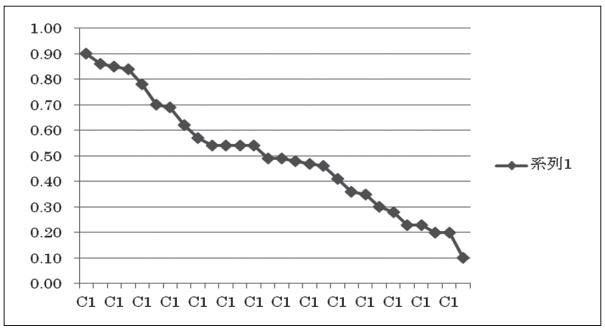

Figure e - Distribution of C1 level items

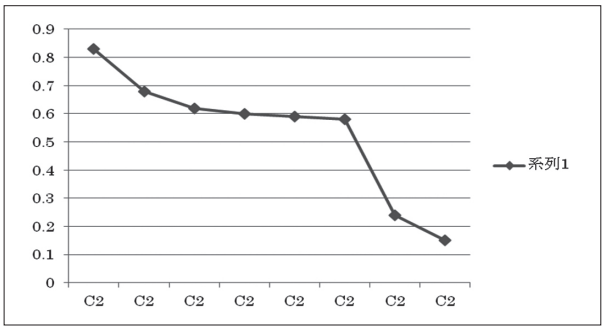

Figure f - Distribution of C2 level items

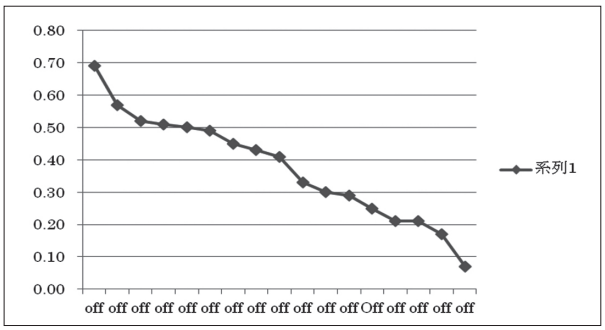

Figure g - Distribution of Off level items

other words, the higher-level words could be easier test items. In fact, one of the easiest items is a C2 level word (0.83). Conversely, there are some lower-level items which were very difficult, with the most difficult item being a B1 word (0.43), or an A2 level word (0.26). Nevertheless, these seem to be rare cases.

More importantly, each level has 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 go either up or down to the next level. The CEFR lexical criterial features levels should be adjusted to the local teaching, learning or testing context to be utilized effectively.

Conclusions and Implications

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) suggest. Teaching materials and methods, and even testing instruments, can be calibrated to the criterial features of each level. The grammatical and lexical properties of English can be presented to learners in ways that are level appropriate, and the 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 and lexical criterial features of each CEFR level for syllabus design, preparation of teaching materials, and assessment will require more information, teachers, learners, and testers will benefit greatly from them. The CEFR lexical 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 and lexical 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.

It may be that correspondences can be established between the tests and the CEFR for vocabulary (and grammar) items, but it may be difficult to establish them in terms of reading between the tests and the CEFR.

Correspondences between tests that are developed on the basis of two different scales can be established. However, linking a test and CEFR in a valid manner requires a great deal more research, including an examination of the use and the social context of the two scales. It may be that the results of one test correspond to the results of another one, but it would mean that correspondences may be established test by test, skill by skill. (cf. Swender et. al, 2012).

To conclude, let us take the six important issues of vocabulary testing (Nation & Stubb, 2011) into consideration, summarized as follows:

1. What is the most appropriate unit of counting when designing a word list or a vocabulary test? (word family or word type (lemma)
2. Should cognates and loanwords be tested?
3. What kind of vocabulary should be tested? The test maker needs to have a policy on what items would be included and not included.

4. What level of difficulty should be tested? Careful thought needs to be given to the item type that is used to make sure that it is suited to the kind of knowledge it is supposed to measure. There should be a trade-off between validity and practicality.
5. Should the testing be group-administered or one-on-one?
6. Should a hard copy or a computerized test be used?

Although some types of multi-word lexical items can be assessed in the same way as single words, others require contextualized test formats. Further study is needed to examine the construct validity of vocabulary and grammar testing for EFL.

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

Acknowledgement:
We are grateful to Haruhiko Mitsunaga (Shimane University) for his help and support to conduct this research.

References

Assessment Systems Corporation. Xcalibre 3 [computer software] http://www.assess.com

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment.* Cambridge, UK: 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.

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

Ekbatani, G. (2011). Measurement and evaluation in post-secondary ESL. New York:
Routledge.

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

Gunning fog index. (n.d.). In Wikipedia. Retrieved from http://en.wikipedia.org/wiki/Gunning_fog_index

Hawkins, J. A., & Filipovic, L. (2012). *EnglishProfile Studies 1: Criterial Features in L2 English*. Cambridge, UK: Cambridge University Press.

Hughes, A. (2003). Testing for language teachers (2nd ed.). Cambridge, UK: Cambridge University Press.

JACET (The Japan Association of College English Teachers). (2003). The JACET List of 8000 Basic Words. Tokyo: JACET.

Kennyon, D. (2012). Using Bachman's assessment use argument as a tool in conceptualizing the issues surrounding linking ACTFL and CEFR. In E. Tschirner (Ed.), *Aligning frameworks of reference in language testing* (pp. 23–34). Tubingen, Germany: Stauffenburg Verlag.

Linacre, L. (2005). WINSTEPS (Version 3.55) [Computer software]. http://www.winsteps.com/winsteps.htm.

Milton, J. (2009). Measuring second language vocabulary acquisition. Bristol, UK: Multilingual Matters.

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

Nation, I. S. P. (2008). Teaching vocabulary: strategies and techniques. Boston: Heinle.

Nation, I. S. P., & Webb, S. (2011). Researching and analyzing vocabulary. Boston, MA: Heinle.

O'Sullivan, B. (Ed.). (2011). Language testing: Theories and practice. Oxford, UK: Palgrave Macmillan.

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

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

Read, J. (2012). Piloting vocabulary tests. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 307–320). UK: Routledge.

Schmitt, N. (2010). Researching vocabulary: A vocabulary research manual. Oxford, UK: Palgrave Macmillan.

Swender, E., Tschirner, E., & Barenfanger, O. (2012). Comparing ACTFL/IRL and CEFR based reading tests. In E. Tschirner (Ed.), *Aligning frameworks of reference in language testing* (pp. 123–138). Tubingen, Germany: Stauffenburg Verlag.

Tschirner, E. (Ed.). (2012). *Aligning frameworks of reference in language testing*. Tubingen, Germany: Stauffenburg Verlag.