jalt journal

The research journal of the Japan Association for Language Teaching

Volume 40 • No. 1 • May 2018

JALT

全国語学教育学会

Japan Association for Language Teaching

¥1,900 ISSN 0287-2420

JALT Journal

Volume 40 • No. 1

May, 2018

Contents

May 2018
Volume 40 • No. 1

Japan Association for Language Teaching

A Nonprofit Organization

The Japan Association for Language Teaching (JALT) is a nonprofit professional organization dedicated to the improvement of language teaching and learning in Japan. It provides a forum for the exchange of new ideas and techniques and offers a means of keeping informed about developments in the rapidly changing field of second and foreign language education. Established in 1976, JALT serves an international membership of approximately 2,700 language teachers. There are 32 JALT chapters and 27 special interest groups (SIGs). JALT is a founder of PAC (Pan-Asian Consortium), which is an association of language teacher organizations in Pacific Asia. PAC holds annual regional conferences and exchanges information among its member organizations. JALT is the Japan affiliate of International TESOL (Teachers of English to Speakers of Other Languages) and a branch of IATEFL (International Association of Teachers of English as a Foreign Language). JALT is also affiliated with many other international and domestic organizations.

JALT publishes *JALT Journal,* a semiannual research journal; *The Language Teacher*, a bimonthly periodical containing articles, teaching activities, reviews, and announcements about professional concerns; and the annual *JALT Post Conference Publication*.

The JALT International Conference on Language Teaching and Learning and Educational Materials Exposition attracts some 2,000 participants annually and offers over 600 papers, workshops, colloquia, and poster sessions. Each JALT chapter holds local meetings and JALT's SIGs provide information and newsletters on specific areas of interest. JALT also sponsors special events such as workshops and conferences on specific themes and awards annual grants for research projects related to language teaching and learning.

Membership is open to those interested in language education and includes optional membership in one chapter and one SIG, copies of JALT publications, and free or discounted admission to JALT-sponsored events. JALT members can join as many additional SIGs as they wish for an annual fee of ¥2,000 per SIG. For information, contact the JALT Central Office or visit the JALT website.

JALT National Officers, 2017–2018

President: . . . . . Richmond Stroupe
Vice President: . . . . . Naomi Fujishima
Auditor: . . . . . Aleda Krause
Director of Treasury: . . . . . Robert Chartrand
Director of Records: . . . . . Maiko K. Nakano
Director of Program: . . . . . Louise Ohashi
Director of Membership: . . . . . Fred Carruth
Director of Public Relations: . . . . . Parisa Mehran

Chapters

Akita, East Shikoku, Fukui, Fukuoka, Gifu, Gunma, Hiroshima, Hokkaido, Ibaraki, Iwate, Kitakyushu, Kobe, Kyoto, Matsuyama, Nagoya, NanKyu, Nara, Niigata, Oita, Okayama, Okinawa, Osaka, Saitama, Sendai, Shinshu, Shizuoka, Tokyo, Tottori, Toyohashi, West Tokyo, Yamagata, Yokohama.

Special Interest Groups

Bilingualism; Business Communication; College and University Educators; Computer-Assisted Language Learning; Critical Thinking; Extensive Reading; CEFR and Language Portfolio; Gender Awareness in Language Education; Global Issues in Language Education; Japanese as a Second Language; Learner Development; Lifelong Language Learning; Literature in Language Teaching; Materials Writers; Mind, Brain, and Education; Mixed, Augmented, and Virtual Realities; Other Language Educators; Pragmatics; School Owners; Speech, Drama, & Debate; Study Abroad; Task-Based Learning; Teacher Development; Teachers Helping Teachers; Teaching Younger Learners; Testing and Evaluation; Vocabulary.

JALT Central Office

Urban Edge Building, 5F 1-37-9 Taito, Taito-ku, Tokyo 110-0016, Japan
Tel.: 03-3837-1630; Fax: 03-3837-1631; Email: jco@jalt.org;
Website: <www.jalt.org>

In This Issue

Articles

The two articles in this issue provide a glimpse of the diverse approaches to research that can appear in *JALT Journal*. In the first, **Tim Stoeckel**, **Tomoko Ishii**, and **Phil Bennett** describe the development and validation of a bilingual version of the New General Service List Test. In the second, **Michiyo Hokamura** uses longitudinal data to examine the development of complexity, accuracy, and fluency in the English writing of two Japanese university students.

Reviews

In this issue, the book reviews introduce seven titles of interest to teacher trainers, language educators, and researchers. In the first review, **George Haikalis** covers an edited anthology of empirical studies investigating team teaching by Local English Teachers (LETs) and Native English-Speaking Teachers (NESTs). The second review by **Caroline Handley** introduces Paul Nation's definition of a word and guide to constructing a word list from a corpus. Third, **Caroline Hutchinson** reviews *Exploring Psychology in Language Learning and Teaching*. In the next review, **Christopher Pirotto** looks at Richard Sampson's latest research into student attitudes and Japanese EFL classroom motivation. **Joseph Poulshock** then examines educating second language teachers from an Oxford Applied Linguistics series title of that very name. Our sixth review is a report on *Linguanomics* by **Dávid Smid**. Lastly, **Bill Snyder** reviews Atsuko Watanabe's case study research into the professional development of Japanese teachers of English through reflective practice.

Editor's Message

After being Associate Editor for 2 years, I am proud that this is my first issue as Editor of *JALT Journal*. I would like to thank the former Editor, Anne McLellan Howard, for all that she has taught me over the last 2 years about editing the journal. As she moves into the position of Consulting Editor, I'm certain that I will continue to be able to rely on her advice. I am also grateful to Greg Sholdt, who will be continuing as Consulting Editor. I would like to thank all those who have submitted articles for consideration in *JALT Journal*, the members of the Editorial Board, our other reviewers, and most definitely the production editors and proofreaders. Without their contributions, it would not be possible to publish the journal.

JALT Journal will continue in its commitment to publish diverse, high quality research relevant for language learning and teaching in the Japanese context. Although such research can be more quantitative or more qualitative, I would especially like to encourage submission of rigorous qualitative research, such as ethnographic research or research on second language interaction, including classroom interaction. Of course, rigorous quantitative research will continue to be welcome. I look forward to 2 (or more) years as Editor.

Articles

A Japanese-English Bilingual Version of the New General Service List Test

Tim Stoeckel
University of Niigata Prefecture
Tomoko Ishii
Meiji Gakuin University
Phil Bennett
University of Niigata Prefecture

This paper describes the development and initial validation of a Japanese-English bilingual version of the New General Service List Test (NGSLT; Stoeckel & Bennett, 2015). The New General Service List (NGSL; Browne, 2013) consists of 2,800 high frequency words and is intended to provide maximal coverage of texts for learners of English. The NGSLT is a diagnostic instrument designed to identify gaps in knowledge of words on the NGSL. The NGSLT is a multiple-choice test that consists of 5 levels, each assessing knowledge of 20 randomly sampled words from a 560-word frequency-based level of the NGSL. A bilingual version of the NGSLT was developed to minimize the risk of conflating vocabulary knowledge with understanding of the answer choices. A validation study with 382 Japanese high school and university learners found the instrument to be reliable (α = .97) and unidimensional and to demonstrate good fit to the Rasch model.

本論文では New General Service List (NGSL) に基づく語彙サイズテスト(NGSLT)の日本語版の開発及び検証を論じる。NGSL (Browne, 2013) は高いテキストカバー率を目指して編集された2800語の高頻度語彙のリストであり、NGSLT (Stoeckel & Bennett, 2015) はそ

のリストについての学習者の知識を診断するテストである。NGSLを560語ごとの５レベルに分割し、各レベルから20語を無作為に抽出し計100問の多肢選択式のテストを作成した。選択肢の理解不足によって不正解になる懸念があるため、日本語版を作成した。大学生・高校生合わせて382人の学習者による検証により、この日本語版の信頼性が高いこと（α = .97）、測定が一次元的に行われていること、またラッシュモデルに適合することが確認された。

Keywords: bilingual tests; New General Service List; New General Service List Test; Rasch model; second language vocabulary testing

Vocabulary is now widely regarded as a critical component of L2 learning (Hunt & Beglar, 2005) with research revealing a close relationship between lexical knowledge and the skills of reading, writing, listening, and speaking (Milton, Wade, & Hopkins, 2010). In the case of reading, learners need to have knowledge of the basic form–meaning relationship of approximately 98% of the running words in a written text to facilitate unassisted comprehension (Schmitt, Jiang, & Grabe, 2011). Because most of this coverage is provided by high-frequency vocabulary, which is typically defined as the most frequent 2,000-3,000 word families (Nation, 2013; Schmitt & Schmitt, 2012), it is important for both pedagogy and research that reliable instruments to measure knowledge of such words be developed. This paper introduces one such instrument. We begin with overviews of two high-frequency vocabulary lists, the General Service List (GSL; West, 1953) and its modern replacement, the New General Service List (NGSL; Browne, 2013). We then outline the development and initial validation evidence of a Japanese–English bilingual version of a diagnostic test of high-frequency terms sampled from the NGSL.

West's Original General Service List

For many years, West's (1953) General Service List was used pedagogically to provide coverage of high-frequency words. The GSL consists of approximately 2,000 headwords plus their related constituents (e.g., *nation* plus *nations*, *national*, *nationally*, and *nationwide*). The criteria for inclusion of words in the GSL included both frequency and a subjective evaluation of how useful each word would be for L2 learners of English. Since its development, the GSL has been used in the creation of both learning materials, such as graded readers, and vocabulary assessment instruments, such as the Vocabulary Levels Test (Schmitt, Schmitt, & Clapham, 2001).

Despite its usefulness, the GSL was probably due for substantial revision. The approximately 2.5-million-word corpus used to create the list is small

by modern standards, meaning the GSL may not be representative of a wide range of language use. Additionally, the list itself has aged and is no longer completely reflective of modern lexis. For example, it contains items such as *telegraph* and *mankind*, which have gradually faded from use, but other terms that have become more widely used such as *computer* and *climate* are absent. Finally, the organization of the GSL may not be optimal for the types of learners (i.e., those of relatively low proficiency) most likely to benefit from a list of high-frequency vocabulary. Originally, entries were grouped as headwords plus inflected forms as well as many frequent and regular derivatives. This was later standardized by Bauman and Culligan (1995) so that each entry included all word forms through level 4 of Bauer and Nation's (1993) word family levels. Though word families appear to be actual psychological constructs (see, e.g., Nagy, Anderson, Schommer, Scott, & Stallman, 1989), and adult native speakers of English familiar with one member of a word family are also likely able to recognize other family members (Tyler & Nagy, 1989), this may not hold true for nonnative speakers unless they are highly proficient in the L2 (Gardner, 2007; McLean, Nation, Pinchbeck, Brown, & Kramer, 2016).

From the perspective of teaching and learning, this means that the burden for mastery of the GSL consists of learning not only each headword but also other family members that may not be readily recognizable from knowledge of the headword. However, for many word families, constituents differ greatly in the text coverage they provide. For instance, in the word family for *hard*, the constituents *hard*, *harder*, and *hardest* represent approximately 95.5% of occurrences in the Corpus of Contemporary American English (COCA; http://corpus.byu.edu/coca/), while the remaining seven word forms (*harden*, *hardness*, *hardship*, plus inflected forms) provide little additional coverage. When the improvement in coverage is so limited, learning those word family members should be deprioritized in favor of learning other more frequently occurring word forms. From the perspective of assessment, if learning one member of a word family does not automatically result in the ability to recognize other family members for many learners, then when tests are configured to measure knowledge of a sampling of headwords, correct responses do not necessarily indicate knowledge of all related family members.

The New General Service List

To address these limitations, a New General Service List was developed by Browne, Culligan, and Phillips (Browne, 2013). The NGSL is derived from an

analysis of a much larger 273-million-word subset of the Cambridge English Corpus (CEC). The complete CEC is comprised of materials with both reported (71.0% of the corpus) and unreported (29.0%) publication dates. Of the former, 85.1% of the words are derived from sources dated 2000 or later (S. Grieves, Cambridge University Press, personal communication, September 23, 2016). This, combined with the fact that the subset of the CEC used to create the NGSL was carefully balanced to include both written and spoken discourse from nine separate subcorpora (Browne, 2013), suggests that the NGSL is representative of a more modern and broader range of English language use than the GSL. Additionally, entries in the NGSL are organized into "modified lemmas" rather than word families. A regular lemma consists of a headword plus its inflected (but not derived) forms. Under the headword *hard*, for example, are only the inflected forms *harder* and *hardest*. In a "modified lemma," orthographically identical headwords and their constituent derivations are grouped together. For instance, the modified lemma for *approach* consists of the nominal inflection *approaches* plus the verbal inflections *approaches*, *approaching*, and *approached*. This modified lemma grouping facilitates accurate text analysis with tools such as the Lextutor VocabProfilers (http://www.lextutor.ca/vp/), which are currently incapable of distinguishing between orthographically identical word forms that belong to separate lemma groupings. It also aligns relatively well with Gardner's (2007) recommendations that words be grouped as base forms plus regular inflections for low-proficiency learners and extended to include irregular inflections and derivational prefixes for those at an intermediate level.

In total, there are 2,800 modified lemmas in the NGSL, ordered according to frequency and dispersion across the various subcorpora used to create the list (Browne, 2014). Though this figure exceeds the 2,000 word families in the original GSL, it may represent a smaller learning burden than the approximately 3,600 lemmas present in the GSL (Browne, 2013). In terms of coverage, at 90.34%, the NGSL offers about 6% more coverage of the subset of the CEC used to create the list than does the original GSL (84.24%; Browne, 2013).

The English Version of the New General Service List Test

The New General Service List Test (NGSLT) is a diagnostic instrument designed to identify gaps in learners' written receptive knowledge of words on the NGSL, to assist in setting vocabulary learning goals, and to aid in designing lexically appropriate educational experiences (Stoeckel & Bennett,

2015). This section describes the test in its original monolingual English format.

The test consists of 100 items, 20 for each of five 560-word frequency-based levels of the NGSL. Items are written with specifications similar to those of the Vocabulary Size Test (VST; Nation & Beglar, 2007). Thus, a multiple-choice format is used in which item stems consist of a target word followed by a short sentence using the word. These decontextualized sentences are intended to indicate the tested word's part of speech, help examinees view it as an authentic element of language, and provide "a little extra associational help in accessing the meaning" (Nation & Beglar, 2007, p. 11). Four answer choices follow in the form of short definitions or synonyms of the tested word and of three other words of similar frequency. To keep answer choices as simple and intelligible as possible, they are written with high-frequency vocabulary. Specifically, whenever possible, items testing words in the first three levels of the NGSL were written only with words from the first two levels. Moreover, items testing words in the fourth and fifth levels were written exclusively with words of higher frequency than the target word. Because of these restrictions, the correct answer is worded only specifically enough to distinguish it from the three distractors. For instance, for the item testing the word *slide*, the correct answer *move* defines *slide* in terms that are only precise enough to distinguish it from the distractors *break*, *make power*, and *become bigger* (for details see Stoeckel & Bennett, 2015). This approach is unavoidable for items testing words that cannot succinctly be defined with high-frequency vocabulary, but it means that examinees are sometimes not required to demonstrate precise meaning recognition. (As described below, the bilingual test format overcomes this limitation by using a direct translation of most tested words.)

The NGSLT is designed to determine whether examinees have made an initial link between the form and meaning of each tested word. To increase the accuracy of the test in this regard, three steps were taken during item writing, each of which was informed by frequency counts or a tally of concordance lines in the COCA. First, when the modified lemma included more than one part of speech, the most frequently occurring part of speech was the form that was tested. Thus, for the headword *approach*, the noun form was tested because it occurs more frequently than the verb form (shown in Figure 1). Second, for the example sentence in most item stems, the most frequently occurring word form for the tested part of speech was used. For *approach*, this meant the singular *approach* rather than the plural *approaches* was utilized. Third, when the tested word had multiple meanings, the most frequently oc-

curring sense was used to define the word. For *approach*, the definition was "way of doing something" rather than "movement toward something" or other less common meanings. This use of frequency as a guide in item writing was intended to increase the likelihood that examinees would be familiar with the word form and meaning sense used in the test.

approach: We like your **approach**.
a. way of doing something
b. part of a book
c. house and land
d. facts and information

Figure 1. Example item from the monolingual version of the NGSLT.

Japanese–English Bilingual Version of the NGSLT

For tests of written receptive vocabulary knowledge such as the NGSLT, bilingual formats have become popular because they are thought to reduce the risk of scores being influenced by poor knowledge of the syntax or vocabulary used in the answer choices (Elgort, 2013; Karami, 2012; Nguyen & Nation, 2011). Supporting this view, research with the VST has shown that scores are generally higher with bilingual versions than with the monolingual variant (Elgort, 2013; McDonald, 2015). Because the NGSLT is intended for low- and intermediate-level learners who may have limited grammatical knowledge or gaps in knowledge of even high-frequency vocabulary, a bilingual format may be particularly suitable for enabling examinees to fully demonstrate their actual lexical knowledge. In terms of washback, the bilingual format may be beneficial in encouraging learners to utilize the L1 to establish initial form–meaning linkage, an approach that is often more effective than using L2 definitions (see Schmitt, 2008).

In developing the Japanese–English bilingual version of the NGSLT, Nguyen and Nation's (2011) guidelines for writing such tests were adopted. Thus, answer choices are usually not direct translations of the definitions that are used in the English version but are instead translations of the words that are defined by the answer choices in the English version. For instance, in the monolingual item shown in Figure 2, "hit this hard" defines *knock*, "follow this" defines *pursue*, and "exchange this for something" defines *switch;* these three words and the target word *justify* are translated directly into Japanese in the bilingual version.

justify: We cannot **justify** this. a. hit this hard b. follow this c. exchange this for something else d. show that this is right	**justify**: We cannot **justify** this. a. 打ち砕く b. 追求する c. 変更する d. 正当化する

Figure 2. A comparison of monolingual and Japanese-English bilingual formatting in the NGSLT.

In addition, four conventions employed by McLean, Ishii, Stoeckel, Bennett, and Matsumoto (2016) in their revisions to the Japanese–English version of the VST were used here. First, when the answer choices in the monolingual version defined words that exist as loanwords in Japanese, the *katakana* forms of these words were avoided in the bilingual format to prevent the use of phonological matching to guess the correct answer or to eliminate distractors. Instead, paraphrases or alternative words with Japanese etymological origins were used. For instance, for the tested word *hall*, rather than using the *katakana* ホール (*hooru*), which is phonologically similar to the tested word, the alternative 集会場 (*shukaijo*) was used. Second, consistency in part of speech and inflection was pursued across the four answer choices. For instance, though Japanese adjectival forms have several possible endings (e.g., ～い [*-i*], ～な [*-na*], ～である[*-dearu*]), an effort was made to use the same ending across the four options wherever it was possible and sounded natural. Third, the answer choices were written so that none would stand out from the others due to a difference in length. Finally, the wording of each item stem was checked to make sure that it sounded natural together with each of the four answer choices from the point of view of a native speaker of Japanese. For this purpose, passive verbs in some item stems were changed to the active voice, which is less awkward in Japanese. For instance, the item stem "**hide:** It was **hidden**" in the monolingual version was changed to "**hide:** Please **hide** these" in the bilingual variant, so that the base form 隠す (*kakusu,* meaning *hide*) could be used instead of its inflected passive form 隠された (*kakusareta,* meaning *hidden*).

Initial Validation Evidence

An initial validation study of the Japanese–English bilingual version of the NGSLT was conducted with a convenience sample of 386 native speakers of

Japanese. These included 285 1st-year students at a prefectural university in northern Japan and 101 students in their 1st, 2nd, or 3rd year of study from three high schools in southern Japan.

The instrumentation consisted of Form A of the Japanese–English bilingual variant of the New General Service List Test (available from the authors).[1] For the university participants, the test was administered by computer via the ClassMarker website (https://www.classmarker.com), which prevented examinees from skipping test questions.[2] For the high school participants, the test was administered in a paper and pencil format because computers were unavailable. These examinees were instructed to answer all test questions; for items testing unknown words, the instructions were to make a best guess after carefully reading all answer choices. Four of these participants responded to fewer than 70% of test items and were therefore removed from the study (final n = 382). For the remaining high school participants, there were a total of 10 unanswered questions (by nine separate persons). The data were analyzed with these values missing and again with imputed randomly generated answers (see Garson, 2012), and there were no discernible differences in any of the analyses. Reported here are the results with imputed answers.

Estimates of reliability were satisfactory for the entire sample (α = .97) and for the separate groups of university (α = .80) and high school (α = .89) students (Table 1). The reliability coefficients for the five individual test levels ranged from .82 to .88 (Table 2), suggesting that this version of the NGSLT also provides reliable estimates of vocabulary knowledge at each level of the test. Table 2 also shows that mean scores across the five levels of the test were consistent with the frequency-based model of vocabulary acquisition in which there is a general trend for higher frequency words to be learned before less commonly occurring lexis (Milton, 2009). Scores of the university students (n = 285) were also found to correlate moderately with performance on the Computerized Assessment System for English Communication (CASEC; http://global.casec.com), a test of general English proficiency, r = .586, p < .001.

Table 1. Descriptive Statistics

Group	*n*	*M*	*SD*	α
All	382	80.1	18.1	.97
University	285	89.4	6.0	.80
High School	97	52.6	12.8	.89

Note. There were 100 items on the test.

Table 2. Descriptive Statistics by Test Level

Test Level	*M*	*SD*	α
1	17.0	3.5	.87
2	17.0	3.8	.88
3	15.7	3.9	.86
4	15.2	3.7	.82
5	15.2	4.3	.86

Note. There were 20 items in each level of the test.

Rasch analysis was performed with the Winsteps software (version 3.92) to assess person fit, item fit, construct dimensionality, and the responsiveness of the instrument to changes in the measured construct. Using standardized outfit values > 2.0 as the criterion for person and item misfit (Wolfe & Smith, 2007), 27 persons were flagged as misfitting the model and were temporarily removed. With this smaller dataset (n = 355), 20 items were identified as having poor fit to the Rasch model. To investigate reasons for this high number of misfitting items, the 27 persons were reinstated and a principle component analysis (PCA) of Rasch residuals was conducted to inspect person and item dimensionality. Rasch analysis identifies the primary dimension in a dataset, presumably lexical knowledge in the case of the NGSLT, and a nonrandom pattern in the residuals would be indicative of a secondary dimension (Linacre, 2007). Using Stevens' (2002) criteria, meaningful dimensions in the PCA were defined as those with components having 10 or more loadings above .40, four or more above .60, or at least three above .80. No secondary dimension was identified in the item residuals. However, in the PCA of person residuals 11 persons (all university students) had loadings above .40 and 16 (all high school students) below -.40 in the first contrast.

This suggested the possibility of differential item functioning (DIF), a situation in which different groups of examinees respond differently to one or more test items even after differences in ability are accounted for (Zumbo, 1999). Thus, a separate calibration *t*-test approach was employed to explore DIF between the high school and university groups. Based on Linacre (1994), the criteria for DIF was set at $p < .01$ and effect size > 1.0 (defined as the group difference in Rasch item measures). Thirty items were found to demonstrate DIF. Fourteen were more difficult than expected for university students and 16 for high school students. Items displaying DIF were examined to determine whether they were sources of unfair bias or were instead indicative of real differences in knowledge between the groups (Zieky, 2006). Consistent with the general pattern of vocabulary growth from high school to university, all test items, including those exhibiting DIF, were more difficult for the high school group in *absolute* terms. For example, although Q74 *currency* exhibited DIF in favor of the high school group ($p < .001$, DIF contrast = -2.79), the university students still performed slightly better in absolute terms (38.2% versus 36.1% correct). Moreover, an examination of each DIF item revealed no obvious cause of unfair bias between the two groups. From this we tentatively conclude that DIF with the present sample was caused by actual dissimilarities in lexical knowledge, perhaps due to differences in curriculum or language exposure.

To remove the effect of DIF in assessing item fit, a separate analysis was conducted for each group. For the high school group, six items misfit the Rasch model (standardized outfit values in parentheses): Q21 *observe* (2.7), Q25 *extra* (2.7), Q27 *solution* (2.1), Q49 *guarantee* (3.6), Q67 *impose* (2.6), Q 99 *accurate* (2.4); for the university group, there were four misfitting items: Q1 *charge* (4.8), Q26 *instance* (2.5), Q52 *label* (2.9), and Q79 *shadow* (2.2). These items will be monitored in future test administrations; however, considering that for each group approximately five items should exceed a standardized outfit value of 2.0 by chance alone, these findings suggest acceptable fit to the Rasch model when the effect of DIF is removed.

Instrument responsiveness, the capacity of an instrument to detect changes in a measured construct (Wolfe & Smith, 2007), was assessed visually with the person–item map shown in Figure 3. Persons are arranged on the left according to ability, and items are arranged on the right according to difficulty. The numbers along the left margin represent the logit-based scale for both person and item measures. In the present dataset, item measures from approximately -3 to +3 logits indicate that the instrument provided coverage over a range of person abilities and is sensitive to changes in the

measured construct. The presence of test items well below the lowest ability person indicates that the test did not have a floor; however, the presence of a number of learners above the most difficult item suggests a ceiling effect as learners gain mastery of the NGSL.

Taken together, the findings of this initial investigation indicate that the instrument appears to have sound measurement properties and behaves in ways that are expected by theoretical understanding of the tested construct. However, the presence of DIF between high school and university learners warrants further investigation.

Score Interpretation and Test Use

The NGSLT assesses written receptive vocabulary knowledge, the kind of lexical knowledge needed for reading (Stoeckel & Bennett, 2015). As stated above, it assesses knowledge of form–meaning linkage. It does not evaluate aspects of vocabulary depth such as collocation or register. Additionally, it does not measure lexical comprehension in listening, nor should it be used to assess productive vocabulary in speaking or writing. Moreover, score interpretations need to account for the overestimation of lexical knowledge due to the use of test-taking strategies and blind guessing that has been observed to occur in multiple-choice vocabulary tests (Gyllstad, Vilkaitė, & Schmitt, 2015). For bilingual multiple-choice tests of written receptive vocabulary knowledge, recent research found that this overestimation is equivalent to approximately 40 to 45% of unknown words (Stoeckel, 2016; Stoeckel & Stewart, 2016).

With this in mind, a good way to use the test is to examine the pattern of scores across the five test levels and to use the point at which scores drop and stay below a threshold of 85 to 90% (i.e., 17 or 18 correct out of 20 at a given level) as a target for intentional vocabulary study. This threshold, and not 100%, is based in part on work by Milton (2009) indicating that even high proficiency learners commonly have small gaps in knowledge of high-frequency vocabulary. The threshold also allows for a small amount of miskeying on the part of examinees. Factoring in the use of test strategies and random guessing, this benchmark represents a level of mastery of perhaps 75% of the entries in a 560-word band, meaning a gap in knowledge of at least 140 words. Because of the importance of high-frequency vocabulary, scores below this threshold indicate a need to study and learn the unknown words at the level in question.

Rasch Measure (logits)	Persons (more ability)	Items (more difficult)
6		
	UUU	
5		
	UUUUUU	
	UUUUUUUUUUUUUU	
4	UUUUUUUUUUUUUUU	
	UUUUUUUUUUUUUUUUUUUU	
	UUUUUUUUUUUUUUUUUUUUUUUUUU	
	UUUUUUUUUUUUUUUUUUUUUUUU	
	UUUUUUUUUUUUUUUUUUUUUUUUUUUU	Q1 Q74
3	UUUUUUUUUUUUUUUUUUUUUUUUUU	Q67
	UUUUUUUUUUUUUUUUUUUUUUUUUUU	
	UUUUUUUUUUUUUUUUUUUUU	Q60 Q68 Q93
	UUUUUUUUUUUUUUUUUUUUUUUU	Q43 Q58
	UUUUUUUUUUUUUUUUUUUUUUUUUUU	Q70 Q76
2	UUUUUUUU	Q10 Q52 Q89 Q94
	UUUUUUUU	
	HHHHUUUU	Q91
	HHUUUU	Q35
	HHHHHHUUUU	Q24 Q47 Q71 Q80 Q84
1	HHHHHH	Q18 Q26 Q49 Q50 Q61 Q85 Q87 Q99
	HHHHHHHHH	Q42 Q90
	HHHHHHHHH	Q13 Q33 Q92 Q97
	HHHHH	Q21 Q31 Q54 Q62 Q63 Q98
	HHHHHHHHH	Q10 Q11 Q19 Q3 Q40 Q5 Q7
0	HHHHHHHHHHH	Q22 Q25 Q27 Q2 Q37 Q38 Q46 Q57 Q66 Q96
	HHHHHHHH	Q17 Q30 Q56 Q77 Q78 Q81
	HHHHHHH	Q45 Q59 Q69
	HHHHH	Q48 Q72 Q83
	HHH	Q23 Q29
-1	HHH	Q6 Q88 Q95
	HHHH	Q20 Q53
	HH	Q55 Q65 Q8
	H	Q12 Q15 Q16 Q73
		Q39 Q51 Q79 Q82 Q9
-2		Q32 Q34
		Q28 Q36
		Q41 Q44 Q64
		Q75
-3		Q86
		Q4
-4		Q14
	(less ability)	(less difficult)

Figure 3. A Rasch person–item map showing the person ability and item difficulty measures for the high school and university students in the validation study. The persons are designated "U" for university and "H" for high school participants. The numbers along the left-hand side are the Rasch-based logit values for persons and items.

Additional Resources

Once a level of the NGSL has been identified for intentional study, learners may benefit from several additional steps. First, it would be useful for examinees to review a complete list of headwords from the targeted NGSL level and to highlight unfamiliar words. This can serve as an initial step in awareness raising and can help to verify whether the target level matches learner needs. Second, it would be valuable for learners to begin a principled program of study of the targeted words (Hunt & Beglar, 2005; Schmitt, 2008). For this, there are a number of free, well-designed resources available on the NGSL website (http://www.newgeneralservicelist.org/), including several different English-only and Japanese–English bilingual spaced repetition flashcard applications that can be used on both PCs and smartphones. Third, periodic reassessment would be helpful to monitor lexical growth, to see whether learning goals have been met, and to guide further goal-setting.

To this end, Form B of the Japanese–English bilingual NGSLT has been completed, and Forms C and D are under development. These instruments use the same test blueprint but each assesses knowledge of 100 different randomly sampled NGSL words. A comparison of the relative difficulty of Forms A and B has been conducted with a sample of Japanese learners using a Rasch-based anchor item approach (see Wolfe, 2000). This examination suggests that Forms A and B yield similar but not identical scores for learners of the same ability. The tests could therefore be considered equivalent for low stakes diagnostic and classroom use but need more careful scrutiny and perhaps a reassignment of some items across the two forms before they could be considered equivalent for high stakes purposes.

Conclusion

This paper introduced and described the ongoing development of a Japanese–English bilingual version of the New General Service List Test. The initial validation evidence presented here suggests that the instrument provides a psychometrically sound measure of written receptive knowledge of words on the NGSL. Test results can be used to establish learning goals, to monitor progress in achieving those goals, and to ascertain whether educational materials are lexically suitable for a given group of learners. Both the monolingual and bilingual versions of Forms A and B of the NGSLT are freely available from any of the authors' Academia.edu profile pages.

Notes

1. The test has been periodically revised to reflect both updates to the NGSL and item-performance data collected from ongoing testing. The version used in the present study is dated March 2016.
2. Research on instructions to skip unknown words or the addition of "I don't know" as an answer choice in multiple-choice vocabulary tests shows that examinees use such conventions differentially (Bennett & Stoeckel, 2012; Stoeckel & Stewart, 2016), which introduces "willingness to skip" as a nonrelevant construct impacting test scores and weakening test validity (Stoeckel, Bennett, & McLean, 2016).

Tim Stoeckel is an associate professor at the University of Niigata Prefecture. His primary interests are in L2 reading and vocabulary development and testing.

Tomoko Ishii is a lecturer at Meiji Gakuin University. Her primary research interests are in memory studies and vocabulary development.

Phil Bennett is an associate professor at the University of Niigata Prefecture. His research interests are mainly in vocabulary development, testing, and metaphorical vocabulary acquisition.

References

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography, 6*, 253-279. https://doi.org/10.1093/ijl/6.4.253

Bauman, J., & Culligan, B. (1995). *The general service list*. Retrieved from http://jbauman.com/aboutgsl.html

Bennett, P., & Stoeckel, T. (2012). Variations in format and willingness to skip items in a multiple-choice vocabulary test. *Vocabulary Education & Research Bulletin, 1*(2), 2-3. Retrieved from https://jaltvocab.weebly.com/uploads/3/3/4/0/3340830/verb-vol1.2.pdf

Browne, C. (2013). The New General Service List: Celebrating 60 years of vocabulary learning. *The Language Teacher, 37*(4), 13-16. Retrieved from https://jalt-publications.org/files/pdf-article/37.4tlt_featureds.pdf

Browne, C. (2014). A new general service list: The better mousetrap we've been looking for? *Vocabulary Learning and Instruction, 3*(2), 1-10. Retrieved from http://www.charlie-browne.com/wp-content/downloadable-files/vli130026.pdf

Elgort, I. (2013). Effects of L1 definitions and cognate status of test items on the Vocabulary Size Test. *Language Testing, 30,* 253-272. https://doi.org/10.1177/0265532212459028

Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics, 28,* 241-265. https://doi.org/10.1093/applin/amm010

Garson, D. (2012). *Missing values analysis and data imputation.* Asheboro, NC: Statistical Publishing.

Gyllstad, H., Vilkaitė, L., & Schmitt, N. (2015). Assessing vocabulary size through multiple-choice formats: Issues with guessing and sampling rates. *ITL - International Journal for Applied Linguistics, 166,* 278-306. https://doi.org/10.1075/itl.166.2.04gyl

Hunt, A., & Beglar, D. (2005). A framework for developing EFL reading vocabulary. *Reading in a Foreign Language, 17,* 23-59. Retrieved from http://nflrc.hawaii.edu/rfl/April2005/hunt/hunt.pdf

Karami, H. (2012). The development and validation of a bilingual version of the Vocabulary Size Test. *RELC Journal, 43,* 53-67. https://doi.org/10.1177/0033688212439359

Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions, 7,* 328.

Linacre, J. M. (2007). *Dimensionality: Contrasts and variances*. Retrieved from http://www.winsteps.com/winman/principalcomponents.htm

McDonald, K. (2015). The potential impact of guessing on monolingual and bilingual versions of the Vocabulary Size Test. *Osaka JALT Journal, 2,* 44-61. Retrieved from http://www.osakajalt.org/storage/Osaka_JALT_Journal_2015.pdf

McLean, S., Ishii, T., Stoeckel, T., Bennett, P., & Matsumoto, Y. (2016). An edited version of the first eight 1,000-word frequency bands of the Japanese-English version of the Vocabulary Size Test. *The Language Teacher, 40*(4), 3-7. Retrieved from https://jalt-publications.org/files/pdf-article/40.4-tlt-art1.pdf

McLean, S., Nation, P., Pinchbeck, G. G., Brown, D., & Kramer, B. (2016). Revisiting the word family: What is an appropriate lexical unit for Japanese EFL learners? Paper presented at the Vocab@Tokyo Vocabulary Conference, Tokyo, Japan. Abstract retrieved from https://drive.google.com/file/d/0B6pqkaroKu33OWZFR3RJX21ZSVU/view

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

Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in English as a foreign language. In R. Chacón-Beltrán, C. Abello-Contesse, & M. Torreblanca-López (Eds.), *Insights into non-native vocabulary teaching and learning* (pp. 83-98). Bristol, England: Multilingual Matters.

Nagy, W., Anderson, R. C., Schommer, M., Scott, J. A., & Stallman, A. C. (1989). Morphological families in the internal lexicon. *Reading Research Quarterly, 24,* 262-282. https://doi.org/10.2307/747770

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge, England: Cambridge University Press.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13. Retrieved from http://jalt-publications.org/tlt/issues/2007-07_31.7

Nguyen, L. T. C., & Nation, I. S. P. (2011). A bilingual vocabulary size test of English for Vietnamese learners. *RELC Journal, 42,* 86-99. https://doi.org/10.1177/0033688210390264

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research, 12,* 329-363. https://doi.org/10.1177/1362168808089921

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95,* 26-43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Schmitt, N., & Schmitt, D. (2012). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching, 47,* 484-503. https://doi.org/10.1017/S0261444812000018

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18,* 55-88. https://doi.org/10.1177/026553220101800103

Stevens, J. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Erlbaum.

Stoeckel, T. (2016, September). A serial multiple-choice format to reduce overestimation of meaning recall knowledge on the Vocabulary Size Test. Paper presented at the 20th Anniversary Conference of the Japan Language Testing Association, Kanagawa, Japan.

Stoeckel, T., & Bennett, P. (2015). A test of the New General Service List. *Vocabulary Learning and Instruction, 4*(1), 1-8. Retrieved from http://vli-journal.org/wp/wp-content/uploads/2015/10/vli.v04.1.2187-2759.pdf

Stoeckel, T., Bennett, P., & McLean, S. (2016). Is "I don't know" a viable answer choice on the Vocabulary Size Test? *TESOL Quarterly, 50,* 965-975. https://doi.org/10.1002/tesq.325

Stoeckel, T., & Stewart, J. (2016, September). The "I don't know" option and L1 answer choices: A comparison of four variants of the Vocabulary Size Test. Paper presented at the Vocab@Tokyo Vocabulary Conference, Tokyo, Japan. Abstract retrieved from https://drive.google.com/file/d/0B6pqkaroKu33OWZFR3RJX21ZSVU/view

Tyler, A., & Nagy, W. (1989). The acquisition of English derivational morphology. *Journal of Memory and Language, 28,* 649-667. https://doi.org/10.1016/0749-596X(89)90002-8

West, M. (1953). *A general service list of English words.* London, England: Longman, Green.

Wolfe, E. W. (2000). Understanding Rasch measurement: Equating and item banking with the Rasch model. *Journal of Applied Measurement, 1,* 409-434.

Wolfe, E. W., & Smith, E. V. (2007). Instrument development tools and activities for measure validation using Rasch models: Part II—validation activities. *Journal of Applied Measurement, 8,* 204-234.

Zieky, M. J. (2006). Fairness review in assessment. In S. M. Downing, & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 359-376). Mahwah, NJ: Erlbaum.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF).* Retrieved from http://faculty.educ.ubc.ca/zumbo/DIF/

The Dynamics of Complexity, Accuracy, and Fluency: A Longitudinal Case Study of Japanese Learners' English Writing

Michiyo Hokamura
Temple University, Japan

In this study I investigated the development of complexity, accuracy, and fluency (CAF) in the English writing of 2 Japanese university students of elementary-to-low-intermediate English proficiency over the course of 1 academic year. Data were analyzed from a dynamic systems theory (DST) perspective using min-max graphs, moving correlation coefficients, and a Monte Carlo analysis. The findings show that (a) the CAF measures followed nonlinear trajectories in their development, (b) they often did not correlate positively with one another, and (c) they sometimes showed signs of improvement during phases of high intravariability. The findings suggest that to understand the developmental trajectories and processes involved in language development more deeply, cross-sectional studies of developmental trends of groups should be supplemented with longitudinal studies that examine language development in individual learners.

本研究では日本人大学生（初級から中級下レベル）二名の英文ライティングの熟達度を一年間にわたり、複雑さ、正確さ、流暢さ (complexity, accuracy, and fluency: CAF) の指標から調査を行った。データ分析はダイナミックシステム理論の観点から、ミニマックスグラフ、移動相関係数、モンテカルロ分析を使用して行った。その結果、(a) 両者のＣＡＦの発達軌跡は非線形であり、(b) ＣＡＦのすべてが互いに正の相関関係をとることはほぼなかった。さらに (c) 学習者個人内のＣＡＦの軌跡が大きく変動した後に発達の兆候がみられることがあった。本結果から、言語発達の軌跡やそのプロセスをさらに理解するためには、グループ間を比較する横断研究を、個々の学習者を調査する長期的研究で補完することを提案する。

JALT Journal, Vol. 40, No. 1, May 2018

Keywords: dynamic systems theory; foreign language writing; longitudinal case study

Second language acquisition (SLA) research is largely concerned with language development in individual learners, but most SLA studies have been focused on general developmental trends of groups. Norris and Ortega (2009) argued that our knowledge about the developmental trends of groups needs to be supplemented with illustrations of individual variation to provide a fuller view of developmental trajectories of learning. Understanding language development in individual learners requires the use of longitudinal designs and more dynamic descriptions explaining variability and nonlinearity in individual learner development (Vyatkina, Hirschmann, & Golcher, 2015). Some studies of L2 writing have used complexity, accuracy, and fluency (CAF) to measure language development. Most such studies, however, have been cross-sectional and contrastive rather than developmental and longitudinal, focusing on only one or two constructs at a time. As a result they have failed to consider CAF as a dynamic and interrelated set of constantly changing subsystems (Norris & Ortega, 2009). To elucidate language development in individual learners, I investigated the development of CAF in English writing from a dynamic systems theory (DST; Abraham, 1994; Lorenz, 1963; Reynolds, 1987; Thom, 1975) perspective in this longitudinal case study.

Dynamic Systems Theory

Systems are sets of interacting variables—also called subsystems—that are themselves parts of a larger system. Dynamic systems are systems that change over time. In many publications the terms *chaos*, *complexity*, *complex adaptive systems*, *nonlinear systems*, and DST are often used interchangeably to refer to a class of theories that are focused on the development of complex systems over time (de Bot, 2008). In this study the term DST is used to encompass all of these notions.

DST started as a branch of theoretical mathematics, and its initial aim was to model the development of complex systems (de Bot & Larsen-Freeman, 2011; de Bot, Lowie, & Verspoor, 2007). The mathematical tools that had been developed based on DST proved to be useful for the analysis of a variety of problems, such as the movement of the moon under the influence of the sun, the Earth, and other planets. Because there are systems on every level in the physical world, the tools and approaches of DST have found application in a wide range of fields ranging from epidemiology to economics

to meteorology, and it has been used to solve practical problems ranging from heartbeat control to drilling holes for oil (de Bot & Larsen-Freeman, 2011). What these fields have in common is that they involve phenomena that do not seem to follow predictable patterns of development and change. DST is now considered more as a set of tools and approaches than as a fixed and all-inclusive theory of change (de Bot, 2008).

SLA as a field is concerned with the development of systems such as individual learners, classes, and ethnic groups, and these systems, too, do not seem to follow predictable patterns of change and development. DST originated and benefited greatly from mathematics and physics. Although there are important differences between SLA and mathematics or physics with respect to the object of inquiry, it has been argued that a DST approach to SLA can provide a framework and the instrumentation needed to merge the social and the cognitive aspects of SLA, showing how the interaction of these aspects can lead to development (de Bot et al., 2007). A DST approach can be applied to SLA research to investigate those systems in a structured way and deepen our understanding of language-learning processes (de Bot, 2008). Dynamic systems have several key characteristics and understanding these characteristics enables researchers to look at dynamic systems in a structured way. In the next section, I describe some of these characteristics.

Key Characteristics of Dynamic Systems

The first characteristic of dynamic systems is the nonlinearity of their development and change. In dynamic systems, the relationship between cause and effect is often nonlinear, that is, there is no proportionate effect for a given cause (de Bot et al., 2007). For example, Meara (2004) simulated the attrition of a lexicon using a simple computer model with a network of 2,500 interconnected words. He studied the effects of switching off some words on other words in the network and showed that switching off one word after another did not lead to a similarly linear decline in the number of words in the lexicon; there was enormous variation in how the network reacted.

The second characteristic of dynamic systems is the interconnectedness of their subsystems. When language is regarded as a dynamic system, subsystems such as the lexical system, the phonological system, and the syntactic system are considered to be interconnected, and changes in one subsystem are likely to influence all the other subsystems (de Bot & Larsen-Freeman, 2011).

The third characteristic of dynamic systems is stabilization. A dynamic system sometimes stabilizes itself in an attractor state, which can be defined

as "the state the system prefers to be in over other states at a particular point in time" (Steenbeek, Jansen, & van Geert, 2012, p. 66). It takes a considerable amount of energy to make a system break away from an attractor state and move on. De Bot and Larsen-Freeman (2011) argued that the phenomenon of fossilization (Selinker, 1972), which is the stabilization of an interlanguage system in a form divergent from the target language system, can be described as the result of an attractor state into which the learner's interlanguage system has settled.

Finally, variability is considered the driving force of development. In other words, variability in a subsystem is seen as an intrinsic property of developmental processes (de Bot et al., 2007; van Dijk, Verspoor, & Lowie, 2011). Therefore, unstable periods can be considered a sign of change and development in DST.

Research Methods From a DST Perspective

Several approaches and methods have been proposed for conducting research from a DST perspective. For example, de Bot and Larsen-Freeman (2011) proposed an approach based on a retrodiction, or "explaining after by before" (p. 20). They suggested that after a change has taken place, researchers can look back and describe what happened and study the patterns of development. Furthermore, it is possible to point to tendencies and patterns of development that have been found, trajectories in the system's development can be observed, and the next state can be explained using the preceding one (de Bot & Larsen-Freeman, 2011). However, the tendencies and patterns are not applicable beyond the particular research context in which they have been studied; thus, when it comes to dynamic systems, researchers should think in terms of what is called *particular generalization* (Gaddis, 2002, p. 62) rather than universal generalization. Researchers might acknowledge tendencies or patterns, but they should resist claiming the applicability of their findings beyond specific times and contexts (Larsen-Freeman & Cameron, 2008).

In the present study I adopted a longitudinal, time series approach, which Larsen-Freeman and Cameron (2008) proposed as a method for research from a DST perspective. This approach enables connections to be made across levels and timescales (Larsen-Freeman & Cameron, 2008) because data are collected at multiple measurement points in chronological order. Furthermore, dense data are available because they are collected at many measurement points. Dense data are required to identify variability, which, as noted above, is the driving force of development in DST (de Bot & Larsen-

Freeman, 2011; Larsen-Freeman & Cameron, 2008). The data in this study were analyzed using tools and methods developed by van Geert and van Dijk (2002). Van Geert and van Dijk (2002) presented several new tools and methods for visualizing and describing intravariability in individual time-serial data collected through repeated observations and applied them to data on the early language development of an infant learning Dutch, her first language. Van Dijk et al. (2011) and Verspoor, Lowie, van Geert, van Dijk, and Schmid (2011) applied the tools and methods to data on the L2 development of six native Spanish learners of English who had been in the United States for less than 3 months and provided detailed instructions on how to use the tools and methods. These tools and methods are described in the Methods section.

Complexity, Accuracy, and Fluency

Complexity, accuracy, and fluency have been used to measure language development in speaking and writing in the field of SLA since the 1990s. Brumfit (1984) made a distinction between fluent and accurate L2 usage to examine the development of L2 speaking in classroom contexts. Skehan (1989) added complexity to fluency and accuracy and used the three constructs in the measurement of learners' performance on tasks. Many researchers believe that L2 proficiency consists of multicomponential constructs and that their principal dimensions can be captured by the CAF model (Skehan, 1989). For example, Housen and Kuiken (2009) argued that "CAF emerge as principal epiphenomena of the psycholinguistic mechanism and processes underlying the acquisition, representation, and processing of L2 knowledge" (p. 462). In other words, complexity and accuracy together reflect the current level of language knowledge, whereas fluency represents learners' control over their linguistic L2 knowledge.

There are many definitions for the components of CAF. In this study complexity is defined as "progressively more elaborate language and a greater variety of syntactic patterning" (Foster & Skehan, 1996, p. 303), and accuracy is defined as "the ability to be free from errors while using language to communicate" (Wolfe-Quintero, Inagaki, & Kim, 1998, p. 33). Fluency is defined as "the number of words or structural units a writer is able to include in their writing within a particular period of time" (Wolfe-Quintero et al., 1998, p. 14). The development of English writing is viewed here as a system, and the elements of CAF are viewed as its subsystems that influence language development.

Research on CAF From a DST Perspective

Several studies on the development of CAF have been conducted from a DST perspective to examine language development in ESL and EFL learners. Spoelman and Verspoor (2010) conducted a 3-year longitudinal case study investigating the developmental patterns of accuracy and complexity in a Dutch student learning to write Finnish, a language with a complex morphology comprising 15 cases. The researchers focused on the use of cases and calculated an overall accuracy rate for the use of this grammatical feature. They also examined three types of complexity—of words, noun phrases, and sentences. The results showed that the development of accuracy and complexity was characterized by peaks and regressions and by progress and backsliding, that there was a complex interaction between these variables, and that there were signs of improvement in intraindividual accuracy and complexity. Polat and Kim (2013) examined the developmental patterns of syntactic complexity, lexical complexity, and accuracy in the English speaking skills of a native speaker of Turkish who was an untutored learner. The researchers found that both syntactic complexity and lexical complexity showed high intraindividual variability and that by the end of the 1-year study period, the participant was capable of producing more complex oral output. Accuracy also showed a great deal of variability, but no clear patterns of development were observed during the study period.

Baba and Nitta (2014) investigated, over the course of 1 academic year, the development of fluency in the English writing of two Japanese university students. The students wrote in English in class for 10 minutes and fluency was measured as the total number of words written in the allocated time. The students' fluency increased over the course of the study; however, the changes in fluency were nonlinear: Individual students followed unique developmental paths. The researcher also found that fluctuation periods were observed before and after the developmental trends and there were critical moments at which the two students changed the way they wrote in English. Based on these findings, the researchers argued that "supra" patterns of L2 writing fluency exist and that these patterns contribute to our understanding of the writing development process.

Gaps in the Literature

The three studies reviewed in the previous section were longitudinal CAF studies on writing or speaking conducted from a DST perspective that have advanced our understanding of language development as a nonlinear process. However, previous researchers have not investigated how all three CAF

constructs change over the same period of time or how they interact with one another. Addressing this gap is important because the three constructs need to be examined together to understand their relationships and how they influence one another. Attentional capacity and working memory are limited (Baddeley, 2007; VanPatten, 1990), and if performance in one area (e.g., fluency) requires attention and working memory involvement, committing attentional resources to this area can negatively affect the other areas (Skehan, 1998). Skehan (1998) proposed that tensions exist between complexity and accuracy and between linguistic form (complexity and accuracy) and fluency. This idea led to the formulation of the trade-off hypothesis (Skehan & Foster, 2001), which predicts that committing attention to one area can cause lower performances in others. Skehan also suggested that higher level performances can occur in two out of the three CAF constructs, but not in all three.

Furthermore, interconnectedness is one key characteristic of dynamic systems. When language development is regarded as a dynamic system, and complexity, accuracy, and fluency are regarded as processes and subsystems underlying the acquisition, representation, and processing of L2 knowledge (Housen & Kuiken, 2009), it then becomes important to investigate the interaction between all three CAF constructs.

Finally, Norris and Ortega (2009) argued that the CAF components are developmental in nature, growing and changing constantly, and interacting with one another in often unpredictable ways. Looking at only one or two constructs at a time does not allow the developmental patterns of language acquisition to be investigated adequately. To fill these gaps, the primary purpose of this study is to investigate changes in all three CAF components as well as their interactions. The following research questions were examined in this study:

RQ1. How do complexity, accuracy, and fluency change over the course of 1 academic year in the English writing of two Japanese learners?

RQ2. Are there significant peaks in the learners' CAF development?

RQ3. How do complexity, accuracy, and fluency interact over the course of 1 academic year?

Method

Participants

Two Japanese 1st-year university students participated in this study. The two female participants are referred to as Maki and Kumi (pseudonyms).

They were chosen from a class of 36 students whom I taught, because they studied diligently and completed all the English writing tasks assigned over the course of 1 academic year (10 months). I obtained their oral consent to use their data for research purposes. Their English proficiency was below Level 3 of the General Tests of English Language Proficiency (G-TELP). Level 3 corresponds to a TOEIC score of between 400 and 600 (General Tests of English Language Proficiency, n.d.). Because the participants did not pass this level, their English proficiency corresponds to a TOEIC score of between 220 and 470 (Educational Testing Service, 2012). The creators of TOEIC claim that learners at this level are capable of minimal communication in ordinary conversations. On the placement test conducted in the first semester, Maki's scores on the G-TELP were as follows: grammar 55/100, listening 29/100, vocabulary and reading 29/100. Kumi's test scores were as follows: grammar 59/100, listening 25/100, vocabulary and reading 42/100. According to a questionnaire administered in the first semester, the two students had had almost no experience writing essays in English in secondary school.

Procedures

The participants were students in an English class that met twice a week for 90 minutes each time. The class was focused on English listening, reading, grammar, and vocabulary. Approximately every three classes, the students had 15 minutes to write an impromptu essay either on a familiar topic such as What I did yesterday, My high school, or My winter vacation, or a description of a series of three pictures; thus, the essays were narratives or descriptions. Each student was required to complete 20 essays, of which 15 required writing on a familiar topic and five required describing pictures. The students were told to focus on content and to write as much as possible. They were given a few minutes of planning time before writing and were allowed to use a dictionary while writing. All essays were collected, checked, and returned with comments and corrections. Models of English essays were sometimes shown to the whole class, and English grammar and vocabulary were also explained when the essays were returned. I noticed that the students' motivation for writing essays in English seemed to have decreased at the end of the first semester. Therefore I began to grade their essays from the first assigned essay in the second semester (the 9th essay) in an effort to motivate them to show their maximum English writing ability.

Data Coding

All the essays written by the two participants were typed and coded for the four CAF measures shown in Table 1. Two units of analysis were used in this study: the T-unit and the clause. A T-unit is defined as a main clause plus any subordinate clauses (Hunt, 1965). Bardovi-Harlig (1992) explained that a sentence has two (or more) T-units when independent clauses (with subjects and finite verbs) are conjoined as shown in Example 1, but a single T-unit when one or more clauses are embedded in an independent clause as shown in Example 2.

Example 1. There was a woman next door, and she was a singer. 2 T-units
Example 2. There was a woman next door who was a singer. 1 T-unit

A clause consists of a subject and a finite verb (Hunt, 1965). Finite verbs are verbs that can stand alone in an independent clause. There are two clauses in Examples 1 and 2.

Table 1. Measures of CAF

Category	Measures
Complexity	Clauses per T-unit (C/T)
	Mean length of clause (MLC)
Accuracy	Error-free clause ratio (EFCR)
Fluency	Words per 15 minutes (W/15M)

Measures of complexity used in this study were clauses per T-unit (C/T) and mean length of clauses (MLC). These two measures were used because they reflect different types of complexity: C/T can be used to measure increases in the number of subordinate clauses whereas MLC can be used to measure increases in pre- and postmodification within a phrase through the use of adjectives, adverbs, prepositional phrases, or nonfinite clauses, and the reduction of clauses into phrases, which helps to condense information (Norris & Ortega, 2009). Accuracy was measured as the error-free clause ratio (EFCR). The clause was chosen as a unit of analysis for measuring accuracy because essays from beginners and low-intermediate learners sometimes do not contain any correct T-units. A clause is, therefore, a better unit of analysis to investigate change in accuracy in beginners. For example,

Ishikawa (1995), who researched the English writing of low-proficiency Japanese learners, argued that the clause can tap into beginning-level writing better than the T-unit because it is a smaller unit than the T-unit and it can, therefore, provide a smaller (i.e., more precise) context for examining language development. The measure of fluency used in this study was the total number of words in the 15-minute writing task (W/15M).

Data Analysis

Several techniques developed by van Geert and van Dijk (2002), van Dijk et al. (2011), and Verspoor et al. (2011) were employed to gain insight into the dynamic developmental processes of Japanese students' English writing. First, moving min-max graphs were used to visualize the degree of variability. Van Geert and van Dijk (2002) explained this technique in the following way:

> This technique shows the data using the bandwidth of observed scores. Instead of displaying measurement points as simple dots, the moving min-max graph presents a score range for each measurement occasion. Instead of a single line graph, the data are presented in a bandwidth of scores. This method uses a moving window, a timeframe that moves up one position (measurement occasion) each time (the size of the window, e.g., five consecutive data points, 1 month, etc. is called its period). Each window partly overlaps the preceding windows, using all the same measurement occasions minus the first and plus the next. For instance, for every set of seven consecutive measurements we calculate the maximum and the minimum values. This is done by way of a predetermined moving window, such that we obtain the following series:
>
> max (t1…t7), max (t2…t8), max (t3…t9), etc.
>
> min (t1…t7), min (t2…t8), min (t3…t9), etc. (pp. 353-354)

In moving min-max graphs, the wider the bandwidth is, the higher the degree of variability. Van Geert and van Dijk (2002) argued that one could take windows of a size of about 1/10 of the entire data set, but in principle, no fewer than five data points should be used. The window size in this study was five consecutive data points because the entire data set was 20 data points (20 essays).

The technique described above is exclusively descriptive, so the observations were tested against chance variation by using resampling methods (van Dijk et al., 2011; van Geert & van Dijk, 2002). A Monte Carlo analysis was conducted to test whether the observed peaks defined as the positive distance between minimum and maximum values in the observations could have occurred by chance. In the analysis, the original data were reshuffled 5,000 times (i.e., 5,000 resampling simulations), and the peaks were tested for statistical significance. In other words, I obtained 5,000 positive distances between the minimum and maximum values by reshuffling the original data 5,000 times and calculated the probability that the peaks in the original data could have been produced by chance. The analysis was performed in Microsoft Excel, using *Poptools*, a statistical add-in developed by Hood (2009) that performs random sampling.

Moving correlation coefficients were used to investigate the dynamic patterns of interactions among the CAF measures. Moving correlation coefficients can display change in correlation coefficients. Moving windows of correlation coefficients were created in the same way that moving windows in min-max graphs were created (Verspoor et al., 2011). The window size was five consecutive data points. For instance, a correlation coefficient between EFCR and W/15M was calculated for every set of five consecutive measurements so that the following series was obtained:

correlation coefficient between EFCR and W/15M (1st essay…5th essay),
correlation coefficient between EFCR and W/15M (2nd essay…6th essay),
correlation coefficient between EFCR and W/15M (3rd essay…7th essay),
etc.

In addition to the moving correlation coefficients, Pearson's correlation coefficients were calculated to examine the overall correlations between all the measures and time. Time was a series of 20 data points (20 essays).

Results

The first research question, which asked how CAF changes over the course of one academic year, was answered by calculating Pearson correlation coefficients among the CAF measures and time and by creating min-max graphs. Table 2 shows the Pearson correlation coefficients. Results obtained for each measure are discussed in detail in the following section.

Table 2. Pearson's Correlations Among Time and All the Measures, Means, and Standard Deviations

Measure	1	2	3	4	5	M	SD
Maki							
1.W/15M	-	.32	.54*	-.49*	.61**	89.55	23.37
2.CT	.32	-	-.04	-.24	.26	1.15	0.12
3.MLC	.54*	-.04	-	-.11	.44	6.80	0.96
4.EFCR	-.49*	-.24	-.11	-	.14	0.44	0.21
5.Time	.61**	.26	.44	.14	-	-	-
Kumi							
1.W/15M	-	-.04	-.38	-.19	-.11	120.90	25.78
2.CT	-.04	-	.02	-.04	.45*	1.35	0.20
3.MLC	-.38	.02	-	.23	.38	7.71	1.04
4.EFCR	-.19	-.04	.23	-	.20	0.60	0.14
5.Time	-.11	.45*	.38	.20	-	-	-

Note. W/15M = words per 15 minutes; C/T = clauses per T-unit; EFCR = error-free clause ratio; MLC = mean length of clause.

*$p < .05$. **$p < .01$.

Words per 15 Minutes as a Measure of Fluency

The wide bandwidth in the moving min-max graph for Maki (Figure 1) shows that the variability in fluency was high between the 7th and the 9th essays. After the phase of high variability, a developmental trend towards greater fluency was observed, as she consistently wrote between 80 and 120 words per 15 minutes, after writing between 60 and 100 words per 15 minutes during the earlier phase of the study. The correlation coefficient between time and W/15M (Table 2) was positive and statistically significant ($r = .61$, $p < .01$), indicating that her fluency improved during the course of the study. Kumi's variability (Figure 1) was high during the early phase of the study, but it decreased after the 9th essay. Her W/15M was between 100 and 180 words during the early phase; in the later phase her W/15M decreased slightly to between 90 and 160 words. After the 17th essay her variability decreased slightly again and her W/15M was between 90 and

150 words. Figure 1 shows that no developmental trend toward better fluency was observed for this participant over the course of the study.

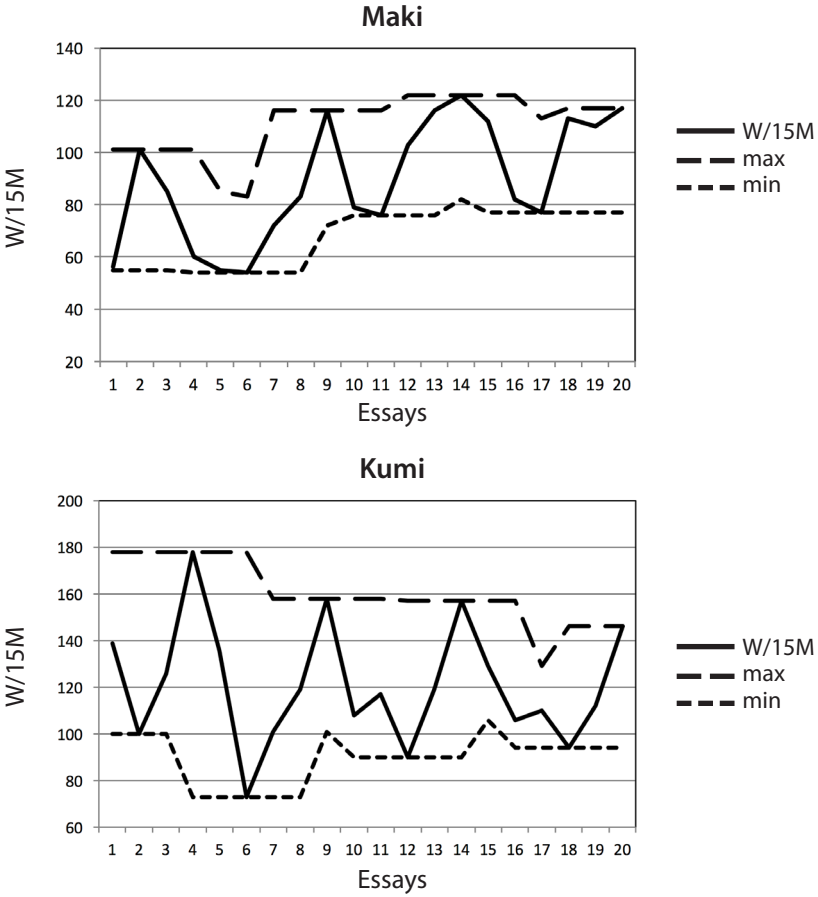

Figure 1. A moving min-max graph (window size of five data points) of the developmental pattern of words per 15 minutes (W/15M). W/15M is a measure of fluency.

Error-Free Clause Ratio as a Measure of Accuracy

Maki's moving min-max graph (Figure 2) shows that the variability in accuracy was high during the early phase and it decreased after the 9th essay. There were almost no error-free clauses in some of her first nine essays with

her EFCR a little above .10. However, after the 12th essay, her EFCR was consistently above .30, so a developmental trend toward better accuracy was observed after the initial phase of high variability. Kumi's EFCR started with low variability and it increased after the 3rd essay. Her EFCR was between .40 and .80 after the phase of low variability. This range was largely preserved toward the end of the study.

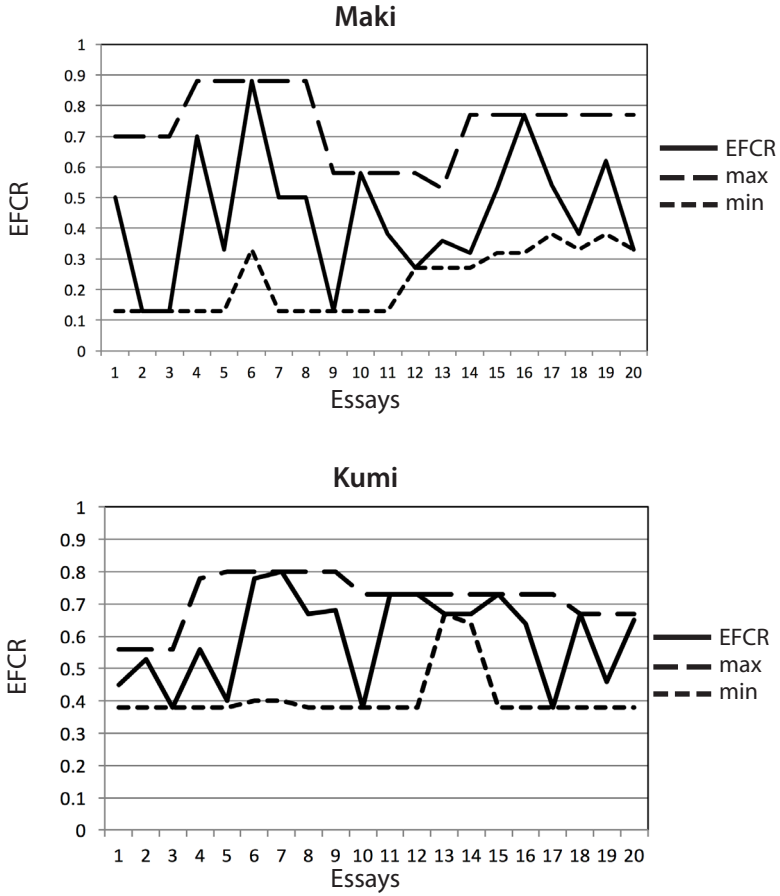

Figure 2. A moving min-max graph (window size of five data points) of the developmental pattern of error-free clause ratio (EFCR). EFCR is a measure of accuracy.

Mean Length of Clause as a Measure of Complexity

Maki's moving min-max graph (Figure 3) shows that the variability in complexity increased after the 9th essay. After the 16th essay, a new phase emerged and her MLC was between six and more than eight words toward the end of the study. Kumi's variability was high between the 1st and the 9th essays and it stabilized toward the end of the study. Her MLC was between six and ten words during the phase of high variability. Because her MLC was consistently above seven words after the phase of high variability, a developmental trend toward more complexity was observed.

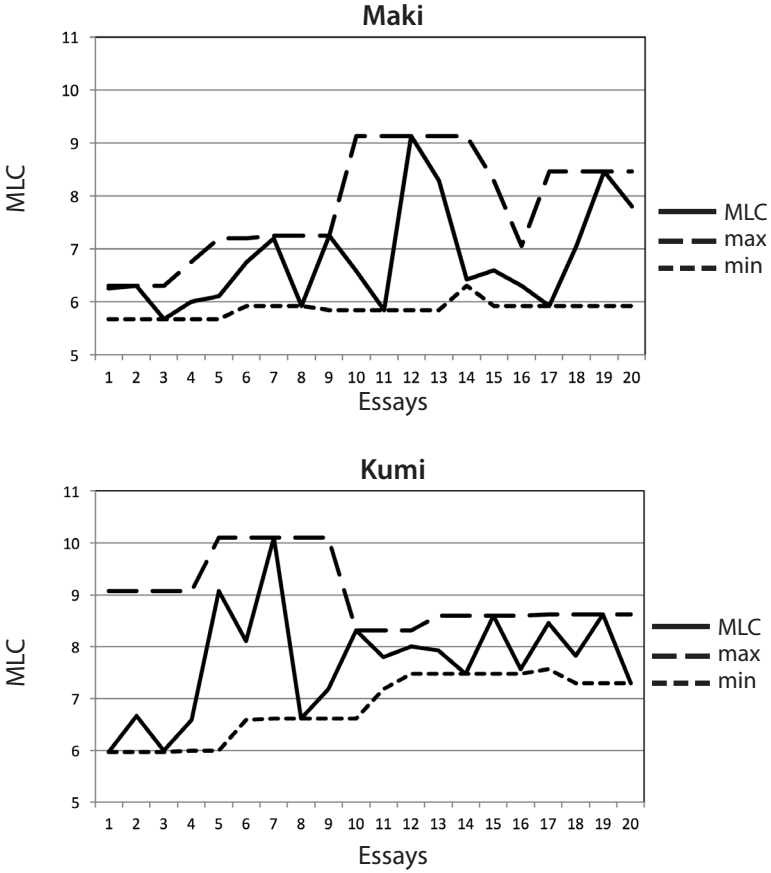

Figure 3. A moving min-max graph (window size of five data points) of the developmental pattern of mean length of clauses (MLC). MLC is a measure of complexity.

Clauses per T-Unit as a Measure of Complexity

Maki's C/T was between 1 and 1.4 over the entire course of the study (Figure 4). In other words, she wrote almost no compound sentences, and her variability was low throughout the study. Kumi's variability in complexity was low during the early phase of the study and her C/T was between 1.1 and 1.6 until the 13th essay. A new phase began after the 16th essay with her C/T ranging between 1.2 and 2 toward the end of the study. The moving min-max graph shows a developmental trend towards greater complexity. Moreover, the correlation between time and C/T (Table 2) was positive and statistically significant (r = .45, p < .05), indicating that her C/T improved during the study. She was writing more and more complex sentences towards the end of the study.

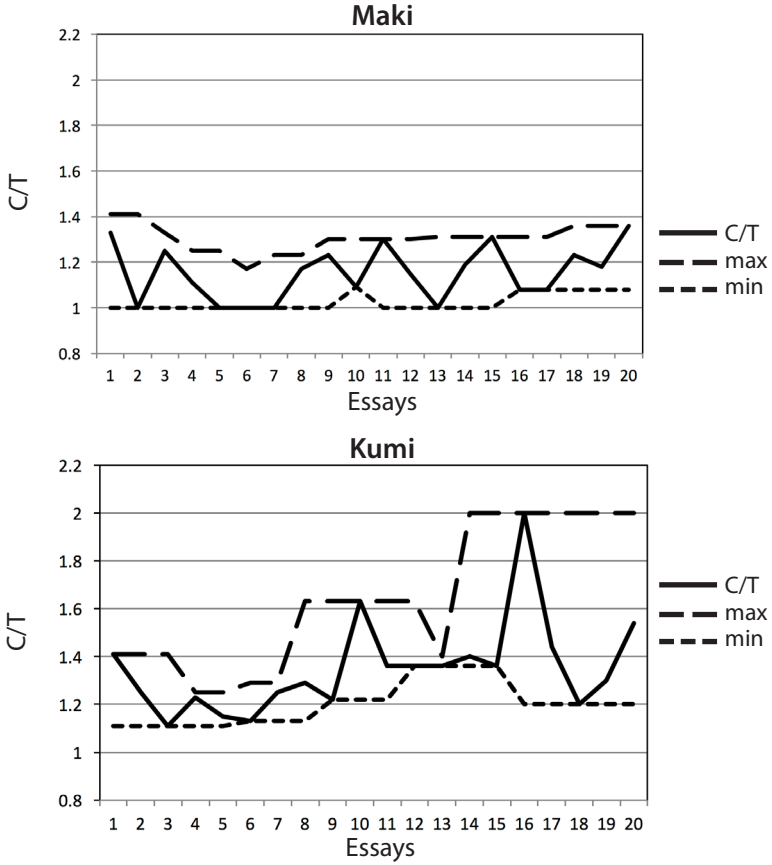

Figure 4. A moving min-max graph (window size of five data points) of the developmental pattern of clauses per T-unit (C/T). C/T is a measure of complexity.

To answer the second research question concerning significant peaks in learners' CAF development, a Monte Carlo analysis was performed for the measures of CAF that showed peaks (Figures 1, 2, and 3) and for statistically nonsignificant correlations with time (Table 2). A Monte Carlo simulation revealed that the peaks in all the measures were statistically nonsignificant (Maki: EFCR $p = .07$, MLC $p = .07$; Kumi: W/15M $p = .47$, MLC $p = .16$). These results indicate that the peaks were likely a result of random fluctuations.

To answer the third research question regarding how complexity, accuracy, and fluency interact over the course of 1 academic year, graphs of moving correlation coefficients between W/15M and the other CAF measures (Figure 5) were created to visualize the changing correlation coefficients. For Maki, the overall correlation coefficient between W/15M and the EFCR (Table 2) was negative and statistically significant ($r = -.49$, $p < .05$), and the moving correlation coefficient remained negative throughout the study (Figure 5). The overall correlation coefficient between W/15M and MLC was positive and statistically significant ($r = .54$, $p < .05$), and the moving correlation coefficient remained largely positive. These results indicate that for Maki, longer essays were associated with longer clauses but more errors. On the other hand, the moving correlation coefficient between W/15M and C/T increased during the early phase of the study, decreased in the later phase, and began to increase again towards the end of the study. These trends show that Maki's longer essays were associated with more complex sentences at first and with simpler sentences later. Toward the end of the study, her longer essays were again associated with more complex sentences. To summarize, fluency and accuracy in Maki's writing appear to have been competing with each other throughout the study, but fluency and MLC interacted positively for most of the study. The correlation between fluency and C/T changed throughout the study. Finally, fluency and both MLC and C/T started to interact positively with one another in the 14th essay and continued until the end of the study.

For Kumi, the overall correlation coefficients between W/15M and the other measures were negative and statistically nonsignificant (Table 2). Figure 5 shows that the moving correlation coefficient between W/15M and MLC was negative for most of the study. This result indicates that longer texts were associated with shorter clauses. The moving correlation coefficients between W/15M and the EFCR and between W/15M and C/T moved in the opposite directions throughout the study. In other words, longer texts were associated with either more complex sentences and more errors or with fewer complex sentences and fewer errors. To summarize, both the

correlations between fluency and accuracy and between fluency and C/T changed throughout the study. However, fluency and MLC appear to have been competing against each other for most of the study.

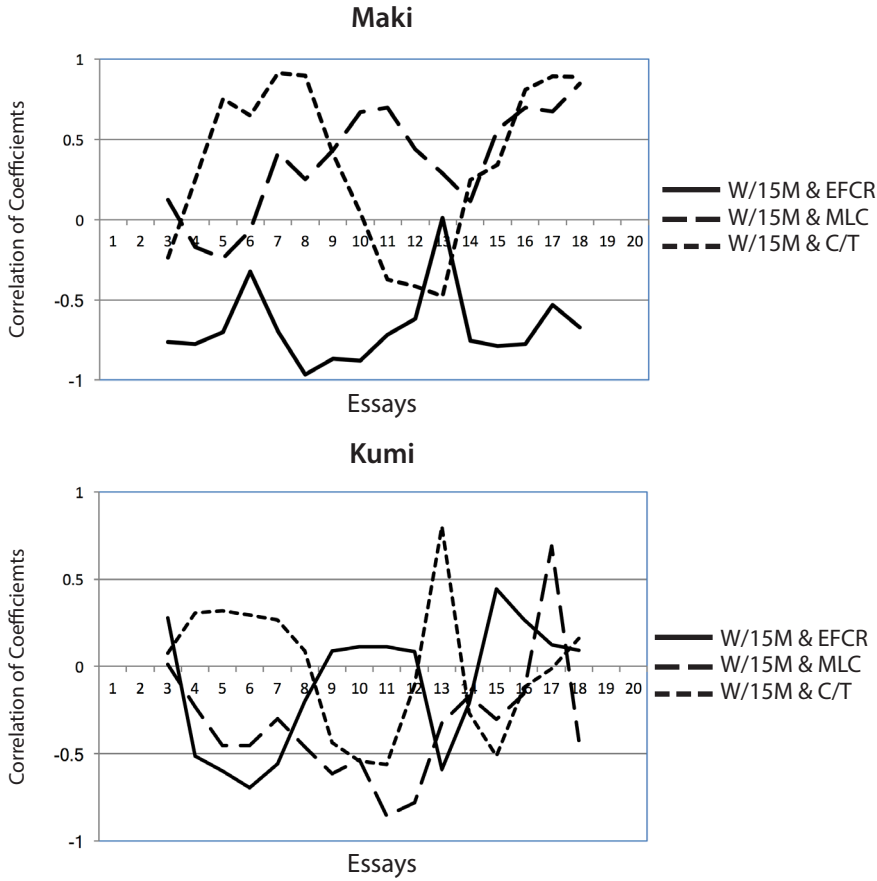

Figure 5. A moving correlation (window of five data points) between W/15M and EFCR, between W/15M and MLC, and between W/15M and C/T. W/15M = words per 15 minutes; EFCR = error-free clause ratio; MLC = mean length of clause; C/T = clauses per T-unit.

Discussion

The primary purpose of this study was to investigate the development of CAF in the English writing of Japanese college students from a DST perspective. One characteristic of dynamic systems is the nonlinearity of development. The findings of this study support this characteristic: The developmental trajectories of all the measures used in this study were nonlinear. For example, although Maki's fluency measure and Kumi's C/T (one of the complexity measures) improved over the course of the study, they never improved linearly. Kumi's fluency measure, which showed no increasing trend over the course of study, also followed a nonlinear trajectory. These results are consistent with those of previous research (Baba & Nitta, 2014; Polat & Kim, 2013; Spoelman & Verspoor, 2010).

Another characteristic of dynamic systems is the interconnectedness of their subsystems. Although only three CAF constructs were investigated in this study, they were found to be interconnected, influencing one another. For example, for most of the study, Maki's fluency and accuracy measures competed against each other, whereas her fluency measure and MLC (one of the complexity measures) positively interacted. In contrast, Kumi's fluency measure and MLC competed against each other for most of the study, and longer texts were associated with negative correlations between the accuracy measure and C/T.

Another characteristic of dynamic systems is stabilization in attractor states. Attractor states were observed in some of the measures used in this study. For example, Maki's MLC in the early essays and Kumi's MLC towards the end of the study remained within steady bandwidths. Maki's C/T throughout the study appeared to remain within steady bandwidths. These observations suggest that the subsystems might have settled into attractor states and stabilized in those states. Maki's MLC might have been pulled out of the attractor state after the 9th essay, as MLC showed variability from the 9th essay to the end of the study.

Furthermore, after the phase of high variability, Maki's fluency and accuracy measures improved, as did Kumi's MLC. Maki's fluency measure showed high variability between the 7th and 9th essays, and a developmental trend was observed after the 12th essay. Maki's accuracy measure showed high variability during the early phase, and then a developmental trend was observed toward the end of the study. Kumi's MLC showed high variability between the 1st and the 9th essays, and then the variability stabilized and a developmental trend was observed toward the end of the study.

These findings support the argument that variability is the driving force of development in dynamic systems (de Bot et al., 2007; van Dijk et al., 2011).

Finally, some changes in CAF were observed after the 9th essay, when I started to grade the participants' essays. A possible explanation for this change is that I told the participants that their essays would be graded based on CAF, and that the grades would influence their final grade in the English class. It is possible that the participants' concern over the final grade made them work harder to obtain a higher grade for the essays.

CAF constructs are rarely positively correlated with one another over the course of the study, a finding that lends support to the trade-off hypothesis (Skehan & Foster, 2001). However, two of the three CAF constructs did sometimes correlate positively. For example, Maki's fluency and complexity measures positively correlated after the 14th essay. Kumi's fluency and accuracy measures positively correlated after the 14th essay. These findings are consistent with Skehan's (1998) suggestion that high-level performance can occur in two out of the three CAF constructs. One exception is the CAF measures for Kumi, which positively correlated around the 17th essay (Figure 5). This finding supports Gunnarsson's (2011) argument that CAF can improve simultaneously when lower level linguistic aspects such as vocabulary, spelling, and grammar are being automatized to improve accuracy and fluency, and enough cognitive capacity is freed up to improve complexity. However, because this trend was observed only toward the end of the study, further research is needed to examine the simultaneous development of CAF.

The results showed how all three CAF constructs changed and interacted with one another over the same period in a structured way. As such, the findings illuminate the developmental patterns of language acquisition of the two participants. Moreover, the findings indicated that examining individual learners' developmental trajectory is pedagogically useful. For example, Maki's C/T remained between 1.0 and 1.4 throughout the study, suggesting that a pedagogical intervention was needed to push her out of the attractor state.

Limitations

This study has several important limitations. First, each writing task used in the study was limited to 15 minutes. As Weigle (2002) argued, however, writing done under timed conditions may not accurately reflect the writers' true ability, because writing ability is not a simple construct; rather, it involves numerous processes such as planning, writing, monitoring, and revising. Although the participants were given a few minutes to plan their responses before writing them, it is not clear whether 15 minutes was sufficient for all

the processes involved in completing a writing task. Second, even though the DST approach highlights the importance of considering a large number of cognitive and social factors in language development, this study focused on only three cognitive variables—complexity, accuracy, and fluency (CAF). The development of English writing was viewed here as a system and the elements of CAF were viewed as its subsystems, which in turn influence language development. However, the development of English writing and its subsystems, CAF, might have been influenced by other subsystems, such as situational, learner, and instructional variables. These variables could include goal orientation, motivation, English proficiency, learning style, physical conditions of the learners, instructional design, and physical environment as well as various political and social conditions. To give just one example, consider the big difference between Maki's and Kumi's outcomes that was observed despite these students' being in the same class with the same teacher and doing the same thing. This difference may be attributable to the differences in their initial English proficiency, which may have influenced the development of CAF. On the placement test conducted in the first semester, their scores were quite different on the vocabulary and reading part. Future research should look at these and other relevant variables that may explain language development.

Conclusion

This study was an investigation of the development of English writing of two Japanese university students over the course of 1 academic year, using the complexity, accuracy, and fluency (CAF) framework. Data were analyzed from a dynamic systems theory (DST) perspective. The findings indicated that the participants' CAF changed as predicted by the key characteristics of dynamic systems and that the three CAF constructs rarely positively correlated with one another over the course of the study. This study underscores the importance of conducting longitudinal studies of individual learners to investigate developmental trajectories and processes of language development. It also supports the use of all three CAF constructs in the investigation of developmental patterns of L2 language acquisition.

Acknowledgments

I am grateful to Katerina Petchko and David Beglar for their invaluable feedback that aided me in writing the manuscript.

Michiyo Hokamura is currently doing doctoral research at Temple University, Japan. Her interest is in the development of second language writing.

References

Abraham, R. (1994). *Chaos, gaia, eros: A chaos pioneer uncovers the three great streams of history*. San Francisco, CA: Harper.

Baba, K., & Nitta, R. (2014). Phase transitions in development of writing fluency from a complex dynamic systems perspective. *Language Learning, 64,* 1-35. https://doi.org/10.1111/lang.12033

Baddeley, A. (2007). *Working memory, thought, and action*. Oxford, England: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198528012.001.0001

Bardovi-Harlig, K. (1992). A second look at T-unit analysis: Reconsidering the sentence. *TESOL Quarterly, 26,* 390-395. https://doi.org/10.2307/3587016

Brumfit, C. (1984). *Communicative methodology in language teaching: The roles of fluency and accuracy*. Cambridge, England: Cambridge University Press.

de Bot, K. (2008). Introduction: Second language development as a dynamic process. *The Modern Language Journal, 92,* 166-178. https://doi.org/10.1111/j.1540-4781.2008.00712.x

de Bot, K., & Larsen-Freeman, D. (2011). Researching second language development from a dynamic systems theory perspective. In M. H. Verspoor, K. de Bot, & W. Lowie (Eds.), *A dynamic approach to second language development* (pp. 5-23). Amsterdam, the Netherlands: John Benjamins. https://doi.org/10.1075/lllt.29.01deb

de Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition, 10,* 7-21. https://doi.org/10.1017/s1366728906002732

Educational Testing Service. (2012). Proficiency scale. Retrieved from http://www.toeic.or.jp/library/toeic_data/toeic/pdf/data/proficiency.pdf

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 18,* 299-323. https://doi.org/10.1017/S0272263100015047

Gaddis, J. L. (2002). *The landscape of history: How historians map the past*. Oxford, England: Oxford University Press.

General Tests of English Language Proficiency. (n.d.). How the G-TELP differs from other tests. Retrieved from http://www.g-telp.jp/english/differs.html

Gunnarsson, C. (2011). The development of complexity, accuracy and fluency in the written production of L2 French. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 247-276). Amsterdam, the Netherlands: John Benjamins.

Hood, G. (2009). Poptools [Computer software]. Canberra, Australia: Pest Animal Control Co-operative Research Centre (CSIRO). Retrieved from http://www.poptools.org

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30,* 461-473. https://doi.org/10.1093/applin/amp048

Hunt, K. W. (1965). *Grammatical structures written at three grade levels*. Champaign, IL: The National Council of Teachers of English.

Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing, 4,* 51-69. https://doi.org/10.1016/1060-3743(95)90023-3

Larsen-Freeman, D., & Cameron, L. (2008). Research methodology on language development from a complex systems perspective. *The Modern Language Journal, 92,* 200-213. https://doi.org/10.1111/j.1540-4781.2008.00714.x

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences, 20,* 130-141. https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2

Meara, P. (2004). Modelling vocabulary loss. *Applied Linguistics, 25,* 137-155. https://doi.org/10.1093/applin/25.2.137

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30,* 555-578. https://doi.org/10.1093/applin/amp044

Polat, B., & Kim, Y. (2013). Dynamics of complexity and accuracy: A longitudinal case study of advanced untutored development. *Applied Linguistics, 35,* 184-207. https://doi.org/10.1093/applin/amt013

Reynolds, C. W. (1987). Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics, 21*(4), 25-34. https://doi.org/10.1145/37402.37406

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching, 10,* 209-231. https://doi.org/10.1515/iral.1972.10.1-4.209

Skehan, P. (1989). *Individual differences in second language learning.* New York, NY: Routledge.

Skehan, P. (1998). *A cognitive approach to language learning*. New York, NY: Oxford University Press.

Skehan, P., & Foster, P. (2001). Cognition and tasks. In R. Peter (Ed.), *Cognition and second language instruction* (pp. 183-205). Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CBO9781139524780.009

Spoelman, M., & Verspoor, M. (2010). Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. *Applied Linguistics, 31,* 532-553. https://doi.org/10.1093/applin/amq001

Steenbeek, H., Jansen, L., & van Geert, P. (2012). Scaffolding dynamics and the emergence of problematic learning trajectories. *Learning and Individual Differences, 22,* 64-75. https://doi.org/10.1016/j.lindif.2011.11.014

Thom, R. (1975). *Structural stability and morphogenesis: An outline of a general theory of models*. Reading, MA: W. A. Benjamin.

van Dijk, M., Verspoor, M., & Lowie, W. (2011). Variability and DST. In M. H. Verspoor, K. de Bot, & W. Lowie (Eds.), *A dynamic approach to second language development methods and techniques* (pp. 55-84). Amsterdam, the Netherlands: John Benjamins. https://doi.org/10.1075/lllt.29.04van

van Geert, P., & van Dijk, M. (2002). Focus on variability: New tools to study intra-individual variability in developmental data. *Infant Behavior & Development, 25*(4), 340-374. https://doi.org/10.1016/S0163-6383(02)00140-6

VanPatten, B. (1990). Attending to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition, 12,* 287-301. https://doi.org/10.1017/S0272263100009177

Verspoor, M. H., Lowie, W., van Geert, P., van Dijk, M., & Schmid, M. S. (2011). How to sections. In M. H. Verspoor, K. de Bot, & W. Lowie (Eds.), *A dynamic approach to second language development methods and techniques* (pp. 129-199). Amsterdam, the Netherlands: John Benjamins. https://doi.org/10.1075/lllt.29.08ver

Vyatkina, N., Hirschmann, H., & Golcher, F. (2015). Syntactic modification at early stages of L2 German writing development: A longitudinal learner corpus study. *Journal of Second Language Writing, 29,* 28-50. https://doi.org/10.1016/j.jslw.2015.06.006

Weigle, S. C. (2002). *Assessing writing*. Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CBO9780511732997

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy and complexity*. Manoa, HI: Second Language Teaching & Curriculum Center, University of Hawaii.

Reviews

LETs and NESTs: Voices, Views and Vignettes. Fiona Copland, Sue Garton, and Steve Mann (Eds). London, England: British Council, 2016. 266 pp.

Reviewed by
George Haikalis
Tennoji Junior High School, attached to Osaka Education University

LETs and NESTs: Voices, Views and Vignettes presents the issues associated with local English teachers (LETs) and native English-speaking teachers (NESTs). NESTs are often teachers who come from "inner circle" countries such as the United States, United Kingdom, Australia, and Canada. As the editors state on page 6, these are most often the type of teachers that employers in the ESL industry have in mind when they advertise teaching positions. Native speakers are often viewed as the authentic embodiment of the standard language (Creese, Blackledge, & Takhi, 2014). The editors also present the issue of native-speakerism (p. 11), which is the common misconception shared by EFL employers around the world that the ideal English teacher is a native speaker. However, a noteworthy addition by Damian J. Rivers in Chapter 4 points out that native-speakerism provides an unfair advantage during recruitment but then a disadvantage after recruitment in the form of limited professional development and policy contribution (pp. 71-73). Moreover, this volume also demonstrates, with firsthand examples, that the situation surrounding LETs and NESTs is changing.

The authors in this edited anthology focus their contributions for the most part on empirical studies in their investigations of both classroom teaching and LET-NEST interactions outside the classroom (e.g., office management issues and socializing). The aim is to provide readers with a contemporary insight into the LET/NEST world in order to uncover current issues. There are several examples of team teaching, LET and ALT (assistant language teacher) interactions, and LET and NEST collaboration. The chap-

ters contain writings from both emerging and well-established researchers who highlight these LET and NEST issues. Chapters 1, 2, and 6 cover team teaching in Japan, Hong Kong, and Korea. Chapters 3 and 5 examine the issue of NESTs and NNESTs (nonnative English-speaking teachers). Chapter 7 gives an overview of EFL and NEST schemes in Europe and Asia. Chapter 8 explores the government initiative in Taiwan to develop English-language skills in the country. Chapter 9 contains cases of team teaching in three different Korean primary schools, which the author uses to provide a better understanding of the interpersonal relationships between LETs and NESTs, focusing especially on issues of differences in power. Chapter 10 has a similar focus but uses research from Vietnamese tertiary institutions. Chapter 11 presents the case of an NNEST and her team-teaching experiences at Japanese primary schools as part of a new government scheme. Chapter 12 examines the factors that foster or inhibit collaboration between local and foreign teachers (including NNESTs) in Hong Kong. In this chapter, Mary Shepard Wong, Icy Lee, and Andy Gao argue that the dominance of NESTs in TESOL is declining due to the increase of NNESTs (p. 211). Chapter 13 considers the value of identity and inequality in the terms *native speaker* and *nonnative speaker*, where author Eljee Javier challenges these distinctions by looking at the role visible ethnic-minority NESTs play in TESOL. The final chapter has a collection of short essays from a number of prominent scholars of native-speakerism who provide recommendations and direction for future research in this field.

The book is accompanied by two separate appendices. One is titled "Investigating NEST Schemes Around the World: Supporting NEST/LET Collaborative Practices" (Copland, Davis, Garton, & Mann, 2016a). In this piece, the authors explore the government programmes in EFL education and observe the interactions between NESTs and LETs. It concludes with recommendations on how to improve these relationships and the quality of the programmes. The second appendix is titled "Materials: Developing Collaborative Practice Between LETs and NESTs" (Copland, Davis, Garton, & Mann, 2016b). This supplement introduces materials to foster the cooperation and understanding between LETs and NESTs and includes ideas and tasks to develop collaboration. For example, there are questionnaires for LETs and NESTs to fill out to see how they manage classrooms or to compare classroom cultures.

Acronyms are used frequently in this book and are a convenient way to introduce the topics in each chapter, for example, the relationship between ALTs or JETs (teachers on the Japan Exchange and Teaching Programme)

and LETs. NNESTs are another topic, although, as the editors suggest, this may carry a negative meaning because of its use of *non*, which can denote a negative or lack of ability (p. 7). Even though NNESTs may be from "outer circle" countries where English has come to play a second language role in a multilingual setting, such as Singapore, Malaysia, Nigeria, India, and the West Indies, this negative bias is said to prevail. However, the editors also mention a paper by Medgyes (1992) who argued that NNESTs are able to be as successful as NESTs because of certain advantages they have, such as the ability to anticipate learner problems and act as a model of a successful learner of English (p. 12). Furthermore, as Adriana González and Enric Llurda explain, there are significantly more people who have learned English as a foreign language than people who have learned it as their first language (p. 91).

The chapters are well researched and many utilize interviews with LETs and NESTs that delve into the relationships between LETs and NESTs and how each are treated in the workplace. Observational data is also used to illustrate these situations. For example, Chapter 6 includes perspectives from native speakers on co-teaching. The chapters focus on government EFL schemes in several different Asian countries, namely Japan, Taiwan, South Korea, and Hong Kong, with even one case in Brunei. Some chapters also focus on cultural biases and provide recommendations on how to cope with them. The final chapter is a particularly valuable resource and an important summary that outlines the issues and gives recommendations and guidance for future research.

I believe the strength of this book lies in its research methodology. NESTs and LETs were interviewed and observed. The interviews and observations were thorough and are supported with other research data. Chapter 11 is especially memorable because it contains an autoethnography of a bicultural and bilingual ALT in Japan through which the author, Chiyuki Yanase, makes a strong case for the use of the L1 in the classroom. Although it is not the focus of this book to address the L1 use issue, it is prominently featured in many of the chapters. The interviews, observations, and other research methods utilized let the reader evaluate the relationships of LETs and NESTs and how they cooperated (or in some cases failed to cooperate) with each other.

I would like to have seen more concern with what role language and culture play in the collaborative efforts between LETs and NESTs. Language and culture are mentioned in several of the chapters in passing, but their role as a potential obstacle in the collaboration between LETs and NESTs warrants a more thorough investigation. Chapter 2, however, contains several

interesting examples of language and culture barriers. For example, in one interview, an LET claimed that because the NEST at her school understood Hong Kong school culture, they had a good working relationship. Meanwhile, another LET lamented over having to exclude the NEST from school activities and meetings because of the language barrier. This issue of language and culture could be further investigated in subsequent research.

Overall, I found this book enlightening. It is somewhat encouraging that my experiences as an ALT in Japan are also fairly common in other Asian countries. However, the reader will see that further work is required if governments are serious about English language education in their countries. This book and the supplements are relevant to scholars studying team teaching between LETs and NESTs and, tangentially, also to those who are interested in the role of the L1 in the classroom. Scholars could use this book as a good starting point for examining a team-teaching issue more closely. Teachers considering a career in ESL education would also benefit from reading this book as a preview to various international teaching contexts. Finally, government and educational administrators would possibly benefit the most because they may not be aware of the situation in their schools or may only be aware of one side of the story.

References

Copland, F., Davis, M., Garton, S., & Mann, S. (2016a). *Investigating NEST schemes around the world: Supporting NEST/LET collaborative practices*. London, England: British Council.

Copland, F., Davis, M., Garton, S., & Mann, S. (2016b). *Materials: Developing collaborative practice between LETs and NESTs*. London, England: British Council.

Creese, A., Blackledge, A., & Takhi, J. K. (2014). The ideal 'native speaker' teacher: Negotiating authenticity and legitimacy in the language classroom. *The Modern Language Journal, 98,* 937-951. https://doi.org/10.1111/modl.12148

Medgyes, P. (1992). Native or non-native: Who's worth more? *ELT Journal, 46,* 340-349. https://doi.org/10.1093/elt/46.4.340

Making and Using Word Lists for Language Learning and Testing. I. S. P. Nation. Philadelphia, PA: John Benjamins, 2016. xiv + 210 pp.

Reviewed by
Caroline Handley
Asia University

In this book, Paul Nation explains how to create word lists from English language corpora, although the same principles could be used to design word lists in other languages. Even though the book is divided into five sections, these are subsumed into two main topics: (a) defining what constitutes a word and (b) constructing a word list from a corpus. Under this division, the book can be understood as consisting of a focus on theory (Chapters 1 to 9) followed by a focus on practice (Chapters 10 to 16). Despite the title, using word lists is the topic of the final chapter only.

As Nation states in the introduction, this book is not an introductory resource regarding the study of vocabulary. It is assumed the reader will already be familiar with vocabulary acquisition in a foreign language and be convinced of the importance of incorporating vocabulary instruction into language courses, areas covered in a previous work (Nation, 2013; see also Schmitt, 2010). Nonetheless, both the overall structure of the book and the format of the individual chapters are extremely well organised, making it easy to read from cover to cover or to use as a reference to find specific information. Reader-friendliness is further increased by the abundance of subheadings and the bullet-point lists of recommendations ending most chapters, which combine a helpful summary with reinforcement of the main points.

In the introduction, Nation not only specifies his aims but also signals his assumptions for the book: That it is a guide primarily for receptive vocabulary knowledge and that it is based on a view of words as single word units with single meanings (excluding homonyms, which have two independent meanings that chanced to converge on the same word form). The reader is not often explicitly reminded of these assumptions later in the book, although they are of paramount importance in assessing the claims made throughout.

After a brief historical overview of word lists, Nation devotes significant space in the first half of the book to describing the theoretical issues in determining what counts as a single word. A naïve reader may think this is an easy matter, but the clear and detailed description of the problems and

choices involved should convince otherwise. In particular, Nation provides a thorough review of the important debate about which unit to use in counting words: type, lemma, or word family.

In his discussion of different senses of the same word form (polysemy), Nation clearly favours treating words as having one core meaning extended into multiple related senses, rather than as having multiple meanings that require separation. In doing so, he possibly underestimates the comprehension problems this can cause foreign language learners, particularly as he ignores metaphorical extensions. As an example, this assumes a learner who knows the noun *tail* can successfully decode the verb *tail* in context. Although he notes that different senses of a word are often translated into distinct words in other languages, he declares that learning a foreign language is about learning "how experience is classified differently in different languages" (p. 52), without acknowledging that this is probably one of the most difficult aspects of foreign language learning (Hendrickson, Kachergis, Fausey, & Goldstone, 2012; see also Pavlenko, 2014, Chapters 2 & 8 in particular).

In fairness to Nation, his reasoning in electing to minimise the issue of polysemy when creating word lists probably has a practical basis. As he states, the computer programmes currently available for extracting word frequency information from corpora are not very good at distinguishing homonyms or proper nouns from common nouns with the same form or at dealing with orthographical errors and alternative spellings (including hyphenated versus unhyphenated compounds). Thus, apart from increasing the many challenges of defining what constitutes a word, admitting polysemy would cause substantial, and at least for the time being, insurmountable problems for building a computer programme that can recognise such distinctions.

In the second half of the book, Nation describes the practicalities of choosing a corpus and constructing word lists from it. He highlights the considerations involved in deciding which corpus to use to create a word list, especially the corpus design and the purpose of the list, as well as the importance of calculating both word frequency and range to determine usefulness for learners. However, he does not assume the reader necessarily intends to make a word list; instead he describes and compares the most well-known and commonly used lists that have been compiled, including specialized lists. He provides an extended explanation and critique of his own BNC/COCA headword lists containing only the headwords of the 25,000 word families (available to download from his webpage: http://www.victoria.

ac.nz/lals/about/staff/paul-nation). In his discussions, he also reminds the reader of the subjective decisions and human input that go into making a word list—not everything is best decided by a computer programme.

Before the brief summary of how word lists can be used in teaching and testing that concludes this book, Nation includes a chapter written by Thi Ngoc Yen Dang and Stuart Webb. This is a research report on creating an essential word list for beginners. Although it differs in style from the rest of the book, it does not feel out of place but rather provides an interesting case study exemplifying the preceding chapters. It also introduces a newer word list that could be of particular use to teachers of beginner- to intermediate-level students (CEFR A1 to B1). In sum, the chapters making up the practice-oriented half of this book are testimony to the years Nation has spent working in this field and the achievements and contributions he and his peers have made.

Nation provides a succinct and accessible description of making word lists for English language teaching, a task that he makes appear easy due to his extensive research and expertise in this area. His humility as a scholar is also refreshingly apparent in his full acknowledgement of others' contributions to the field and his own past errors, his unpretentious writing style, and the accreditation of his PhD students as coauthors whose research informed his ideas in several chapters. For any educator who wishes to learn about using corpora and making word lists, this book is both an invaluable guide and a pleasurable read.

References

Hendrickson, A. T., Kachergis, G., Fausey, C. M., & Goldstone, R. L. (2012). Re-learning labeled categories reveals structured representations. In *Proceedings of the Thirty-Fourth Annual Conference of the Cognitive Science Society* (pp. 1668-1673). Sapporo, Japan: Cognitive Science Society. Retrieved from: http://www.indiana.edu/~pcl/papers/relearningcategories.pdf

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge, England: Cambridge University Press.

Pavlenko, A. (2014). *The bilingual mind: And what it tells us about language and thought*. Cambridge, England: Cambridge University Press [Kindle SDK version]. https://doi.org/10.1017/CBO9781139021456

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, England: Palgrave Macmillan. https://doi.org/10.1057/9780230293977

Exploring Psychology in Language Learning and Teaching. Marion Williams, Sarah Mercer, and Stephen Ryan. Oxford, England: Oxford University Press, 2015. ix + 171pp.

Reviewed by
Caroline Hutchinson
Juntendo University

> *It's the end of the summer semester, post-examination. The stress has taken a toll on students, and in some cases on their personal relationships. It's hot, and the sounds of construction are painfully loud through the open windows of the classroom. You've been looking forward to these final classes as a chance to unwind together and look to the future, but the lesson is falling flat. You're probably wondering what's going wrong, and what you can do about it.*

This is the scenario that opens Chapter 1 of *Exploring Psychology in Language Learning and Teaching*. It is a story heard in staff rooms the world over, and it illustrates the power individual and group psychology can have to make or break a lesson. Like others in the Oxford Handbooks for Language Teachers series, this book is designed to bridge theory and practice, giving practical classroom examples to contextualize psychological theory along with reflection activities aimed at deepening understanding of student and teacher beliefs.

In Chapter 1, the authors trace major developments in the field of psychology and how they have affected pedagogical beliefs and practice. Behaviourist approaches, largely based on research conducted on animals, saw learning in terms of response to stimuli; these led to the adoption of the audiolingual method in language teaching, with its emphasis on repetition or substitution. Cognitive psychology shed greater light on mental processes, emphasizing strategies and personalization in language learning. In the 1960s, humanistic scholars stressed the role played by learners' emotions and beliefs. Sociocultural theorists considered the force exerted on the individual by other people, institutions, and cultural beliefs and highlighted the benefits of supportive collaboration with the teacher and other learners. The chapter ends with a consideration of emerging fields in psychology, which stress the dynamic and complex relationships between individuals and their contexts.

In Chapter 2, the authors turn their attention to the role of groups and contexts and how teachers can bring out the best in their classes. The authors caution against overgeneralization when considering cultural values, stressing instead the "nested systems" (Davis & Sumara, 2006), such as nation, community, class at school, and family, that make up individual identity. They urge sensitivity to our own cultural lens as well as that of students and the promotion of open dialogue about students' own perspectives. They also discuss the locus of power within the classroom, considering different leadership styles adopted by teachers and the social nature of learner autonomy.

The topic of Chapter 3 is the self and what teachers can do to encourage a "healthy but realistic" (p. 58) sense of self in learners. Identity is complex, and some more global parts of the self may be resistant to change. The authors stress, however, that teachers can foster positive self-efficacy by teaching students strategies, challenging negative beliefs, and encouraging the belief that students can succeed through their own efforts.

In Chapter 4, the authors consider different types of beliefs: what constitutes knowledge, implicit beliefs about how able we are to develop our skills (what Dweck [2006] calls a "growth mindset," as opposed to a "fixed mindset"), and how we explain success and failure in our learning. They also stress that teachers should be aware of their own beliefs and the influence these can have on our methodology and feedback.

The focus in Chapter 5 is on anxiety and affect. The writers provide insights from neuroscience about how our brains evaluate events and argue that emotions can override other cognitive processes in ways that are difficult for teachers to predict. They then suggest techniques that teachers might use to encourage positive emotional outlooks in students, for example, through helping learners to work to their strengths or through creating opportunities to engage with meaningful issues beyond the classroom.

In Chapter 6, the writers explore theories of motivation and the unpredictable ways in which individual motivation may change over time. Williams, Mercer, and Ryan emphasize the importance of understanding learners' goals and frustrations without attempting to force the learners in directions contrary to where they want to go. As in previous chapters, the authors also turn the lens on teachers, arguing that when we feel competent, autonomous, and comfortable in our teaching context, we can best convey this enthusiasm to our students.

In Chapter 7, the focus is on the actions students take while learning, in particular self-regulation—the ways students manage their own learning. One way in which students can consciously control their learning is by using

strategies, and the authors provide a number of real student examples to illustrate how personal and creative these can be. They argue that awareness of strategies and a metacognitive understanding of the learning process are vital to successful language learning and have a major impact on students' sense of competence and agency in learning.

The concluding chapter brings together the aspects of psychology considered in each of the chapters of the book and introduces the concept of "L2 willingness to communicate" (MacIntyre, Dörnyei, Clément, & Noels, 1998) to illustrate how all the concepts presented in the book can influence communication behavior. Williams, Mercer, and Ryan argue that as teachers, we should neither oversimplify nor feel overwhelmed by the burden of understanding our students and their needs. Instead, we should provide opportunities for learners to communicate who they are, to make choices, and to become aware of themselves as learners, while promoting a positive group dynamic and reducing anxiety in the classroom.

In *Exploring Psychology in Language Learning and Teaching*, the authors do not claim to offer definitive answers to our classroom dilemmas; rather, they argue persuasively that the interactions between aspects of individual psychology and between individuals as members of groups are complex and unpredictable. In that sense there are no definitive answers. Our job as teachers is to educate ourselves, to reflect on experience and theory, and to remain open to what learners themselves have to say. This book is an excellent starting point for teachers at any level, offering concise explanations and realistic suggestions supported by relevant examples and reflection activities throughout. Practitioners hoping to do further research will also benefit from its extensive citations and references, and for teacher trainers, the online resources on the Oxford University Press companion website provide links to related discussion questions and supplementary activities.

References

Davis, B., & Sumara, D. (2006). *Complexity and education: Inquiries into learning, teaching, and research*. New York, NY: Routledge.

Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York, NY: Random House.

MacIntyre, P. D., Dörnyei, Z., Clément, R., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a second language: A situational model of second language confidence and affiliation. *The Modern Language Journal, 82,* 545-562. https://doi.org/10.1111/j.1540-4781.1998.tb05543.x

Complexity in Classroom Foreign Language Learning Motivation: A Practitioner's Perspective from Japan. Richard J. Sampson. Bristol, England: Multilingual Matters, 2016. xi + 227 pp.

Reviewed by
Christopher Pirotto
Fukui University of Technology

Second language learning motivation has been a topic widely investigated and discussed by researchers and teachers alike. Richard Sampson's previous research into Japanese EFL classroom motivation and student attitudes (see Sampson 2010, 2012) inspired him to conduct the yearlong research project that resulted in *Complexity in Classroom Foreign Language Learning Motivation*. Sampson conducted five action research cycles and used his own journal entries, student journals, and survey data to address the main aims of his book, which are "(i) to gain a clearer picture of the ways in which self-concept affects language learner motivation; (ii) to explore the ways in which teacher-instigated change action might affect students' motivation; and (iii) to generate a more complex, holistic understanding of dynamic motivation in the class group" (p. 5). According to Sampson, the narrative-based approach of how he came to understand language motivation in his classroom adds to previous research like that of Miyahara (2015) that used a narrative approach to tell the L2 identity stories of several students.

In Chapter 1, the author explains what inspired him to conduct the research, states the aims of the book, briefly introduces the reader to complex systems theory, and provides the layout of the chapters that follow. After the introduction, the book is split into three parts. The layout used by the author allows readers to easily choose the sections that appeal to them.

Part 1, "成長 – *seichou* – growth," contains four chapters. In Chapter 2, "Groundings from Foreign Language Learning Motivation Research in Japan," the author provides a detailed narrative of the motivation literature unique to the Japanese EFL context. In Chapter 3, "A Move to Socio-dynamic Motivation," the literature surrounding Dörnyei's (2009) L2 Motivational Self System is explored, with emphasis on possible-self theory. By stating that current literature regarding EFL students in Japan fails to "consider in any depth the interplay between these learner-internal and -external influences" (p. 35), the author identifies a hole in the literature that he hopes to

fill. The author adds that the research on this topic has been mostly quantitative and does "not describe adequately the subjective understandings of the dynamic processes involved in language learning motivation" (p. 35). Chapter 4, "Research Design," begins with a philosophical discussion of and introduction to action research and complex systems theory. The chapter concludes with the study design and data collection methods. Chapter 5, "Action Research Narrative," provides a chronological narrative of the five classroom-based action research cycles implemented over the academic year. This chapter would be of specific interest to language teachers as the narrative details how students responded to specific lessons and activities in the classroom.

Part 2, "再見 – *saiken* – re-view," begins with Chapter 6, "Revisiting Complex Systems Theory," which introduces the reader to complex systems theory, its terms, and how it can be used to help understand classroom language-learning motivation. In Chapter 7, "Class Group as an Open System," Sampson discusses how systems outside the classroom can affect the motivation of learners inside the classroom and also ways that learners bring themselves into the classroom. Some specific examples discussed include students' past experiences in the language classroom, changing relationships among students, and students' hobbies outside the classroom. In Chapter 8, "Co-adaptation Between Self and Environment," the author continues to draw on complex systems theory to discuss three nested motivational states (English use, interaction, and realisations) that he found to be coadaptive with parts of the L2 Motivational Self System. Using complex systems theory to analyze the interaction between nested motivational states and the L2 Motivational Self System provides a "realistic vision of the interaction between self and environment in classroom language learning" (p. 123). In Chapter 9, "Motivational Phase-shifts and Self-organisation Across the Class Group," Sampson analyzes how certain transitions in the classroom, such as new types of lessons and class groups, can bring about motivational changes not only in individuals but also within the class system as a whole. In the final chapter of Part 2, "Novel Motivational Emergence in the Class Group," the author continues to draw on complex systems theory when he seeks to understand the final motivational emergence of his class by analyzing interactions in the class over the final 6 weeks.

Part 3, "相互 – *sougo* – reciprocity," consists of two chapters. In Chapter 11, "The Landscape of Classroom Motivation," the author draws on his findings to suggest a way to better theorize motivation in the language classroom. His theory is founded on self-ideas, perceptions of experiences, and the class

environment. In the final chapter, "Conclusion and Iteration," the author discusses possible pedagogical implications for the classroom along with the methodological strengths and weaknesses of his research that can inform future studies into second language learning motivation.

This book is not a step-by-step guide to increase L2 motivation in the classroom nor is it a manual of classroom activities and lessons to increase L2 motivation. However, the appendix includes the activities that were assigned to the students and the author discusses how certain lessons affected the L2 motivation of students. These resources may be of particular interest to those educators who have similar teaching contexts.

Although the book reads as if the author intended the target audience to be academics with an advanced knowledge of complex systems theory, I believe that the way the research was carried out makes it valuable for classroom teachers and action researchers. One of the greatest strengths of this book is the way the author uses journal entries from everyone involved in the learning process to provide an in-depth look into how specific activities, events, perceptions, and experiences have an impact on L2 learning motivation. Although no two learners or groups of students are the same, educators with students of a similar age will find that the journal entries provide valuable insight into what their own students may be experiencing or thinking.

The author's use of complex systems theory as a way to help understand the motivation of his class over a yearlong course is a valuable addition to the foreign language learning motivation literature. I recommend this book to anyone who wishes to learn more about classroom language learning motivation, especially in the Japanese context.

References

Dörnyei, Z. (2009). The L2 Motivational Self System. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity, and the L2 self* (pp. 9-42). Bristol, England: Multilingual Matters.

Miyahara, M. (2015). *Emerging self-identities and emotion in foreign language learning: A narrative-oriented approach*. Bristol, England: Multilingual Matters.

Sampson, R. J. (2010). Student-negotiated lesson style. *RELC Journal, 41,* 283-299. https://doi.org/10.1177/0033688210380555

Sampson, R. J. (2012). The language-learning self, self-enhancement activities, and self-perceptual change. *Language Teaching Research, 6,* 313-331. https://doi.org/10.1177/1362168812436898

Educating Second Language Teachers. Donald Freeman. Oxford, England: Oxford University Press, 2016. xxi + 290 pp.

Reviewed by
Joseph Poulshock
Senshu University

In this significant book, Donald Freeman sets forth his extensive knowledge and expertise regarding the challenge of educating second language teachers. In four major parts, Freeman (1) describes the ways language teacher education differs from other kinds of teacher education; (2) discusses three major views about how people learn language teaching; (3) examines the mental processes of thinking, knowing, and reflecting about teaching; and (4) introduces a design theory for teacher education.

Design theory stands as the main feature of the book, and with it, Freeman provides a conceptual framework for thinking about language teacher education. He sets this framework within sociocultural theory, where student teachers work through social processes to make sense of the culture of teaching. However, readers will not find practical tools about methods, techniques, and activities for educating teachers. Rather, they will find conceptual tools for making sense of teacher education.

In *Educating Second Language Teachers*, Freeman advocates a descriptive or nonprescriptive approach to language teacher education. That is, he does not prescribe what ought to be. Though some educators may disagree with this descriptive approach, Freeman argues that as teacher educators, we do not prescribe what teachers should study or know. When we prescribe, we limit what is valuable and effective; we esteem one way of teaching over others. Instead, Freeman describes a social process in which teacher educators provide their students with "tools and opportunities to use them" (p. 229).

It must be emphasized that Freeman does not speak of practical or material tools. Rather, he opens a box of conceptual tools—metatools—which he also calls *social facts*. For example, some teachers may employ a technique called presentation, practice, and production (PPP). When teachers use PPP in a lesson, it is just a practical tool, but PPP becomes a conceptual tool or social fact when teachers explain it, use language about it, create verbal plans to use it, or discuss and evaluate its use. Thus, teachers use these conceptual tools as social facts to make sense of teaching.

From Freeman's perspective, teachers call on these social facts in community, which expresses itself in two ways: through activity and through explanation. In the *community of activity*, teachers do all the visible actions connected with teaching, and when doing practice teaching, student teachers act as if they are teachers as they rehearse for the future. In the *community of explanation*, teachers and student teachers think, talk, and write about their actions, with the goal of making their actions intelligible and meaningful.

Freeman suggests a nuance to explanation that he calls *articulation*. As students practice teaching, they act as if they belong to the community, but they often explain social facts in fuzzy and imprecise ways. For Freeman, this fuzziness is not explanation, but articulation. When student teachers begin to explain social facts well, their articulations become explanations. Freeman says that student teachers tend to explain social facts imprecisely, but as they grow, they explain them with more accuracy. By giving feedback, teacher educators can help student teachers learn to explain better, thus enabling them to fully enter the teaching community.

In the first three parts of the book, Freeman takes us on a grand and masterful tour of second language education, covering abstract and philosophical sociocultural theory and significant historical developments in the field. Teacher educators at universities will find this helpful, especially if they value educational theory and philosophy, but short-term trainers and workshop leaders may find these ideas less useful. In the culminating fourth part of the book, which may be the most beneficial section, Freeman introduces his design theory for teacher education. He says that "the central process in the design theory is how articulation becomes explanation" (p. 261).

Each part of the book gives a wealth of information, but there is one major problem. Though Freeman demonstrates a masterful understanding of the philosophy, history, and practice of second language teacher education, he often writes in academese, a style that is "a betrayal to our calling to enhance the spread of knowledge" (Pinker, 2014b, para. 61). These are strong words, but maybe we can blame the abstract nature of the subject for sowing the confusion. Even Freeman admits that social theory is wordy and complex. As a result, readers for whom English is a second and foreign language may get lost in the book's thick abstraction. Simply put, the book reads harder than necessary. For example, in a typical passage, we can see the problem of long nominalizations, which Sword (2016) dubbed as zombie nouns because they suck the action and vitality out of writing:

> Pluralization refers to the ways in which the *unitary view of language teaching knowledge* was both broadened and specialized with these different roles in diverse work environments. *Diversification of uses of teaching knowledge* drove pluralization in how that knowledge was defined and articulated. (p. 194, italics added)

Sword (2016), Williams (2009), and Pinker (2014a) all told us how to fix this problem. As a minimal first step, we make subjects actors, and we put actions in verbs (not hiding actions in nominalizations). However, Freeman often presents actorless sentences of 20-25 words with zombie-noun subjects, which readers must hold in working memory until the verb appears at sentence end. Many readers will need to reread these complex sentences because they place a heavy tax on mental computational resources (see Gibson, 1998).

Some readers may respond, "Book reviews should focus on ideas, not writing style." However, on the back flap, the eminent author Jack Richards endorses Freeman's book as an "invaluable resource for teachers, teacher educators, and researchers." In essence, this is true, yet it may not be good buying advice for some would-be readers because many of the book's most salient points are often buried in obscure layers of academese. Researchers, teacher educators, and teachers may be able to dig up the treasures in this book, and it would be worth the effort, but some of them may find the digging rather hard, especially those who do not have a sufficient level of English or patience.

Thus, despite the masterful account of second language education in an important theoretical work that targets all kinds of language teaching professionals, this book may be best suited for learned scholars who can decipher academese. The rest of us will have to wait for the succinct, stylish, and more straightforward version.

References

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68,* 1-76. https://doi.org/10.1016/S0010-0277(98)00034-1

Pinker, S. (2014a). *The sense of style: The thinking person's guide to writing in the 21st century*. New York, NY: Viking Adult.

Pinker, S. (2014b). Why academics stink at writing. *The Chronicle of Higher Education*. Retrieved from https://stevenpinker.com/files/pinker/files/why_academics_stink_at_writing.pdf

Sword, H. (2016). *The writer's diet: A guide to fit prose* [Kindle version]. Chicago, IL: University of Chicago Press. https://doi.org/10.7208/chicago/9780226352039.001.0001

Williams, J. M. (2009). *Style: The basics of clarity and grace*. New York, NY: Pearson Longman.

Linguanomics: What Is the Market Potential of Multilingualism? Gabrielle Hogan-Brun. London, England: Bloomsbury, 2017. xiii + 168 pp.

Reviewed by
Dávid Smid
Meisei University and Eötvös Loránd University

The title of Gabrielle Hogan-Brun's book starts with a newly coined portmanteau: an innovative example of linguistic blending whereby words from two distinct languages are merged—the Latin word for *language* and the English word *economics*—to denote a recently emerging interdisciplinary field dedicated to examining "the relationship between economic and linguistic variables" (p. 153). Given the linguistic makeup of *linguanomics*, it may come as no surprise that the focus of the book is on multilingualism from the perspective of economics. In the current era of globalization, contact between people of different cultural and linguistic backgrounds has become increasingly common. Under such circumstances, questioning the market value of languages other than English—the language most commonly associated with current globalization practices (Dewey, 2007)—is pertinent. Thus, Hogan-Brun's venture—as she herself is multilingual (see Assimil, 2017)—seems to be timely and highly relevant for everyone affected by the presence of multilingualism, be it through a workplace, organization, or country.

As the author explains in the Afterword, the purpose behind writing the book was to draw readers' attention to how the operation of societies, at both the organizational and the individual levels, is interwoven with lan-

guages and what the economic consequences are. The book is intended to be practical, that is, the emphasis is on the application of theory into practice. The practice-oriented approach is achieved via two aspects. First, we are given numerous illustrative examples of the market value multilingualism can have from a wide range of contexts. Second, we are invited to uncover our own attitudes and express our opinions about the economic aspects of languages as the writer raises several questions but remains objective when—or if—she answers them. Therefore, the reader should not expect clear-cut answers to questions, such as what value is associated with a particular language or what foreign language one should choose to learn for securing a high-paying job in the future, as these questions are context dependent. Rather, the book should be seen as "a source of ideas" (p. 130) to help broaden one's perspectives on the globalized world as well as prompt further thinking or even research about one's context and its characteristics.

The contents of the book are organized in a logical and reader-friendly manner, and the author—despite the complexity of the two fields in question (i.e., languages and economics)—manages to maintain a clear, easily comprehensible style, all of which make the reading a joyful and engaging experience. The book starts with a short lead-in section, in which the reader is given a few glimpses of how the relationship between languages and economics correlates with our everyday lives. Here, the two fundamental ideas of the book are also formulated: "languages have market value" and "language skills should be seen as assets" (p. xiii). The rest of the work is divided into five chapters. The aforementioned ideas are studied from various angles, which means that other fields, such as history, marketing, business, or education, are also reflected in the discussion. A reader accustomed to more academic pieces of writing may be surprised to notice that no references are cited within the main text. One can gain access to the sources used only by consulting a separate References section at the end of the book. In a similar vein, in the absence of footnotes or end-of-chapter notes, one is left with an extensive Notes section after the main text, which can detract from the reading experience. However, both of these shortcomings can be justified given the practical approach of the book. Finally, the inclusion of a short Glossary and an Index must be praised for further enhancing the usefulness of the work.

Quite aptly, the first of the five chapters begins with a historical overview of cross-cultural trading, highlighting the fact that achieving success in business between different nations has always been contingent on the effective use of languages and communication skills, although what counts as effective has been subject to change over time and space. In the second chapter, the

focus is shifted to the present era and the economic consequences of making choices about the use of languages at the governmental and organizational levels. Recurring countries and organizational bases in the discussion include Canada, Switzerland, and the European Union, but a broad range of settings are mentioned (e.g., Peru, Kenya, Japan, and New Zealand). As throughout the book, here the author also aims to provide a comprehensive account, and we are reminded that language policies are often established at the expense of minor languages. Moving on with the issues of language diversity and decision making, the next chapter is dedicated to how the presence of multiple languages is managed in a wide range of contexts (e.g., telemedicine, aviation, and maritime industries) for the sake of minimizing the occurrence of miscommunication. There is considerable examination of how multilingual skills are promoted within the education sector (ranging from primary to tertiary levels), and thus, how countries—from the Americas to Europe and Africa to Asia and Australia—seek to invest in their citizens through language education. In the fourth chapter, the author assesses the benefits of learning additional languages, pinpointing the growing need for multilingual speakers in the global economy. The economic prospects of people competent in multiple languages are further explored in the final chapter of the book: Hogan-Brun envisages the diversification of the world economy in terms of *linguistic commodities*, that is, languages with a superior value in the global market, as a possible scenario for the future.

Although it is clear that the aim of the author was to provide a practical account by putting some theoretical concepts into practice, it might have been reasonable to include an introductory chapter on the field of language economics (for an overview, see Grin, 2003). This would have afforded a better opportunity for the reader to follow the discussion in a well-oriented manner. A more thorough introduction to some linguistics- and economics-related theoretical concepts—such as globalization, linguistic diversity, language commodification, or multilingual competence—could also have been included because they form the basis for the discussion. The brief glossary will probably not satisfy the needs of those readers who are looking for a more scholarly handbook on language economics. On a similar note, more enthusiastic readers might wish for the expansion of some parts of the book: Issues raised concerning education or technology are areas that would benefit from more attention, considering their crucial links to the topic in question.

All things considered, Hogan-Brun's book *Linguanomics: What is the Market Potential of Multilingualism?* can be regarded as an impactful and

practical introduction to language economics considered from the perspective of multilingualism. After finishing the book, the reader is equipped with a new lens with which they are able to see the market value of languages in a global economy-driven world more clearly. Another merit of the book is the author's objectivity, by which the reader is invited to participate in the discussion. Given the relevance of the book's topic and the approach adopted, it is recommended not just to those with basic interests in the fields of economics or languages, but to everyone who seeks some new perspectives related to the current, interconnected state of the world.

References

Assimil. (2017, July 27). Linguanomics: Interview with Gabrielle Hogan-Brun [Blog post]. Retrieved from http://fr.assimil.com/blog/linguanomics-interview-with-gabrielle-hogan-brun-2/

Dewey, M. (2007). English as a lingua franca and globalization: An interconnected perspective. *International Journal of Applied Linguistics, 17,* 332-354. https://dx.doi.org/10.1111/j.1473-4192.2007.00177.x

Grin, F. (2003). Language planning and economics. *Current Issues in Language Planning, 4,* 1-66. https://dx.doi.org/10.1080/14664200308668048

Reflective Practice as Professional Development: Experiences of Teachers of English in Japan. Atsuko Watanabe. Bristol, England: Multilingual Matters, 2016. ix + 190 pp.

Reviewed by
Bill Snyder
Kanda University of International Studies

Although reflective practice (RP) has become a standard component of many teacher education and professional development programs, it has also come under critique for lacking definition and methodological rigor. Mann and Walsh (2013) argued that for RP to fulfill its potential in professional education, it needs to be reformed as a more rigorously data-driven process. These changes in practice should include a greater focus on spoken, collaborative processes of reflection and the use of data from a wider range of contexts. It is only through such refocusing of RP that it can become an

evidence-based process that informs practitioners beyond the individual case. The publication of Atsuko Watanabe's *Reflective Practice as Professional Development* makes an important contribution to meeting the goals set by Mann and Walsh. Following a brief overview of the text, the bulk of this review will focus on those contributions.

Watanabe's book presents a multiple-case study regarding the impact of three "reflective interventions" (journaling, focus group interviews, and individual interviews; p. 17) on the professional development of seven Japanese English teachers. The overall goal of the study was to explore the value of RP as a form of professional development for teachers in the Japanese context. The first three chapters provide background to the study, including reviews of relevant literature on RP and an exploration of issues. Chapter 4 introduces the reflective continuum, Watanabe's conceptualization of reflection as a recursive process involving five major forms of reflection: description, reconfirmation, *hansei*, reinterpretation, and awareness. *Hansei* is a form of "self-critical reflection" commonly practiced in schools and workplaces in Japan, in which a person acknowledges some fault and pledges to improve (p. 27). This is usually seen as a negative, problem-focused process rather than a developmental one. Chapters 5 and 6 cover the analyses of data from novice and experienced teachers respectively, focusing on changes in teacher identity arising from RP. Chapter 7 presents analyses of the development of teacher cognition—the mental resources and processes underlying teacher thought and decision making (see Borg, 2006)—in each participant. Chapter 8 highlights the value of the reflective interventions, focusing on the challenges of using them and on the differences among the participants in the individual interviews. The concluding chapter presents implications for using RP for professional development in Japan, including a reconceptualization of teacher development as "repertoire expansion," which Watanabe characterizes as the expansion of a teacher's "expert cognition and their professional identity in ways that allow them to be more versatile in a wider variety of contexts" (p. 167). In this view, the source of development is seen as internal to each teacher, arising from their unique experiences, and providing "not just model solutions to problems, but a more integrated sense of how to grow as a teacher" (p. 169).

Watanabe's study clearly illustrates the impact of context on RP. Watanabe designed her study taking into account key Japanese concepts including *honne* ("one's real intent") and *tatemae* ("one's public front"; p. 27) as well as *hansei*. Her selection of reflective interventions and the ways in which individual and focus group interviews were carried out were intended to

create environments in which participants could express *honne* and in which reflection would not necessarily carry the negative connotations of *hansei*. However, as Watanabe acknowledges, in the final analysis, "it was not possible . . . to truly know if the data represents *honne* or *tatemae*" (pp. 166-167) and that what might be more important is how the interventions gave the participants the opportunity to determine "what they *wanted* to say rather than what they were expected to say" (p. 167; italics original). Moreover, *hansei* came to be seen by Watanabe as a necessary stage of the reflective continuum for her participants because of how embedded it was in their experiences, ultimately granting them "autonomy in their journey towards embracing reflective practice" (p. 167).

Another Japanese cultural concept that emerged as a significant influence on RP over the course of the study was *kotodama*, a Shintoist concept that posits that "verbalisation of inner thoughts leads to mystical power" (p. 34). The presence of the words in the world may bring about unwanted outcomes or create situations in which the speaker cannot act constructively. This understanding shapes linguistic avoidance practices in certain Japanese contexts. Watanabe found that this concept may have influenced what some of her participants choose to write in their journals, avoiding topics that they felt unable to take any constructive action on. As Watanabe notes, RP as a Western practice might see topic choice in reflection as "value free" (p. 35), but such a stance might not hold true for Japanese teachers.

Watanabe's data collection methods, especially the focus group and individual interviews, provide a rich grounding for her analyses of spoken language data that arise in collaborative contexts, as Mann and Walsh (2013) called for. In her data collection process, Watanabe included three focus group interviews in order to create a less isolating environment for the participants, one in which they might be more likely to express *honne* with peers than they would with the researcher in individual interviews, and to "give them a forum to debrief each other" (p. 19) regarding their participation in the research. There were also six monthly, open-ended, individual interviews with each participant that were intended to help Watanabe "get to know [the participants] better and the views they expressed" (p. 20) more directly than through the journals.

Watanabe also used both of these interventions to invite greater collaboration with her participants and give them greater voice in the outcomes of the study. Two of the three focus group interviews included time spent reflecting on participation in the study. The fifth of the six individual interviews was built around participants' identification of reflective themes

through a review of previous interview transcripts. In their final journal entry, participants were also asked to reread their journal entries and identify reflective themes for use in Watanabe's analysis.

Extended samples of interaction between Watanabe and the participants in the individual interviews are provided throughout the text as support for the analysis and show how Watanabe as researcher took on a mentoring role, especially for the novice participants. This use of the data allows readers to see RP in action. It is unfortunate that similar extended samples of the interactions among participants in the focus group interviews are not provided. This lack may be a result of Watanabe's focus on the participants as separate cases in her analysis. However, providing such data might have allowed readers to see more clearly how peer collaboration played a role in each participant's professional development.

Watanabe critiques current models of RP for seeing development as linear, culminating in a level of criticality. She presents her reflective continuum as an alternative, recursive model of development with multiple points of access and no defined end point. As such, her model resembles more open-ended approaches to RP, such as double-looped learning (Argyris & Schon, 1974). This conceptualization of RP, supported by data showing the differences among her participants in their development, may be a more accurate representation of how RP works in the real world.

Overall, *Reflective Practice as Professional Development* should be recommended to anyone with an interest in RP, especially in carrying out RP as an evidence-based practice. Watanabe's detailed presentation of her cases and data collection procedures as well as the rich contextualization of the entire study and focus on open-ended professional development will provide a model for future research in RP that can only make the achievement of Mann and Walsh's (2013) goal for the field more likely.

References

Argyris, C., & Schon, D. A. (1974). *Theory in practice: Increasing professional effectiveness*. San Francisco, CA: Jossey-Bass.

Borg, S. (2006). *Teacher cognition and language education*. London, England: Continuum.

Mann, S., & Walsh, S. (2013). RP or 'RIP': A critical perspective on reflective practice. *Applied Linguistics Review, 4,* 291-315. https://doi.org/10.1515/applirev-2013-0013

Information for Contributors

All submissions must conform to *JALT Journal* Editorial Policy and Guidelines.

Editorial Policy

JALT Journal, the refereed research journal of the Japan Association for Language Teaching (*Zenkoku Gogaku Kyouiku Gakkai*), invites empirical and theoretical research articles and research reports on second and foreign language teaching and learning in Japanese contexts. Submissions from Asian and other international contexts are accepted if applicable to language teaching in Japan. Areas of particular interest include but are not limited to the following:

1. Curriculum design and teaching methods
2. Classroom-centered research
3. Intercultural studies
4. Testing and evaluation
5. Teacher training
6. Language learning and acquisition
7. Overviews of research and practice in related fields

The editors encourage submissions in five categories: (a) full-length articles, (b) short research reports (*Research Forum*), (c) essays on language education framed in theory and supported by argumentation that may include either primary or secondary data (*Perspectives*), (d) comments on previously published *JALT Journal* articles (*Point to Point*), and (e) book and media reviews (*Reviews*). Articles should be written for a general audience of language educators; therefore, statistical techniques and specialized terms must be clearly explained.

Guidelines

Style

JALT Journal follows the *Publication Manual of the American Psychological Association*, 6th edition (available from APA Order Department, P.O. Box 2710, Hyattsville, MD 20784, USA; by email: <order@apa.org>; from the website: <www.apa.org/books.ordering.html>). Consult recent copies of *JALT Journal* or *TESOL Quarterly* for examples of documentation and references. A downloadable copy of the *JALT Journal* style sheet is also available on our website at <http://jalt-publications.org/jj/>.

Format

Full-length articles must not be more than 20 pages in length (6,000 words), including references, notes, tables, and figures. *Research Forum* submissions should not be more than 10 pages in length (3,000 words). *Perspectives* submissions should not be more than 15 pages in length (4,500 words). *Point to Point* comments on previously published articles should not be more than 675 words in length, and *Reviews* should generally range from 500 to 1000 words. All submissions must be word processed in A4 or 8.5 x 11" format with line spacing set at 1.5 lines. **For refereed submissions, names and identifying references should appear only on the cover sheet**. Authors are responsible for the accuracy of references and reference citations.

Submission Procedure

Please submit the following materials, except for reviews, as email attachments in MS Word format to the appropriate editor indicated below:

1. Cover sheet with the title and author name(s), affiliation(s), and contact information of corresponding author.
2. Manuscript, including title, abstract, and keywords, with no reference to the author. Do not use running heads. Follow the *JALT Journal* style sheet.

If the manuscript is accepted for publication, a Japanese translation of the abstract will be required. Authors will also be asked to provide biographical information. Insert all tables and figures in the manuscript. Do not send as separate files.

Submissions will be acknowledged within 1 month of their receipt. All manuscripts are first reviewed by the Editor to ensure they comply with *JALT Journal* Guidelines. Those considered for publication are subject to blind review by at least two readers, with special attention given to (1) compliance with *JALT Journal* Editorial Policy, (2) the significance and originality of the submission, and (3) the use of appropriate research design and methodology. Evaluation is usually completed within 3 months. Each contributing author of published articles and Book Reviews will receive one

complimentary copy of the *Journal* and a PDF of the article (Book Reviews are compiled together as one PDF). *JALT Journal* does not provide off-prints. Contributing authors have the option of ordering further copies of *JALT Journal* (contact JALT Central Office for price details).

Restrictions

Papers submitted to *JALT Journal* must not have been previously published, nor should they be under consideration for publication elsewhere. *JALT Journal* has First World Publication Rights, as defined by International Copyright Conventions, for all manuscripts published. If accepted, the editors reserve the right to edit all copy for length, style, and clarity without prior notification to authors. Plagiarism, including self-plagiarism, will result in articles not being published or being retracted and may also result in the author(s) being banned from submitting to any JALT publication

Full-Length Articles, Research Forum, Perspectives, and Point to Point Submissions

Please send submissions in these categories or general inquiries to:

jj-editor@jalt-publications.org

Eric Hauser, *JALT Journal* Editor

Japanese-Language Manuscripts

JALT Journal welcomes Japanese-language manuscripts on second/foreign language teaching and learning as well as Japanese-language reviews of publications. Submissions must conform to the Editorial Policy and Guidelines given above. Authors must provide a detailed abstract in English, 500 to 750 words in length, for full-length manuscripts and a 100-word abstract for reviews. Refer to the Japanese-Language Guidelines (following page) for details. Please send Japanese-language manuscripts to:

jj-editorj@jalt-publications.org

Yo In'nami, *JALT Journal* Japanese-Language Editor

Reviews

The editors invite reviews of books and other relevant publications in the field of language education. A list of publications that have been sent to JALT for review is published bimonthly in *The Language Teacher*. Review authors receive one copy of the *Journal*. Please send submissions, queries, or requests for books, materials, and review guidelines to:

jj-reviews@jalt-publications.org

Greg Rouault, *JALT Journal* Reviews Editor

Address for Inquiries about Subscriptions, Ordering *JALT Journal*, or Advertising

JALT Central Office
Urban Edge Building 5F
1-37-9 Taito, Taito-ku, Tokyo 110-0016, Japan
Tel.: 03-3837-1630; Fax: 03-3837-1631
(From overseas: Tel.: 81-3-3837-1630; Fax: 81-3-3837-1631)
Email: jco@jalt.org URL: www.jalt.org

日本語論文投稿要領

*JALT Journal*では日本語で執筆された論文、研究報告、実践報告、書評等を募集しています。文体:一般的な学術論文のスタイルを用い、章立ての仕方や参考文献のデータの書き方などは、*Publication Manual of the American Psychological Association* (6th ed.)の定める方式に合わせて下さい。不明の場合は、*JALT Journal* の英語論文を参考にするか、日本語編集者までお問い合わせ下さい。なお、*JALT Journal*の読者は現場の教師が主なので、特殊な専門用語や統計的手法は、わかりやすく定義するか説明を加えるなどして下さい。原稿: 長さは、参考文献リストも含め18,000字 (書評の場合は1,500字) 以内です。A4の用紙に横書きで、1行40字、1ページ30行で印刷して下さい。手書きの原稿は受け付けません。

提出するもの:
以下の原稿を電子メールの添付書類、あるいは郵送でお送りください。

- 執筆者の名前と所属機関名を書いた表紙
- MS-Word で保存した本文(執筆者は無記名のこと)
- 執筆者連絡先(住所、電話番号、ファックス、e-mail アドレス)
- 400字以内の和文要旨
- 英文のタイトルと、500〜750語の英文要旨(書評の場合は100語程度の英文要旨)
- 100字以内の執筆者略歴
- 審査を経て掲載の認められた草稿は、図表などを全て写植版にしたものにして提出すること

査読: 編集委員会で投稿要領に合っているかどうかを確認したあと、少なくとも二人の査読者が査読を行います。査読者には執筆者の名前は知らされません。査読の過程では特に、原稿が*JALT Journal*の目的に合っているか、言語教育にとって意味があるか、独創性はあるか、研究計画や方法論は適切か等が判定されます。査読は通常二か月以内に終了しますが、特に投稿の多い場合などは審査にそれ以上の時間がかかることがあります。

注意: *JALT Journal*に投稿する原稿は、すでに出版されているものや他の学術雑誌に投稿中のものは避けて下さい。*JALT Journal*は、そこに掲載されるすべての論文に関して国際著作権協定による世界初出版権を持ちます。なお、お送りいただいた原稿は返却しませんので、控を保存して下さい。

投稿原稿送り先またはお問い合わせ:

〒112-8551 東京都文京区春日1-13-27中央大学理工学部英語教室
JALT Journal 日本語編集者　印南　洋
電話: 03-3817-1950
jj-editorj@jalt-publications.org

JALT Journal 第 40 巻　第 1 号

2018年 4月20日　印刷
2018年 5月1日　発行
編集人　ハウザー　エリック
発行人　リチモンド・ストゥループ
発行所　全国語学教育学会事務局
〒110-0016　東京都台東区台東1-37-9 アーバンエッジビル5F
TEL(03) 3837-1630; FAX(03) 3837-1631
印刷所　コーシンシャ株式会社
〒530-0043　大阪市北区天満1-18-4天満ファーストビル301 TEL(06)6351-8795

watch,

http://bit.ly/jaltYTchan

like,

https://www.facebook.com/JALT2018

follow,

https://twitter.com/jaltconference

and go . . .

diversity & inclusion

#jalt2018

JALT2018

44th Annual International Conference on Language Teaching and Learning & Educational Materials Exhibition

November 23 – 26, 2018

Shizuoka Convention & Arts Center "GRANSHIP"
Shizuoka City, Shizuoka Prefecture, JAPAN

https://jalt.org/conference/jalt2018