jalt journal

The research journal of the Japan Association for Language Teaching

Volume 45 • No. 2 • November 2023

JALT

全国語学教育学会

Japan Association for Language Teaching

¥1,900 ISSN 0287-2420

JALT Journal

Volume 45 • No. 2

November 2023

Contents

November 2023
Volume 45 • No. 2

Article

Perspectives

Expositions

Reviews

JALT Journal Information

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 3,000 language teachers. JALT has 32 JALT chapters and 30 special interest groups (SIGs) and 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 Postconference Publication*.

The JALT International Conference on Language Teaching and Learning and Educational Materials Exposition attracts some 2,000 participants annually and offers more than 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 copies of JALT publications, free or discounted admission to JALT-sponsored events, and optional membership in one chapter and one SIG. For an annual fee of ¥2,000 per SIG, JALT members can join as many additional SIGs as they desire. For information about JALT membership, contact the JALT Central Office or visit the JALT website.

JALT National Officers (2023)

President: . . . . . *Clare Kaneko*
Vice President: . . . . . *Kenn Gale*
Auditor: . . . . . *Robert Chartrand*
Director of Treasury: . . . . . *Michael Mielke*
Director of Records: . . . . . *Samantha Kawakami*
Director of Program: . . . . . *Wayne Malcolm*
Director of Membership: . . . . . *Julie Kimura*
Director of Public Relations: . . . . . *William Pellowe*
Acting Director of Program: . . . . . *Chelanna White*

Chapters

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

Special Interest Groups

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

JALT Central Office

Level 20, Marunouchi Trust Tower–Main,
1-8-3 Marunouchi, Chiyoda-ku, Tokyo 100-0005 Japan
Tel.: (+81) 3-5288-5443; Email: jco@jalt.org;
Website: <https://jalt.org>

In This Issue

Articles

In the full-length research article, **Noriko Iwamoto** explores the association between self-assessment and L2 speaking proficiency by analyzing their relationship to English-speaking anxiety and motivation, and self-esteem. Iwamoto identifies trends in overestimation and underestimation by learners depending on their English-speaking proficiency, and presents data on the relationship between students' self-assessments and teacher assessment of student proficiency.

The Perspectives article by **Tim Stoeckel**, **Stuart McLean**, **Paul Raine**, **Hung Tan Ha**, **Nam Thi Phuong Ho**, and **Young Ae Kim** unpacks an online vocabulary testing platform. Stoeckel and colleagues introduce the online resource as one option for teachers and researchers to create, deliver, and score contextualized meaning-recall tests for vocabulary learning. This discussion includes a framework for understanding form-recognition, form-recall, meaning-recognition, and meaning-recall formats with a focus on the amount of contextualization provided in a given test item.

The three *Expositions* articles are by **Paul Nation**, **Shelley Staples**, and **Laurence Anthony**, and incoming *JALT Journal* Book Reviews Editor **Melodie Cook**. In the first article, Nation explains why Japanese learners of English have difficulties with word parts, and how limited knowledge of word parts impacts vocabulary size, reading in English, and language proficiency in general. The second *Expositions* article by Staples and Anthony clarifies the relevance of corpus linguistics to English language learning and teaching. The authors also note the multi-modal quality of corpus data, and the potential of corpus-based analysis to reveal systemic patterns in everyday language use not necessarily obvious to learners or teachers. In the final *Expositions* article, Cook provides a concise history of academic book reviews, which serves to clarify their intended purposes as valuable scholarly contributions. This discussion grounds Cook's outline for *JALT Journal*'s new, critically-oriented approach to book reviews. The underlying intentions behind this new approach are to provide potential reviewers a platform upon which they can conduct thorough and informed critical analyses of new contributions to our field of research, and to further affirm *JALT Journal*'s importance to the development of applied linguistic research in Japan and beyond. To facilitate this process, Cook ends her article with guidelines for future book reviewers to consider.

Reviews

This issue features three book reviews. To begin, **Ian Allensworth** summarizes Ema Ushioda's call for a new ethical agenda for research into language learning motivation which draws specifically from critical language theory, and even more broadly from critical theory. **John Bankier** then reviews an edited volume of nine studies into social networks in language learning and language teaching. And finally, **Lachlan Jackson** takes up David Block's call for researchers to situate political economy in general, and social class in particular, as central to their work in an emerging neoliberal world with growing inequalities.

Parting Acknowledgements and Farewell from Greg Rouault

In this issue, my last as Reviews Editor (since starting back in May 2012), I would like to take the time to recognize the materials publishers for their support, without which a book reviews column would not be possible. Special appreciation goes to my Assistant Reviews Editors, Bill Snyder and John Nevara (twice!), for their collaborative input and sage counsel. Further acknowledgements go out to the current and past editorial and production teams at *JALT Journal* and in particular to past JJ Reviews Editors, Bill Perry and Yuriko Kite, for entrusting me to carry on with their great work. Finally, the active engagement and feedback from JALT members, readers of *JALT Journal*, and the review authors themselves have made this volunteer opportunity a most satisfying experience . . . One which without Paul Lewis's warmth, Malcolm Swanson's patience, Theron Muller's training, and Scott Gardner's subtlety would not have come to be as a member of JALT Publications.

— Greg Rouault, Reviews Editor (May 2012 - Nov 2023)

From the Editors

The *JALT Journal* team would like to begin this issue's editorial note by extending our sincere appreciation and gratitude to **Greg Rouault**, who has served as *JALT Journal*'s Reviews Editor from May 2012 to the current November 2023 issue. Greg's decade-long service to *JALT Journal* is noteworthy. He has been an important contributor to the health and strength of the journal, a thread between various phases in the journal's evolution, and a supporter of and facilitator for more than 150 academics, some of

whom began their scholarly careers with book reviews. Academic discourse would simply not exist without the volunteer spirit and actions of countless academics worldwide, and Greg's contribution has been impressive in this regard to say the least. Thank you, Greg. We wish you all the best in your future endeavors. Taking Greg's place is incoming Reviews Editor Melodie Cook, who is a former editor (May/November issues 2014 and 2015) of the *JALT Journal*. We look forward to working with her in her new role, and we encourage readers to read her Expositions article about book reviews in the current issue. *JALT Journal* has another new member of the editorial team, with Joe Geluso joining us as English-language Assistant Editor.

As we all continue to navigate the personal, professional, and academic changes and developments resulting from shifting back to pre-pandemic practices, the *JALT Journal* editorial team would like to thank the Publications Board, Editorial Advisory Board, additional readers, copy editors, and the production team for their unwavering support. We would like to reiterate that *JALT Journal* maintains its commitment to publishing high-quality English-language and Japanese-language research relevant for language learning and teaching within the Japanese context and beyond. We invite readers to read our "Aims and Scope" section in the backmatter, and consider submitting their research for publication in *JALT Journal*. Submissions in English should be submitted to our online submission platform at https://jalt-publications.org/content/index.php/jj/information/authors in either Rich Text or Microsoft Word Format (NOT PDF). Materials in Japanese should be emailed to the Japanese language-editor, Kiwamu Kasahara at jalt-pubs.jj.ed.j@jalt.org. Please refer to the Back Matter for further details. In addition, from the May 2024 issue, the Appendix section for all articles will transition to an online-only format. These will be made accessible through the *JALT Journal* website.

— Dennis Koyama, *JALT Journal*, Editor
— Jeremie Bouchard, *JALT Journal*, Associate Editor
— Joe Geluso, *JALT Journal*, Incoming Assistant Editor

JALT Journal Online

https://jalt-publications.org/jj

45 Volumes

84 Issues

Hundreds of articles!

1979~2023

Article

Relationship Between L2 Proficiency and Psychological Traits With Self-Assessment Bias Among L2 Speakers

Noriko Iwamoto
Toyo University

Self-assessment is sometimes used to assess second language (L2) skills, but it has a degree of error that is possibly caused by L2 learners themselves, resulting from their L2 proficiency level and psychological traits. This study, involving 196 Japanese university students, calculated the self-assessment bias of L2 speakers using many-facet Rasch measurement. Correlation analysis explored the relationship between self-assessment bias and L2 speaking proficiency with psychological traits including self-esteem, English speaking anxiety, and English speaking motivation. The results showed that self-assessment bias was related to the belief in effortism, in which a person correlates ability with intensity of effort, and the Dunning-Kruger effect, in which high-proficiency learners tend to underestimate and low-proficiency learners tend to overestimate their abilities. Specifically, overestimation related to low proficiency prevents L2 learners from accurately assessing their performance, whereas underestimation related to L2 learners' belief that they lacked the effort to improve their speaking skills.

自己評価は時折第二言語技能の評価に使用されるが、ある程度の誤差が生じる。おそらくこの誤差には、第二言語 (L2) 習熟度や心理的特徴など、学習者自身によってもたらされる誤差が含まれる。本研究では196名の日本人大学生を対象に、多相ラッシュ分析を使ってL2スピーキング自己評価のバイアスを測定した。そして相関分析を用いて、自己評価バイアス、スピーキング習熟度、心理的特徴（自尊心、L2スピーキングの不安とモチベーション）の関係を調査した。その結果自己評価バイアスは、努力主義の信念とダニング・クルーガー効果と関連があることが明らかとなった。特に英語習熟度の低い学習者は自身のスピーキングパフォーマンスを過大評価

https://doi.org/10.37546/JALTJJ45.2-1

JALT Journal, Vol. 45, No. 2, November 2023

する傾向があり、一方でスピーキング技能を伸ばす努力が不足していると考える学習者は過小評価する傾向にあることがわかった。

Keywords: effort; English speaking anxiety; English speaking motivation

Self-assessment is sometimes used to assess second language (L2) skills. Self-assessment is considered beneficial for teachers because it provides them with "some idea of how the students view their own language abilities and development" (Brown, 2005, p. 58). It is also beneficial for learners because, by reflecting on their own skills and improvements, they increase their autonomy and motivation for learning (Brown & Hudson, 1998; Oscarson, 1989).

The question then arises whether learners can accurately evaluate their own L2 skills. High correlations greater than .70 between self-assessment and L2 proficiency have been reported in some studies (AlFallay, 2004; Babaii et al., 2016; Bachman & Palmer, 1989; Le Blanc & Painchaud, 1985). Other studies have found low correlation coefficients below .30 or no correlations (Brantmeier, 2006; Jafarpur, 1991; Trofimovich et al., 2016). In a meta-analysis involving ten studies, Ross (1998) reported a correlation between self-assessment and language performance of .633. Li and Zhang (2020), covering 67 studies, reported .466 and maintained that criterion type, training, and instruments are factors that exert a moderating effect.

Apart from the factors mentioned by Li and Zhang (2020), the inaccuracy of self-assessment can be attributed to L2 learners themselves, including their L2 proficiency and psychological traits. For example, L2 proficiency may influence self-assessment because of the Dunning-Kruger effect (Kruger & Dunning, 1999), in which high-proficiency learners tend to underestimate and low-proficiency learners tend to overestimate their abilities. L2 learners' psychological traits, such as self-esteem, L2 anxiety, and L2 motivation, have also been found to be related to their self-assessment (AlFallay, 2004; MacIntyre & Doucette, 2010; MacIntyre et al., 1997; Masgoret & Gardner, 2003). However, studies investigating the relation between self-assessment and psychological traits have often utilized L2 learners' self-assessment raw scores and tend to ignore self-assessment bias, or the degree of deviation from actual ability.

The present study used the many-facet Rasch measurement (MFRM) to calculate self-assessment bias-size measures of L2 speaking performance. The research examined L2 speaking skill because among the four skills, speaking seems to be most strongly connected to self-assessment, for

Underhill (1987) stated that when people talk with others, consciously or unconsciously, they are constantly assessing themselves in terms of how successfully they are communicating. Regarding psychological traits, the study chose self-esteem, L2 anxiety, and L2 motivation because self-esteem is considered to be related to the act of self-assessment (Heine et al., 2001), whereas anxiety and motivation have been extensively investigated in L2 acquisition literature. Therefore, this work dealt with the self-assessment bias of L2 speaking to investigate how the degree of deviation in the self-assessment of L2 speaking skills would relate to L2 learners' psychological traits, including self-esteem, L2 speaking anxiety, and L2 speaking motivation, as well as L2 speaking proficiency.

Literature Review

Most studies on the self-assessment of L2 skills have focused on the validity of self-assessment and often found overestimation of self-assessments compared with teacher ratings (Barrot, 2015; Hung et al., 2016; Jafarpur, 1991; Jassen-van Dieten, 1989; Suzuki, 2015; Trofimovich et al., 2016), probably owing to the "above-average effect," in which people tend to overestimate their own abilities and rate themselves above average (Dunning et al., 1989). Other researchers reported on the underestimation of self-assessments, especially among Asian students probably due to a modesty bias of their cultures (Aryadoust, 2015; Chen, 2008; Matsuno, 2009; Rian et al., 2014; Suzuki, 2009).

Markus and Kitayama (1991) explained that, in European and American cultural contexts, high self-esteem is a prerequisite for participating in independent and mutually approving relationships, and people in these relationships tend to view themselves positively. By contrast, in East Asian contexts, people tend to have lower self-esteem and tend to be self-critical, which is indispensable for mutually sympathetic relationships (Heine et al., 2001). Therefore, the modesty bias caused by lower self-esteem appears to be related to self-assessment. Indeed, because of their lower self-esteem, Asian participants were often found to evaluate their traits, abilities, or performance lower than Western counterparts (Farh et al., 1991; Heine et al., 2001). Therefore, self-esteem appears to be related to the act of self-assessment. However, few studies have investigated the correlation between self-esteem and self-assessment of L2 skills. Only AlFallay (2004) investigated the correlations between self-esteem and self-assessment of L2 presentation skills, reporting that participants with high self-esteem give higher self-assessment scores than the teacher-assessment, whereas participants

with low self-esteem are the most accurate in their self-assessment.

The overestimation and underestimation of L2 self-assessments can also be explained in terms of proficiency. Higher- and lower-proficiency learners tend to underestimate and overestimate their abilities, respectively (Barrot, 2015; Saito et al., 2020; Suzuki, 2015). Kruger and Dunning (1999) called this tendency the "Dunning-Kruger effect" in self-assessment and stated that those in the bottom quartile tend to overestimate their abilities because their incompetence prevents them from accurately evaluating their abilities, whereas top-quartile participants underestimate their abilities because they tend to consider their proficiency as similar to that of their peers.

To some extent, some researchers have investigated the influence of psychological traits on self-assessment of L2 skills. For L2 acquisition, L2 anxiety is one of the most important psychological factors influencing learners' L2 learning and performance. MacIntyre and Gardner (1994) defined it as "the feeling of tension and apprehension specifically associated with second language contexts, including speaking, listening, and learning" (p. 284). Horwitz et al. (1986) developed the Foreign Language Classroom Anxiety Scale (FLCAS) as a measure of anxiety specific to foreign language learning. L2 speaking is considered the most anxiety-provoking skill (Horwitz et al., 1986; Phillips, 1992), and L2 anxiety has shown a negative relation with L2 learners' self-assessment of L2 speaking (Clément et al., 1994; Gardner & MacIntyre, 1993; MacIntyre et al., 1997).

L2 motivation is another factor that strongly influences L2 learning. Gardner and MacIntyre (1993) described the motivated individual as "one who wants to achieve a particular goal, devotes considerable effort to achieve this goal, and experiences satisfaction in the activities associated with achieving this goal" (p. 3). Based on the socio-educational model, Gardner (1985) developed a multicomponential motivation questionnaire called the Attitude/Motivation Test Battery (AMTB) and found that motivation is positively correlated with L2 skills. Masgoret and Gardener (2003) conducted a meta-analysis of the studies by Gardner and his associates that used the AMTB. By investigating 75 independent samples in their meta-analysis, Masgoret and Gardener found that motivation is more strongly correlated with the self-rating of L2 skills compared with the other two achievement measures of course grades and objective measures.

Studies investigating self-assessment of L2 skills in relation to psychological traits have often calculated correlations between self-assessment raw scores and psychological traits. However, they have not focused on bias, or the degree to which self-assessment deviates from criterion measures.

Meanwhile, Saito et al. (2020) used overconfidence scores, calculated by subtracting the mean external listeners' score from their self-assessment score; they found no significant correlation between overconfidence scores and the promotional orientation variable, which represents more enjoyment and less anxiety. Their findings differ from those of previous studies that used self-assessment raw scores, in which psychological traits are often correlated with self-assessment. An implication is that self-assessment and self-assessment bias-size measures may not be the same.

Therefore, this study calculated the self-assessment bias of L2 speaking skills using MFRM and investigated how the degree of self-assessment inaccuracy is related to psychological traits, including self-esteem, English speaking anxiety, and English speaking motivation, as well as L2 speaking proficiency. The study posed the following research questions:

RQ1. To what degree do Japanese students' self-assessments of their L2 oral performance differ from teacher-assessments?

RQ2. To what degree do self-esteem, English speaking anxiety, English speaking motivation, and L2 speaking proficiency relate to self-assessment bias-size measures of L2 oral performance?

Method

Participants

The participants in this study included L2 learners as well as raters. As for the L2 learners, 196 students (53 females and 143 males) participated. They majored in science and engineering or information science at two private universities in Japan. Most participants had the English skills of level A2 or B1 of the Common European Framework of Reference for Languages. The participants had studied English for at least six years in English courses in Japanese schools, where reading and grammar were more often the focus of instruction than communication skills. Given the EFL context, most of the participants had had few opportunities to speak English in their daily lives.

Next, four raters participated in the study. Two raters were L1 English speakers, and two raters were L1 Japanese speakers. All raters were English teachers at Japanese universities and had experience testing and marking Japanese students' speaking abilities. A summary of the raters' profiles is presented in Table 1. Rater 2, an L1 Japanese speaker, earned his M.A. and Ph.D. degrees from an American university and had experience teaching at an American university for 20 years, thus he helped to back-translate the questionnaire.

Table 1
Raters' Profiles

Rater	Gender	Age	Nationality	Educational background	Teaching position	Teaching experience
1	Female	30s	American	M.A. in English Literature	Adjunct professor	3 years
2	Male	60s	Japanese	Ph.D. in Philosophy	Professor	35 years
3	Male	50s	Canadian	M.A. in TESOL	Adjunct professor	20 years
4	Female	40s	Japanese	M.A. in TESOL	Associate professor	13 years

Instruments

This study used a questionnaire to measure self-esteem, English speaking anxiety, and English-speaking motivation. The self-esteem scale was based on the self-esteem scale of Rosenberg (1965). The English-speaking anxiety scale was based on the FLCAS (Horwitz et al., 1986). The English-speaking motivation scale was based on the motivation items from the AMTB (Gardner, 1985); modifications were made based on the questionnaire used by Gardner et al. (1997) and Irie (2005). The English-speaking motivation scale consisted of three components: attitude toward speaking English, desire to learn to speak English, and motivational intensity. These English questionnaire items were translated to Japanese by the author, and the Japanese translation was backtranslated to English by a bilingual professor, who also served as Rater 2. The back-translated questionnaire was compared with the original, and some modifications were made. The items were rated using a six-point Likert scale, where 1 = strongly disagree, 2 = disagree, 3 = slightly disagree, 4 = slightly agree, 5 = agree, and 6 = strongly agree.

The participants' speaking abilities were assessed using a two-part oral proficiency test consisting of an interview and picture task. Each participant took part in a two-part five-minute oral interview. During the first part, the participant was asked to respond to five questions, such as "What is your hobby?" and "What did you do last weekend?" In the latter part, the participant was asked to tell a story in English while looking at a four-panel cartoon adapted from the pre-first level interview questions of the Eiken Test in

Practical English Proficiency (Obunsha, 2010). This test was used because it is the most widely known English proficiency test in Japan, supported by Japan's Ministry of Education, Culture, Sports, Science and Technology; many Japanese students are familiar with it.

The research assessed the participants' performance using the Kanda English Proficiency Test (KEPT) scale used by Bonk and Ockey (2003). This scale was selected because it was developed to assess the speaking skills of Japanese university students. KEPT has been successfully used as a diagnostic and placement test for Japanese university students (Ockey, 2009, 2011; Van Moere, 2006). The original KEPT scale consists of five categories designed to assess oral discussion skills in a group of four students. The present study excluded one category, communicative strategies, which assesses how students respond to and interact with other students. This study only conducted individual oral interviews in which no interactions with other students were involved. Thus, the remaining four categories of grammar, vocabulary, fluency, and pronunciation were used. The English descriptions were translated to Japanese using a back-translation method with the help of Rater 2. Although the original KEPT has six levels, the first level, "Does not discuss," was deleted because all participants spoke English during the individual interviews. Therefore, the remaining five levels were used: 1 = very weak, 2 = weak, 3 = fair, 4 = good, and 5 = very good. Each level is accompanied by a description of the performance for that level. Because raters can use half points from levels 1 to 4, KEPT is a nine-point scale. The same scale was used for both the students' self-assessment and the teachers' assessment.

Data Collection

Data were gathered from May to August 2011. The study obtained written consent from those who agreed to participate in the study. They completed the questionnaire and submitted it during the individual oral interview with the author, a Japanese teacher of English. The interviews were audio-recorded with an IC recorder. After completing the oral interview, each participant was given a self-assessment sheet and was instructed to evaluate the four categories of grammar, vocabulary, fluency, and pronunciation by finding the description of the KEPT scoring levels that best matched their performance. The participants did not receive self-assessment training, which often improves the accuracy of self-assessment (Babaii et al., 2016; Chen, 2008), given that the objective of this study was to determine how the degree of self-assessment inaccuracy is related to psychological traits and proficiency.

Before conducting the teacher-assessment, each rater received an explanation of the oral interview and rating scale by the author. Using the audio-recorded data, four raters independently assessed each participant's oral proficiency using the same scoring rubric. MFRM does not require every rater to assess the complete data set, only that there is sufficient overlap (Linacre & Wright, 2002). Thus, to save time and labor, Rater 1 assessed students 1 to 146, Rater 2 assessed students 26 to 196, Rater 3 assessed students 97 to 196, and Rater 4 assessed students 1 to 25 and 97 to 196.

Data Analysis

The study analyzed the collected questionnaire data using Winsteps 3.80.1 (Linacre, 2013). The Rasch rating scale model provides several advantages in analyzing Likert-scale data over using raw scores (Apple, 2013). First, it changes ordinal raw scores to interval measures called logits. When the average is set at 0 logits; positive logits represent higher than average scores and negative logits, lower (Bond et al., 2021). The Rasch model also indicates the relative difficulty of each item and places both persons and items on the same single logit scale (Bond et al., 2021). Second, the Rasch model provides fit statistics that allow the identification of poorly performing items and raters. Reasonable infit and outfit MNSQ values fall within .5 and 1.5 (Linacre, 2007), which was used for this study. Third, researchers can check the dimensionality of the items hypothesized to measure the same trait using Rasch principal components analysis (PCA) of item residual analysis (Bond et al., 2021) which is generally used to determine unidimensionality. The criteria for determining unidimensionality are that over 50% is necessary for the variance explained by measures, and that the first contrast should account for either less than 10% of the variance and/or the eigenvalue should be less than 3.0 (Linacre, 2007). The present study used these criteria.

Participants' speaking data were examined using MFRM, which is an extension of the Rasch model. In addition to person ability and item difficulty, MFRM can assess other variables such as tasks and raters (Linacre, 2014). MFRM was used for L2 speaking assessment in this study because of its advantages over conventional approaches. First, MFRM can provide estimates of ability adjusted for rater bias; in contrast, speaking scores in a conventional approach using raw scores are likely to be degraded owing to differences in rater severity/leniency (Bond et al., 2021). Second, the joint calibration of facets allows rater severity to be placed on the same scale as ratee performance and task difficulty. This enables researchers to "draw

useful, diagnostically informative comparisons among the various facets" (Myford & Wolfe, 2003, p. 404).

The study calculated students' ability measures using four teachers' ratings, processed via Facets 3.80.0 (Linacre, 2017). Next, the self-assessment ratings were separated from the teacher-ratings by weighing teacher-assessment scores at .001, and the self-assessment measures were calculated. Then, bias-size measures were calculated by subtracting teacher-assessment measures from self-assessment measures. A positive bias-size measures indicated a more lenient self-rater relative to the teacher raters. A negative bias-size measures indicated a more severe self-rater compared with the teacher raters.

Preliminary Analysis

The study applied a Rasch measurement model for the questionnaire items using Winsteps. Three self-esteem items did not fit the Rasch model and were deleted. To check for unidimensionality, the study conducted Rasch PCA for each construct, which were found to meet the criteria. Therefore, five factors were used in the main analysis: self-esteem, English speaking anxiety, attitude toward speaking English, desire to learn to speak English, and motivational intensity (see Appendix A for the descriptions of items with the Rasch analysis results).

Results

Research Question 1

The study used MFRM; the modeled facets were the raters, students, and assessment categories. Scale 9 was not used by any rater, probably because the participants were all science majors whose English-speaking abilities were not as high as English language majors for whom Bond and Ockey (2003) originally created the KEPT rubric. The eight-point scale met Linacre's (2002) criteria for effective category functioning: at least 10 responses were made for each category, the outfit MNSQ was below 2.00, and the step difficulty of each category advanced by at least .25 logits (Wolfe & Smith, 2007). Therefore, an eight-point scale was used in this study.

Table 2 provides a rater measurement report for teacher- and self-raters, and Table 3 provides that for teacher-raters only. The current study adopted Linacre's (2007) infit and outfit MNSQ criterion of .50 to 1.50, which indicates that the items do not greatly diverge from Rasch model expectations. As shown in Tables 2 and 3, the infit and outfit MNSQ values in the present

study met this criterion. The reliability estimate of .99 means that the raters were separated into different levels of severity. The significant chi-square value showed that all raters were not equally severe.

Table 2

Calibration Report for Teacher-Raters and Self-Raters

Rater	Logit Measure	SE	Infit MNSQ	Outfit MNSQ
1	.53	.04	.80	.81
2	-.22	.04	.90	.92
3	.59	.05	.78	.78
4	.39	.04	.86	.86
Self	1.24	.03	1.42	1.46

Note. Fixed (all same) chi-square = 823.3; df = 4, p < .001; separation = 11.29; reliability = .99.

Table 3

Calibration Report for Teacher-Raters

Rater	Logit Measure	SE	Infit MNSQ	Outfit MNSQ
1	.63	.05	.89	.91
2	-.39	.04	1.08	1.09
3	.69	.06	.97	.97
4	.42	.05	.98	.98

Note. Fixed (all same) chi-square = 358.1; df = 3, p < .001; separation = 8.70; reliability = .99.

Bias-size measures were calculated by subtracting the teacher-assessment measures from the self-assessment measures. Pearson correlations between teacher-assessments (TA) and self-assessments (SA) were calculated. The results showed that TA and SA were moderately correlated (r = .44, p < .001).

Research Question 2

To answer Research Question 2, two groups were created based on the bias-size measures from both ends. AlFallay (2004) used subjects with the

highest and lowest 25% scores to represent the opposing groups. This grouping method was used in the present study. An overestimation (OE) group and an underestimation (UE) group were created based on the highest and lowest 25% bias-size measures. The OE group included the top quarter students, or 51 students with bias-size measures above 1.4 logits, and the UE group included the bottom quarter students, or 50 students with bias-size measures below -.85 logits. The descriptive statistics of each variable for each group, as well as for all students, are shown in Table 4. A MANOVA was conducted to determine the effect of bias-size measures (overestimation or underestimation) on the TA, SA, and five variables. Wilks's Λ was significant, $F(7, 93) = 15.97$, $p < .01$, $\eta^2 = .55$. As a follow-up test, a series of ANOVAs were conducted for each dependent variable. Using the Bonferroni method, each ANOVA was tested at the .007 level (.05/7). The results showed that the ANOVAs for TA and SA were significant: TA, $F(1, 99) = 18.15$, $p < .007$, $\eta^2 = .16$; SA, $F(1, 99) = 50.06$, $p < .007$, $\eta^2 = .34$. Thus, UE students had significantly higher average teacher-assessment measures than OE students, whereas their self-assessment measures were significantly lower than those of OE students. However, the results showed no significant differences in psychological traits between the two groups, indicating that those who overestimated and underestimated their abilities tended to have similar levels of self-esteem, anxiety, and motivation.

Table 4

Descriptive Statistics for All, Overestimation, and Underestimation Groups

	All		Overestimation		Underestimation	
	M	*SD*	*M*	*SD*	*M*	*SD*
TA	-.03	1.54	-.68	.22	.63	.22
SA	-.00	2.79	1.66	.41	-2.50	.42
SE	.04	1.64	.11	.25	.03	.25
ANX	.54	1.49	.34	.23	.85	.23
ASE	.42	1.59	.47	.24	.40	.24
DLSE	.67	1.37	.63	.19	.88	.19
MI	-.95	1.87	-.95	.28	-1.19	.29

Note. TA = Teacher-Assessment; SA = Self-Assessment; SE = Self-Esteem; ANX = English Speaking Anxiety; ASE = Attitude toward Speaking English; DLSE = Desire to Learn to Speak English; MI = Motivational Intensity.

Next, the correlations between the eight variables were calculated for all students and for each group, and the results are shown in Tables 5, 6, and 7. Several differences were observed between the two groups. First, the bias-size measures of the OE group were negatively correlated with TA (r = -.54, p < .01), whereas those of the UE group were positively correlated with TA (r = .64, p < .01). The negative relation for the OE group indicated that those with lower speaking abilities were likely to overestimate their own performance. Meanwhile, the positive relation for the UE group indicated that those with lower speaking ability tended to assess themselves lower. Second, unlike the results of OE students, which showed no correlations between bias-size measures and SA, UE students had a high correlation between them (r = .95, p < .01). An implication is that the greater their negative bias, the lower their self-assessment measures, whereas the smaller their negative bias, the higher their self-assessment. However, such a regular and systematic relation was not observed among the OE students. Finally, no correlations were found between the bias-size measures and psychological traits of OE students, meanwhile the bias-size measures of UE students were significantly correlated with the two variables of desire to learn to speak English and motivational intensity.

Table 5

Correlations Among Teacher-Assessment, Self-Assessment, Bias-Size Measures, and Five Psychological Traits (N = 196)

Variable	1	2	3	4	5	6	7	8
1. TA	——							
2. SA	.44**	——						
3. Bias	-.12	.84**	——					
4. SE	.08	.05	.00	——				
5. ANX	-.09	-.18**	-.15	-.18*	——			
6. ASE	.24**	.21**	.09	.29**	-.23**	——		
7. DLSE	.33**	.20**	.02	.15*	-.05	.55**	——	
8. MI	.34**	.32**	.15*	.04	-.18*	.36**	.44**	——

Note. TA = Teacher-Assessment; SA = Self-Assessment; Bias = Bias-Size Measures; SE = Self-Esteem; ANX = English Speaking Anxiety; ASE = Attitude toward Speaking English; DLSE = Desire to Learn to Speak English; MI = Motivational Intensity.

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

Table 6

Correlations Among Teacher-Assessment, Self-Assessment, Bias-Size Measures, and Five Psychological Traits with Overestimation Group (N = 51)

Variable	1	2	3	4	5	6	7	8
1. TA	——							
2. SA	.85**	——						
3. Bias	-.54**	-.00	——					
4. SE	.12	.22	.13	——				
5. ANX	-.12	-.16	-.02	-.12	——			
6. ASE	.33**	.36*	-.05	.36**	-.06	——		
7. DLSE	.43**	.46**	-.07	.21	-.01	.65**	——	
8. MI	.28*	.32**	-.01	-.02	-.15	.24	.47**	——

Note. TA = Teacher-Assessment; SA = Self-Assessment; Bias = Bias-Size Measures; SE = Self-Esteem; ANX = English Speaking Anxiety; ASE = Attitude toward Speaking English; DLSE = Desire to Learn to Speak English; MI = Motivational Intensity.

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

Table 7

Correlations Among Teacher-Assessment, Self-Assessment, Bias-Size Measures, and Five Psychological Traits with Underestimation Group (N = 50)

Variable	1	2	3	4	5	6	7	8
1. TA	——							
2. SA	.84**	——						
3. Bias	.64**	.95**	——					
4. SE	.01	-.08	-.12	——				
5. ANX	-.14	-.13	-.10	-.31*	——			
6. ASE	.24	.24	.20	.20	-.31*	——		
7. DLSE	.27	.33*	.32*	.13	-.07	.55**	——	
8. MI	.43**	.42**	.35*	-.08	-.12	.48**	.44**	——

Note. TA = Teacher-Assessment; SA = Self-Assessment; Bias = Bias-Size Measures; SE = Self-Esteem; ANX = English Speaking Anxiety; ASE = Attitude toward Speaking English; DLSE = Desire to Learn to Speak English; MI = Motivational Intensity.

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

Discussion

Research Question 1

Research Question 1 compared students' self-assessment of L2 oral performance with teacher-assessment. The present study identified a correlation of .44. The moderate correlation suggested that self-assessments of L2 oral performance by EFL learners may not be very reliable. The results of rater severity showed that self-assessment (1.24 logits) was more severe than teacher-assessments (-.22 to .59 logits). An examination of the bias-size measures showed that the number of students who overestimated their ability was larger than those who underestimated it. For example, 71 students had bias-size measures above 1.0, whereas 41 students had measures below -1.0 logits. Although more students overestimated their proficiency, some students made excessive underestimation: The least ability measures were -6.23 logits by teacher-assessment and -9.73 logits by student-assessment. Twelve students assessed themselves lower than -6.23 logits. In contrast, the highest ability measures were not greatly different between the teacher-assessment (3.92 logits) and self-assessment (4.28 logits). Thus, excessively severe self-assessment measures made by some students might have caused a greater severity of self-assessment logits compared with the teacher-assessment logits.

Research Question 2

Research Question 2 examined the relation of self-assessment bias with L2 proficiency and psychological traits. As teacher-assessment fit the Rasch model (Table 3), teacher-assessment in this study is considered to be a reliable measure of learners' L2 speaking proficiency. This is because MFRM can produce person ability measures that are adjusted for rater bias as long as raters are internally consistent (Bond et al., 2021). First, the results for all students were checked. The correlation results, shown in Table 5, revealed that self-assessment was correlated with L2 proficiency and all psychological variables except self-esteem, whereas bias-size measures were correlated with anxiety and motivational intensity only. Self-assessment and bias-size measures showed different results, which could indicate that these two are not the same.

Regarding psychological traits, English speaking anxiety was negatively correlated with bias-size measures. This showed that those with greater anxiety tended to underestimate their own performance. This finding is in accordance with past studies that found a negative relationship between

anxiety and self-assessment of L2 speaking (Clément et al., 1994; Gardner & MacIntyre, 1993; MacIntyre et al., 1997).

In addition to anxiety, bias-size measures were correlated with motivational intensity. The positive correlation indicated that those who said they extended more effort tended to overestimate their own speaking performance. This seems to imply Japanese people's belief in effortism, a tendency to value effort (Okawa, 2016). Kariya (1995) explained that unlike Western societies, Japanese society tends to believe that students' academic achievements are mainly derived from their efforts rather than their innate abilities. For example, Sudo's (2015) empirical study of 3,436 Japanese elementary school students revealed that although only one-third of the students' academic achievements were significantly related to their studying hours, 76.5% believed that anyone could be good at school subjects if they put in enough effort. From this result, Sudo (2015) highlighted Japanese people's tendency to blame students' lack of effort for their poor grades while paying little attention to their innate abilities. Likewise, the participants in the present study probably connected their L2 speaking ability to how much effort they had made.

Next, the participants from the top and bottom quartiles, 51 OE students and 50 UE students, were compared to investigate the students whose self-assessment measures greatly deviated from the teacher-assessment. The UE group had significantly higher average L2 speaking ability than the OE group, but the former evaluated their performance significantly lower than the latter. SA and TA were correlated highly in both groups (r = .85 for OE and .84 for UE, p < .01), compared with all students (r = .44, p < .01), which was an unexpected result because the OE and UE students had extreme bias. A possible explanation is that the OE and UE students each had the same tendency of either overestimation or underestimation for their self-assessments, which might have caused greater correlational values. This may imply that high correlational values do not necessarily indicate the accuracy of self-assessment. Bond et al. (2021) also pointed out the problem with intercorrelations between judge ratings because "they can demonstrate only consistency among the rank orders of candidates. They do not tell us anything about the severity or leniency differences between judges" (p. 147). This also supports the importance of calculating bias-size measures; the inaccuracy of self-assessment may not be detected from the results of correlation analysis.

Several different correlation results were found between the OE and UE groups. First, the correlation coefficient for teacher-assessment and bias-

size measures was -.54 for OE students. This negative relation indicated that among OE students, those with lower L2 speaking proficiency tended to overestimate their own performance. This demonstrated the Dunning-Kruger effect, and because the OE students had lower proficiency than average, as Kruger and Dunning (1999) stated, their incompetency might have prevented them from accurately evaluating their performance. On the contrary, the teacher-assessment and bias-size measures for UE students had a positive relation of .64, indicating that among UE students, those with lower proficiency tended to underestimate their performance. Thus, the Dunning-Kruger effect, where lower-proficiency participants tended to overestimate their ability, was not observed among UE students. A possible explanation may be that the participants conducted self-assessments immediately after their oral performance had finished. Other studies conducted prior practice and/or peer-assessment, after which participants could think about their performance more objectively. Meanwhile, the self-assessments in the present study might have been more influenced by participants' immediate subjective feelings. If the participants were not satisfied with their performance during the interview, their negative feelings on their performance might have caused them to evaluate themselves significantly lower. Indeed, after the interview, some participants who could not speak English well hung their heads or lamented, "Oh, my English ability is so poor!" The disheartening feeling that they could not speak English as expected might have caused them to have a greater negative bias toward their own L2 performance. Thus, among UE students, those with lower proficiency tended to assess themselves lower than necessary.

Second, the correlation between bias-size measures and self-assessment differed between the two groups. The UE group had a high correlation of .95, whereas the OE group showed no significant correlation. UE students with a positive bias toward their L2 oral performance tended to give a higher self-assessment, whereas UE students with a negative bias tended to give a lower self-assessment. Meanwhile, OE students did not show such systematic relations. The inconsistency in the bias size for OE students could signify their incompetence in evaluating L2 oral performance properly owing to their lower proficiency (Kruger & Dunning, 1999).

Finally, the relation between bias-size measures and psychological traits differed between the two groups. Although the bias size of the OE group had no significant relation with any psychological traits, that of the UE group showed a weak relation with desire to learn to speak English (DLSE) and motivational intensity (MI), implying that when they had little desire

or made little effort to improve their spoken English, they were likely to underestimate their performance. Therefore, the belief in effortism seems to be related especially to underestimation. Although the degrees of DLSE and MI were not significantly different between the UE and OE students, only the UE students' DLSE and MI were related to self-assessment bias size. Thus, UE students may be more likely to feel that their lack of a great desire or sufficient effort to improve in speaking English could indicate their low level of speaking ability, leading them to underestimate their ability. In other words, they seem to believe that if they have a stronger desire and make more effort, then their English-speaking ability will improve. This idea is in accordance with a previous finding that Japanese self-perception tends to be critical because Japanese people strongly believe in improvement and achievement (Heine et al., 1999). Hung et al. (2016) stated that highly proficient learners underestimate their L2 abilities because they set high standards for themselves. Thus, for UE students, although their speaking ability is higher than average, they are critical of and underestimate their own L2 speaking skills, which may represent their dissatisfaction with the intensity of their desire and efforts and their belief in the possibility of improvement. Although lower self-esteem is considered to be related to self-criticism (Markus & Kitayama, 1991), in the present study, no significant correlations were found between self-esteem and bias-size measures in either OE or UE group, as both had the same level of self-esteem. Thus, contrary to the hypothesis, lower self-esteem showed no relation to underestimation, but belief in effortism seemed more related to underestimation.

This study's results have several implications. First, this study calculated bias-size measures, instead of raw scores for self-assessments, to elucidate the degree to which self-assessment could deviate from actual ability. Indeed, the use of bias-size measures obtained results different from those obtained using self-assessment measures. For example, all students' self-assessment measures were correlated with L2 proficiency, whereas the bias-size measures were not. Their self-assessment measures were correlated with most psychological traits, but the bias-size measures were correlated only with anxiety and motivational intensity. Research on the relation between self-assessment and psychological traits has mainly utilized self-assessment raw scores to explore for correlations. These previous studies revealed that L2 learners with certain psychological traits give higher or lower self-assessment scores for their abilities. By calculating self-assessment bias-size measures, the present study could explore how the degree of overestimation and underestimation is related to one's psychological traits. Therefore, in-

vestigation of self-assessment bias-size measures will provide new insights into complex self-assessment behaviors among L2 learners with different degrees of anxiety and motivation, along with other psychological traits.

Second, L2 proficiency was negatively correlated with the bias-size measures of OE students, indicating that lower-proficiency students tended to have a greater overestimation of their speaking abilities. As Kruger and Dunning (1999) explained, students' lack of competence probably prevents them from assessing their own ability accurately. Overestimation of one's own L2 ability based on one's incompetence is a problem requiring attention. If learners cannot realize their weaknesses, then future improvements could be stalled. As some studies have reported that the accuracy of self-assessment improves after training (Babaii et al., 2016; Chen, 2008), lower-proficiency students especially require training that may help them correctly view their own L2 ability.

The third implication is that L2 learners' self-assessment bias may be related to their (lack of) effort in speaking English. In particular, higher-proficiency learners had a tendency to view their L2 speaking performance negatively when they considered that they had not made sufficient efforts to improve their English speaking skill. Japanese people's belief in effortism encourages and motivates students to study harder, but at the same time it can be problematic because it diverts attention from actual abilities (Sudo, 2015). Although teachers should emphasize the significance of making an effort, they should also teach learners to view their actual abilities more accurately, without taking into account how much effort has been made.

Conclusion

This study analyzed the self-assessment bias-size measures for L2 oral performance of 196 students and considered their relations with self-esteem, English speaking anxiety, English speaking motivation, and L2 speaking proficiency. The results showed that self-assessment was moderately correlated with teacher-assessment. The self-assessment bias is believed to be attributed to weak influences of the belief in effortism and the Dunning-Kruger effect.

This study had two limitations. First, it involved only science majors from two Japanese universities; the inclusion of more participants with different majors and/or proficiencies is needed for generalization. For example, those majoring in English may have more confidence in their L2 abilities and evaluate their English skills differently. Second, because culture can influence the evaluative attitude of one's own abilities (Heine et al., 2001), participants

from different backgrounds should be included to elucidate the influences of culture, such as modesty bias and effortism, on self-assessment.

Despite these limitations, this research is among the few studies that have investigated the relation between self-assessment bias, L2 speaking proficiency, and psychological traits. Self-assessment is considered a highly complex metacognitive task (Butler & Lee, 2006) that cannot be explained by only a few factors. Thus, other variables might influence self-assessment. For example, qualitative data, such as interviews, might shed light on the factors that cause inaccuracy in self-assessment. Accurate assessment of one's own language skills is important for L2 acquisition; by recognizing their own strengths and weaknesses, learners can become more autonomous and monitor their own L2 skills. Given the EFL context of Japan, L2 learners have few opportunities to speak English outside their classrooms, unlike L2 learners in ESL contexts who can use their L2 in daily life and monitor their own speaking skills every day. If EFL learners learn to assess their own speaking proficiency, such as by becoming better informed of their biases, then they can gauge their own progress. Further research on the self-assessment of speaking skills is needed to enhance the L2 oral proficiency of EFL learners.

Noriko Iwamoto is a professor of English as a Foreign Language (EFL) in the Faculty of Science and Engineering at Toyo University. Her research interests include self-assessment of L2 skills and the influence of L2 affective factors on second language learning.

References

AlFallay, I. (2004). The role of selected psychological and personality traits of the rater in the accuracy of self- and peer-assessment. *System, 32*(3), 407–425. https://doi.org/10.1016/j.system.2004.04.006

Apple, T. M. (2013). Using Rasch analysis to create and evaluate a measurement instrument for foreign language classroom speaking anxiety. *JALT Journal, 35*(1), 5–28. https://doi.org/10.37546/JALTJJ35.1-1

Aryadoust, V. (2015). Self- and peer assessments of oral presentations by first-year university students. *Educational Assessment, 20*(3), 199–225. https://doi.org/10.1080/10627197.2015.1061989

Babaii, E., Taghaddomi, S., & Pashmforoosh, R. (2016). Speaking self-assessment: Mismatches between learners' and teachers' criteria. *Language Testing, 33*(3), 411–437. https://doi.org/10.1177/0265532215590847

Bachman, L. F., & Palmer, A. S. (1989). The construct validation of self-ratings of communicative language ability. *Language Testing, 6*(1), 14–29. https://doi.org/10.1177/026553228900600104

Barrot, J. (2015). Self- and teacher assessment of speaking performances: An investigation of interrater consistency and agreement. *Journal of Language and Literature, 6*(4), 193–200.

Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th ed.). Routledge. https://doi.org/10.4324/9780429030499

Bonk, W. J., & Ockey, G. J. (2003). A many-facet Rasch analysis of the second language group oral discussion task. *Language Testing, 20*(1), 89–110. https://doi.org/10.1191/0265532203lt245oa

Brantmeier, C. (2006). Advanced L2 learners and reading placement: Self-assessment, CBT, and subsequent performance. *System, 34*(1), 15–35. https://doi.org/10.1016/j.system.2005.08.004

Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment.* McGraw-Hill.

Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly, 32*(4)*,* 653–675. https://doi.org/10.2307/3587999

Butler, Y. G., & Lee, J. (2006). On-task versus off-task self-assessments among Korean elementary school students studying English. *The Modern Language Journal, 90*(4), 506–518. https://doi.org/10.1111/j.1540-4781.2006.00463.x

Chen, Y.-M. (2008). Learning to self-assess oral performance in English: A longitudinal case study. *Language Teaching Research, 12*(2), 235–262. https://doi.org/10.1177/1362168807086293

Clément, R., Dörnyei, Z., & Noels, K. A. (1994). Motivation, self-confidence, and group cohesion in the foreign language classroom. *Language Learning, 44*(3)*,* 417–448. https://doi.org/10.1111/j.1467-1770.1994.tb01113.x

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57*(6), 1082–1090. https://doi.org/10.1037/0022-3514.57.6.1082

Farh, J., Dobbins, G. H., & Cheng, B. (1991). Cultural relativity in action: A comparison of self-ratings made by Chinese and U.S. workers. *Personnel Psychology, 44*(1), 129–147. https://doi.org/10.1111/j.1744-6570.1991.tb00693.x

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitude and motivation.* Edward Arnold.

Gardner, R. C., & MacIntyre, P. D. (1993). A student's contribution to second language learning. Part II: Affective variables. *Language Teaching, 26*, 1–11. https://doi.org/10.1017/S0261444800000045

Gardner, R. C., Tremblay, P. F., & Masgoret, A. (1997). Towards a full model of second language learning: An empirical investigation. *The Modern Language Journal, 81*(3), 344–362. https://doi.org/10.1111/j.1540-4781.1997.tb05495.x

Heine, S. J., Kitayama, S., & Lehman, D. R. (2001). Cultural differences in self-evaluation: Japanese readily accept negative self-relevant information. *Journal of Cross-Cultural Psychology, 32*(4)*,* 434–443. https://doi.org/10.1177/0022022101032004004

Heine, S. J., Markus, H. R., Lehman, D. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review, 106*(4), 766–794. https://doi.org/10.1037/0033295X.106.4.766

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*(2), 125–132. https://doi.org/10.2307/327317

Hung, Y., Samuelson, B. L., & Chen, S. (2016). Relationships between peer- and self-assessment and teacher assessment of young EFL learners' oral presentations. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives* (pp. 317–338). Springer. https://doi.org/10.1007/978-3-319-22422-0_13

Irie, K. (2005). *Stability and flexibility of language learning motivation: A multimethod study of Japanese junior high school students* (Publication No. 3176834) [Doctoral dissertation, Temple University]. ProQuest Dissertations & Theses Global. https://www.proquest.com/dissertations-theses/stability-flexibility-language-learning/docview/305393271/se-2

Jafarpur, A. (1991). Can native EFL learners estimate their own proficiency? *Evaluation and Research in Education, 5*(3), 145–157. https://doi.org/10.1080/09500799109533306

Janssen-van Dieten, A.-M. (1989). The development of a test of Dutch as a second language: The validity of self-assessment by inexperienced subjects. *Language Testing, 6*(1), 30–46. https://doi.org/10.1177/026553228900600105

Kariya, T. (1995). *Taishū kyōiku shakai no yukue* [Prospect of public education society]. Chukou Shinsho.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessment. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. https://doi:10.1037/0022-3514.77.6.1121

Le Blanc, R., & Painchaud, G. (1985). Self-assessment as a second language placement instrument. *TESOL Quarterly, 19*(3), 673–687. https://doi:10.2307/3586670

Li, M., & Zhang, X. (2020). A meta-analysis of self-assessment and language performance in language testing and assessment. *Language Testing, 38*(2), 189–218. https://doi.org/10.1177/0265532220932481

Linacre, J. M. (2002). Optical rating scale category effectiveness. *Journal of Applied Measurement, 3*, 85–106.

Linacre, J. M. (2007). *A user's guide to WINSTEPS: Rasch model computer program*. https://www. winsteps.com

Linacre, J. M. (2013). *WINSTEPS* (version 3.80.1) [Computer software]. Beaverton. https://www. winsteps.com

Linacre, J. M. (2014). *A user's guide to FACETS: Rasch model computer program.* https://www. winsteps.com

Linacre, J. M. (2017). *FACETS* (version 3.80.0) [Computer software]. Beaverton. https://www. winsteps.com

Linacre, J. M., & Wright, B. D. (2002). Construction of measures from many-facet data. *Journal of Applied Measurement, 3*(4), 486–512.

MacIntyre, P. D., & Doucette, J. (2010). Willingness to communicate and action control. *System, 38*(2), 161–171. https://doi.org/10.1016/j.system.2009.12.013

MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning, 44*(2), 283–305. https://doi.org/10.1111/j.1467-1770.1994.tb01103.x

MacIntyre, P. D., Noels, K. A., & Clément, R. (1997). Biases in self-ratings of second language proficiency: The role of language anxiety. *Language Learning, 47*(2), 265–287. https://doi.org/10.1111/0023-8333.81997008

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224–253. https://doi.org/10.1037/0033295X.98.2.224

Masgoret, A., & Gardner, R. C. (2003). Attitudes, motivation, and second language learning: A meta-analysis of studies conducted by Gardner and associates. *Language Learning, 53*(1), 123–163. https://doi.org/10.1111/1467-9922.00227

Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing, 26*(1)*,* 75–100. https://doi.org/10.1177/0265532208097337

Myford, C. M., & Wolfe, E. W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement, 4*(4), 386–422.

Obunsha. (2010). *Eiken jun ikkyū nijishiken mensetsu taisaku kanzen yosōmondai* [The Eiken test in practical English proficiency pre-1 level interview test questions]. Obunsha.

Ockey, G. J. (2009). The effects of group members' personalities on a test taker's L2 group oral discussion test scores. *Language Testing, 26*(2), 161–186. https://doi.org/10.1177/0265532208101005

Ockey, G. J. (2011). Self-consciousness and assertiveness as explanatory variables of L2 oral ability: A latent variable approach. *Language Learning, 61*(3), 968–989. https://doi.org/10.1111/j.1467-9922.2010.00625.x

Okawa, K. (2016). *Ganbaru koto/ganbaranai koto no shakai-gaku: Doryoku shugi no yukue* [Sociology of making effort/not making effort: Prospect of effortism]. Harbestosha.

Oscarson, M. (1989). Self-assessment of language proficiency: Rationale and applications. *Language Testing, 6*(3), 1–13. https://doi.org/10.1177/026553228900600103

Phillips, E. M. (1992). The effects of language anxiety on students' oral test performance and attitudes. *The Modern Language Journal, 76*(1), 14–26. https://doi.org/10.2307/329894

Rian, J., Hinkelman, D., & Cotter, M. (2014). Self-, peer, and teacher assessments of student presentation videos. In P. Clements, A. Krause, & H. Brown (Eds.), *JALT2014 Conference Proceedings* (pp. 688–697). JALT.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton University Press. https://doi.org/10.1515/9781400876136

Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experimental factors. *Language Testing*, *15*(1), 1–20. https://doi.org/10.1177/026553229801500101

Saito, K., Trofimovich, P., Abe, M., & In'nami, Y. (2020). Dunning-Kruger effect in second language speech learning: How does self-perception align with other perception over time? *Learning and Individual Differences*, *79*, 101849. https://doi.org/10.1016/j.lindif.2020.101849

Sudo, K. (2015). *Shōgakusei no doryoku shugi no keisei yōin to kiketsu* [Formation factors and consequences of elementary school students' effortism]. Benesse. https://berd.benesse.jp/up_images/research/Survey_on_learning_report_3.pdf

Suzuki, M. (2009). The compatibility of L2 learners' assessment of self- and peer revisions of writing with teachers' assessment. *TESOL Quarterly, 43*(1), 137–148. https://doi.org/10.1002/j.1545-7249.2009.tb00233.x

Suzuki, Y. (2015). Self-assessment of Japanese as a second language: The role of experiences in the naturalistic acquisition. *Language Testing, 32*(1), 63–81. https://doi.org/10.1177/0265532214541885

Trofimovich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2016). Flawed self-assessment: Investigating self- and other-perception of second language speech. *Bilingualism: Language and Cognition, 19*(1), 122–140. https://doi.org/10.1017/S1366728914000832

Underhill, N. (1987). *Testing spoken language*. Cambridge University Press.

Van Moere, A. (2006). Validity evidence in a university group oral test. *Language Testing, 23*(4), 411–440. https://doi.org/10.1191/0265532206lt336oa

Wolfe, E. W., & Smith, E. V. Jr. (2007). Instrument development tools and activities for measure validation using Rasch models: Part II-Validation activities. *Journal of Applied Measurement, 8*(2), 204–234.

Appendix

Questionnaire Item Descriptions and Rasch Results

Items	Descriptions	logits
Self-Esteem		
person reliability = .83, item reliability = .98		
SE1	I believe that I have a number of good qualities.	-.98
SE2	I am able to do things as well as most other people.	-.95
SE3	I feel useful most of the time.	-.01
SE4	I feel that I am a person of worth.	-.11
SE5	I respect myself.	1.03
SE6	I am able to do things better than other people.	.34
SE7	I have more good points than weak points.	.69

English Speaking Anxiety person reliability = .77, item reliability = .95		
ANX1	I would feel nervous speaking English with native speakers of English.	-.46
ANX2	I feel nervous about speaking English in class activities.	.13
ANX3	I lack confidence in my English-speaking abilities.	-.51
ANX4	I worry that my English teacher thinks that my English-speaking level is low.	.72
ANX5	I worry that I will make mistakes when I speak English	-.01
ANX6	I feel nervous having a conversation in English.	.13
Attitude Toward Speaking English person reliability = .82, item reliability = .89		
ASE1	I enjoy speaking English.	-.25
ASE2	I enjoy speaking English more than reading English.	-.11
ASE3	I enjoy speaking English more than writing English.	-.44
ASE4	I look forward to my English-speaking classes.	.38
ASE5	I enjoy English speaking classes more than other classes.	.17
ASE6	I look forward to opportunities to speak English.	.25
Desire to Learn to Speak English person reliability = .79, item reliability = .97		
DLSE1	Speaking English is important for engineers.	-.81
DLSE2	I would take an English conversation course in school, even if it were not required.	.54
DLSE3	I wish I had more classes in which I could speak English.	.74
DLSE4	I really want to learn to speak English better.	-.44
DLSE5	I believe that Japanese students should be taught to speak English at school.	-.34
DLSE6	My desire to learn to speak English is increasing.	.30

Motivational Intensity person reliability = .84, item reliability = .88		
MI1	I think I try to speak English more than other students.	.08
MI2	I look for opportunities to speak English outside of class.	-.24
MI3	I spend a long time studying English.	.13
MI4	I study English more than most of my classmates.	.44
MI5	I often think about how I can improve my English-speaking skills.	-.48
MI6	I work hard to become an excellent speaker of English.	.09

Perspectives

A Contextualized Meaning-Recall Vocabulary Testing Platform

Tim Stoeckel
University of Niigata Prefecture

Stuart McLean
Kindai University

Paul Raine
Ritsumeikan University

Hung Tan Ha
University of Economics Ho Chi Minh City and *Victoria University of Wellington*

Nam Thi Phuong Ho
University of Economics Ho Chi Minh City

Young Ae Kim
Kindai University and *Kyoto Seika University*

In contextualized vocabulary assessment, target words appear in extended context. Compared to tests employing single-word or limited-context prompts, research suggests that contextualized assessment is more reliable and demonstrates better concurrent validity. In meaning-recall vocabulary assessment, examinees retrieve

https://doi.org/10.37546/JALTJJ45.2-2

JALT Journal, Vol. 45, No. 2, November 2023

target-word meaning from memory and typically demonstrate knowledge via a written L2-to-L1 translation. Compared to multiple-choice formats, meaning-recall yields more reliable data, correlates more strongly with reading comprehension, and is less influenced by guessing and test strategies. To facilitate these approaches to vocabulary assessment, this article introduces a resource for teachers and researchers to create, administer, and mark contextualized meaning-recall tests. Users input a passage, select target items, and share the test URL with examinees. Examinees then provide L1 translations or L2 synonyms, definitions, or explanations of target words in input boxes below the lines of text. Raters mark responses online, and these judgments can be saved for partial automatic marking in future test use.

文脈化された語彙測定では、ターゲット項目が段落の中に現れる。単一語彙または限られた文脈の項目を用いるテストに比べて、文脈化された測定はより信頼性が高く、より優れた併存的妥当性を示している。意味想起語彙テストでは、受験者はターゲット語彙の意味を思い出し、通常は第二言語（L2）から第一言語（L1）への書記による翻訳で知識を示す。多肢選択形式と比較して、意味想起はより信頼性のあるデータをもたらし、読解力とより強い相関を示し、推測やテスト戦略の影響を受けにくい。語彙測定へのこれらのアプローチを支持するために、本稿では教師や研究者が文脈化された意味想起テストを作成、実施、採点するためのリソースを紹介している。利用者は文章を入力し、問題項目を選択、テストのURLを受験者と共有する。受験者は、テキストの下にある入力ボックスに対象語のL1翻訳やL2同義語、定義、またはその説明を入力する。採点者はオンラインで回答を評価し、正答とみなされる解答は将来のテスト利用時の自動採点のために保存できる。

Keywords*:* contextualized meaning-recall test; meaning-recall; vocabulary assessment

The availability of second language (L2) vocabulary assessment instruments of both breadth and depth has increased markedly over the past 20 years. One aspect of vocabulary knowledge commonly assessed with these tests is the form-meaning link, or the ability to associate meaning with the written or spoken form of a word (Jiang, 2002; McLean, Kramer, & Beglar, 2015). One kind of form-meaning assessment that has attracted recent attention is that of meaning-recall knowledge, or the ability to retrieve the meaning of an L2 word from memory upon seeing or hearing the word form. Meaning-recall is typically assessed by eliciting L2-to-L1 translations, or L2 synonyms or explanations of tested words. Meaning-recall tests are better predictors of reading ability than multiple-choice instruments (i.e., meaning-recognition; Stewart et al., 2023; Zhang & Zhang, 2022), making them attractive for many research purposes (Stewart et al., 2021; Stoeckel et al., 2021).

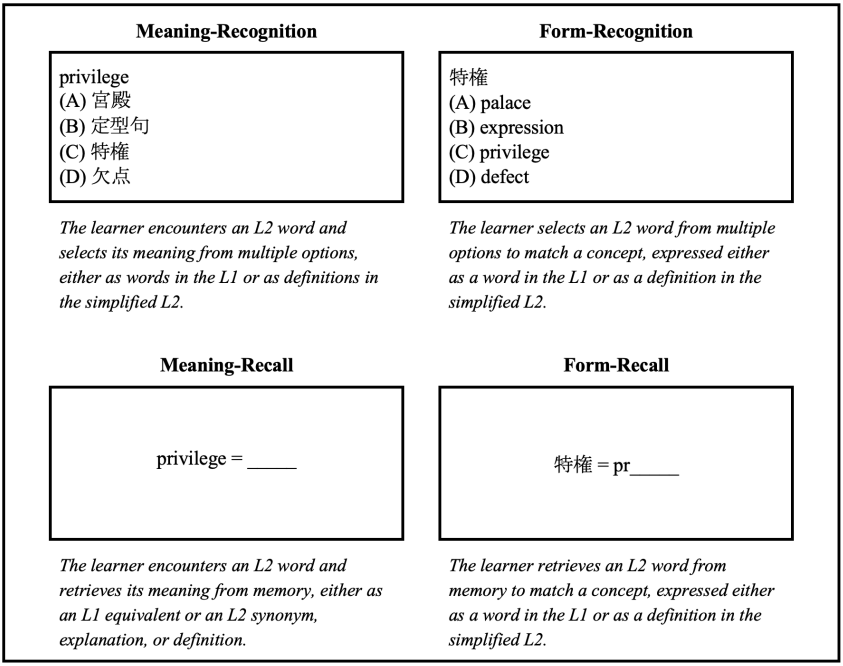
A drawback of meaning-recall assessment, it is sometimes argued, is that it is less practical, requiring more time to administer and mark tests (Webb, 2021a). One testing site, Vocableveltest.org (McLean et al., 2021), addresses

this problem with automated marking based on meticulously created banks of possible correct responses. Though this does not eliminate hand-marking, as novel responses do occur, it greatly reduces the time needed for scoring. A limitation of this tool, however, is that test makers must select from pre-existing lists of target words and test items. Though Vocableveltest.org offers an extensive array of choices, teachers and researchers may at times wish to assess other words and, more importantly, in contexts other than those appearing in the existing item bank. To address this shortcoming, the present paper introduces a contextualized meaning-recall test (CMRT) platform designed to complement the assessment options offered by Vocableveltest.org. It differs from Vocableveltest.org in that test makers input their own items and tests, meaning they can assess any target word or phrase desired. Moreover, vocabulary can be tested in contexts ranging from a single word to full-length passages. Though not as robust as the automated marking feature of Vocableveltest.org, the CMRT platform also allows for partial automated marking by saving manual ratings of responses for future test use. A beta version of the platform is currently available at https://cmrt.vocableveltest.org/. Though it shares a partial domain name with Vocableveltest.org, the two resources operate independently.

A Framework for Contextualized Assessment of Form-Meaning Knowledge

We begin by reviewing contextualized assessment of form-meaning knowledge. In form-meaning tests, vocabulary knowledge can be assessed at the levels of form-recognition, form-recall, meaning-recognition, and meaning-recall (Schmitt, 2010; Figure 1). The difference between recall and recognition is that in recall, examinees retrieve word meaning or form from memory, while in recognition they select meaning or form from a list of options. The difference between tests of form and meaning is that in the former, word meaning is provided in the test prompt and examinees must demonstrate knowledge of the L2 word form, whereas in the latter, the L2 form is provided, and examinees demonstrate understanding of its meaning.

Figure 1
Four Test Modalities of Form-Meaning Knowledge



Note. The meaning-recognition item is adapted from the New General Service List Test (Stoeckel et al., 2018).

Within these four modalities, several variations exist. First, as depicted in Figure 1, bilingual formats are sometimes used to reduce the risk of target word knowledge being conflated with the ability to understand other elements of the test item (e.g., Nguyen & Nation, 2011). Second, also shown in Figure 1, in form-recall tests, one or more letters of the target word are sometimes provided to limit possible correct responses to only the target item (e.g., Schmitt et al., 2021). Third, scoring of recall tests sometimes (e.g., Schmitt et al., 2021), but not always (e.g., Stoeckel et al., 2019), requires correct spelling of the desired response.

A fourth difference, and a focus of the present paper, is the amount of context in the test items. Existing tests of form-meaning knowledge range from those providing the target item devoid of any context to those in which the target structure is embedded in substantial context that may aid lexi-

cal inferencing. Figure 2 depicts four levels of contextualization for each of the previously described aspects of form-meaning knowledge. These levels are admittedly somewhat arbitrary but are meant to represent important points along a continuum of possibilities. At level 1, only the target item is provided in the item stem. This level of contextualization appeared in early versions of the TOEFL (Read, 2000) and is employed in the Vocabulary Levels Test and its variants (VLT; Nation, 1983; Schmitt et al., 2001). At level 2, the stem contains a short sentence with only enough context to orient test-takers to the target item's part of speech. The Vocabulary Size Test (VST; Nation & Beglar, 2007) is an example of an instrument using this level of contextualization. Level 3 item stems are also one sentence in length while level 4 are a paragraph or more. Levels 3 and 4 differ from level 2 in that they may contain information to aid lexical inferencing. Inference-generating information may be intentionally included in all test items, as in Sasao and Webb's (2018) Guessing from Context Test. Alternatively, the presence or amount of information to aid inferencing may vary across items. Laufer's (1989) vocabulary measure is a good example of this. Her purpose was to determine which words were known in a normal reading passage, and content words were target items, whether they could be inferred from their context in the passage or not.

To our knowledge, the two highest levels of contextualization are used only in meaning-recognition tests. An example of level 3 is Sasao and Webb's (2018) aforementioned diagnostic test of lexical inferencing ability, and examples of level 4 can be found in standardized tests like the TOEFL. As displayed in Figure 2, at level 4, multiple target words can be assessed in a single prompt. This is a good way to balance the provision of extended context with practicality. The absence of more extensive contextualization in form-recognition and form-recall tests is understandable because in these modalities item stems are often in the L1, where single words can readily be understood in isolation. As for meaning-recall, there are potentially good uses for contextualized assessment at levels 3 and 4, but to date such tests are rarely, if ever, used. In the next sections we explore these possible assessment applications by taking a closer look at the two defining characteristics of the CMRT platform introduced in this paper: contextualization in vocabulary testing and meaning-recall assessment.

Figure 2

A Framework of Vocabulary Assessment Item Formats Measuring Form-Meaning Knowledge at Four Levels of Contextualization

	Levels of Contextualization			
	1	**2**	**3**	**4**
Meaning-Recognition	privilege (A) 宮殿 (B) 定型句 (C) 特権 (D) 欠点	It is a privilege. (A) 宮殿 (B) 定型句 (C) 特権 (D) 欠点	Having access to clean water and good food is a privilege that many people around the world do not have. (A) 宮殿 (B) 定型句 (C) 特権 (D) 欠点	Having access to clean water and good food is a (1) privilege that many people around the world do not have. While we (2) satisfy our thirst with a simple turn of the (3) tap, many people face the daily challenge of finding safe water sources. It's crucial for us to recognize this (4) disparity and work towards ensuring everyone's right to these essential resources. 1. (A) 宮殿 (B) 定型句 (C) 特権 (D) 欠点
Form-Recognition	特権 (A) palace (B) expression (C) privilege (D) defect	それは特権です。 (A) palace (B) expression (C) privilege (D) defect		
Meaning-Recall	privilege = _____	It is a privilege. privilege = _____		
Form-Recall	特権 = pr____	It is a pr_____. (特権)		

Note. The example level 2 meaning-recognition item comes from the New General Service List Test (Stoeckel et al., 2018). All other example items are extrapolated from that.

Contextualization in Vocabulary Testing

Test Uses

The amount of context provided in vocabulary test items should reflect the purpose and intended consequences of testing (Read & Chapelle, 2001). Formally, Read and Chapelle distinguish between context-independent and context-dependent vocabulary assessment. In the former, the expected response can be made without reference to context while in the latter, understanding of contextual information in the test item is necessary to answer correctly. Thus, levels 1 and 2 in our framework are context-independent while levels 3 and 4 could be either, depending on whether test items can be answered without comprehending the context provided in the stem. Examples of context-dependent items might be found in a test of lexical inferencing in which the target items are pseudowords whose meanings can only be worked out from the context provided. Let us now consider the possible role of context for several vocabulary testing purposes.

Assessing Lexical Inferencing Ability

Obviously, when we wish to assess the ability to guess words from context, a context-dependent item format is indispensable. Lexical inferencing is an important vocabulary-development strategy (Nation, 2008), so there is utility in diagnostically assessing this skill and helping students become better at it (Nation, 2013).

Testing Isolated Knowledge of the Form-Meaning Link

In contrast, when we wish to measure understanding of the form-meaning link in isolation, context-independent assessment is required so that examinees are unable to employ the separate skill of guessing unknown items from context (Schmitt, 2010). Accordingly, size and levels tests such as the VST (Nation & Beglar, 2007) and VLT (Nation, 1983) are typically context-independent.

Measuring Vocabulary Knowledge for Reading

There are arguments for and against contextualized vocabulary assessment for the receptive skills. In fluent reading, context usually offers relatively little support for understanding word meaning because automaticity in word recognition and meaning retrieval is required to free up cognitive resources for text-level meaning construction (Grabe, 2009). Therefore, it is

sometimes claimed that context-independent tests are better gauges of the lexical understanding typically employed in reading (Cameron, 2002).

However, in coverage-comprehension studies, in which researchers investigate how differences in comprehension correspond with small changes in the percentage of words known in a text, there may a case for assessing knowledge of lexis in the context of a study's reading passage. Word meaning may be understood when assessed in a non-contextualized manner but not when used with a specialized meaning in the passage (Webb, 2021b). Alternatively, a word may be understood in the supportive context of a natural text but not in a discrete point test item. Though research on previously unread text has not found a significant difference in scores on fully-contextualized (i.e., level 4) and non-contextualized vocabulary tests (Henning, 1991), just one study has examined this issue. If researchers wish to measure the precise percentage of words known in a particular text, perhaps context should be considered. Indeed, in previous coverage-comprehension research, both approaches to vocabulary measurement have been used (see Laufer, 1989 and Schmitt et al., 2011).

Testing the Assumptions of the Word Family

A word family consists of a base form (e.g., *use*) together with related inflectional (e.g., *used, uses*) and derivational forms (e.g., *useful, useless*). The precise members of a family depend on the definition used (see Bauer & Nation, 1993), but a general assumption underlying the word family is that when a learner knows the meaning of one member, they should also be able to receptively understand other members when encountered in a meaningful context (Nation, 2015). Thus, contextualized vocabulary assessment may be preferred in studies investigating this assumption of the word family (Laufer et al., 2021). Such an approach might yield different results from research that has assessed word knowledge with no supportive context and found relatively low correspondence between baseword and derivational form knowledge (e.g., Ward & Chuenjundaeng, 2009). This is uncertain, however, because, as discussed below, comparisons of tests with different levels of contextualization have yielded inconsistent findings (Henning, 1991; Laufer, 2023).

Promoting Positive Washback

Washback is the effect that tests have on teaching and learning. Although tests are probably not administered solely for their washback, selection of item format can be influenced by the perceived washback a test has (Read

& Chapelle, 2001). There are divergent views regarding the washback of context in vocabulary test items. When vocabulary items in the TOEFL were changed from discrete point to those embedded in reading passages, it was thought to bring about positive washback in that it would encourage test-takers to learn to deal with vocabulary in communicative contexts (Read & Chapelle, 2001). Similarly, Qian (2008) has stated that non-contextualized vocabulary testing can have negative washback if it encourages the study of words in isolation. Nation (2013), however, disagrees with this view, citing research that the use of word cards and other forms of limited-context study are effective for learning new words (de Groot, 2006; Elgort, 2011).

Research on Contextualization in Tests of Form-Meaning Knowledge

Several studies have directly compared levels of contextualization in assessment of form-meaning knowledge. This research has almost exclusively investigated meaning-recognition item types and paints a moderately favorable picture for the use of increased contextualization in vocabulary assessment.

Of the areas explored, two have not been impacted by differences in contextualization. The first is the correlation between vocabulary and reading test scores. This research has compared vocabulary assessment at contextualization levels 1 and 3 (Qian, 2002), 2 and 4 (Ushiro et al., 2009), and 3 and 4 (Qian, 2008). In each case, the vocabulary-reading correlation did not significantly differ for the compared vocabulary measures. Second, though only levels 3 and 4 have been compared, differences in context have not been found to influence item discrimination as estimated with point-biserial correlations (Qian, 2008). This means that test items employing the compared levels of contextualization did not differ in their capacity to distinguish learners on the basis of vocabulary knowledge.

Research has also identified two areas that have been affected by the level of contextualization in vocabulary items. The first is concurrent validity. In Henning's (1991) aforementioned comparison of meaning-recognition items at each of our four levels of contextualization, level 4 scores correlated most strongly with a criterion vocabulary measure, with the difference between levels 1 and 4 reaching significance. Second, added context may favorably impact test reliability. Henning (1991) found that estimates of internal reliability consistently increased with contextualization across tests with the same number of items. The differences were significant for level 4 in comparison to levels 1 and 2 and nearly significant relative to level 3.

Finally, research has yielded inconsistent results on the impact of changes in contextualization on item difficulty. Henning (1991) assessed the same words across five item types[1] at all four levels of contextualization and found no significant difference in mean scores. On the other hand, Laufer (2023) found a significant difference in scores when testing knowledge of the same pseudowords at three levels of contextualization. She provided learners with the meanings of 22 pseudo-basewords (e.g., *stace*) and then tested their ability to understand derivations of those words (e.g., *stacement*) at contextualization levels 1-3. Mean scores at level 1 were significantly lower than at levels 2 and 3. Perhaps these disparate findings can be explained by differences in the item stems used in the two studies. Whereas Henning's contained only the target word and (at levels 2-4) the context in which it was embedded, each of Laufer's item stems reminded test-takers to consider context, as in the following example:

If *stace* means "to participate," what does *stacement* mean in the following sentence?
Full and active *stacement* in school activities is required of all students.
Stacement means _____.

Another possible explanation is research showing that meaning-recognition formats, like those employed by Henning, mostly measure isolated vocabulary knowledge even when extensive contextualization is used (Ushiro et al., 2009).

Meaning-Recall Vocabulary Assessment

Considerations in Choosing Between Meaning-Recall and Meaning-Recognition

As with levels of contextualization, the type of form-meaning knowledge assessed ought to be guided by the purpose and intended consequences of testing (Schmitt et al., 2020). Because both meaning-recall and meaning-recognition assess receptive lexical knowledge, these two modalities are frequently compared, and decisions regarding which to use are often made by weighing practicality and accuracy. In the following paragraphs, we discuss these two factors together with a third consideration, washback.

Practicality refers to the ease with which tests are designed, administered, and scored (Brown, 2004). Regarding design, meaning-recall tests are clearly more practical owing to the time and expertise needed to write

good distractors for meaning-recognition tests (Rodriguez, 2005). For test administration, however, meaning-recognition is quicker (McLean et al., 2020) because test-takers only select responses rather than translate target-words. Regarding scoring, meaning-recognition is also quicker – indeed, it is instantaneous in computer-administered tests. As previously mentioned, the scoring of meaning-recall tests has become easier with automated marking, but currently novel responses still require human attention. When there are numerous examinees or when results are needed quickly, meaning-recognition tests remain the more practical option. However, for classroom assessment purposes such as achievement tests, and for many research applications, any difference in test practicality may be outweighed by considerations of accuracy and washback.

The accuracy of a language test is based on its capacity to (a) detect knowledge when it is present and (b) detect the absence of knowledge when it is absent. These are referred to as sensitivity and specificity, respectively (Eckes, 2017). It is sometimes claimed that meaning-recognition is more sensitive than meaning-recall (Webb, 2021a), as evidenced by the many studies showing that learners achieve higher scores on meaning-recognition tests (e.g., Kremmel & Schmitt, 2016; Laufer & Goldstein, 2004; Stoeckel et al., 2019; Stoeckel & Sukigara, 2018). However, meaning-recognition tests are influenced by the use of construct-irrelevant test strategies and blind guessing (Gyllstad et al., 2015; McDonald, 2015; McLean, Kramer, & Stewart, 2015), indicating that a portion of the score difference between the two test formats is due to decreased specificity rather than increased sensitivity of the meaning-recognition measure. Hence, some scholars consider meaning-recall to be the more accurate of the two test formats, at least as a measure of the lexical knowledge used in reading (Kremmel & Schmitt, 2016; McLean, 2021; Schmitt, 2019; Stoeckel et al., 2021). Perhaps an indication of how widely this second view is held, meaning-recall tests are commonly employed as criterion measures in validation studies of meaning-recognition tests (e.g., Kremmel & Schmitt, 2016; Stoeckel et al., 2019; Webb et al., 2017), but rarely, if ever, the other way around.

An overlooked factor favoring the use of meaning-recall is washback. Compared to meaning-recognition, meaning-recall is a stronger form of lexical knowledge (Laufer & Goldstein, 2004) that correlates more strongly with receptive language ability (McLean et al., 2020; Zhang & Zhang, 2022). There is, therefore, good reason to encourage learners to master vocabulary to the level of meaning-recall, and perhaps meaning-recall vocabulary assessment would have that effect. While this may be difficult to enact in large-

scale educational testing, it should be considered for smaller-scale uses like classroom progress tests and quizzes.

Research Comparing Meaning-Recall and Meaning-Recognition

The studies comparing meaning-recall and meaning-recognition vocabulary measures have produced relatively consistent findings. First, meaning-recall has better internal reliability. Though statistical significance has gone unreported, this has been found in each study that reported the reliability of both measures and that assessed the same words under the two item formats (McLean et al., 2020; Stoeckel et al., 2019; Stoeckel & Sukigara, 2018). Second, meaning-recall appears to be a better predictor of reading comprehension. Although some studies have lacked statistical significance (Jeon & Yamashita, 2014; Laufer & Aviad-Levitzky, 2007), others have found a clear contrast (Zhang & Zhang, 2022) with large effect sizes (McLean, et al., 2020). Third, as previously stated, meaning-recall tests require more time to administer. Note, however, that this difference has reached statistical significance for multiple-choice but not matching formats (McLean et al., 2020). Fourth, when the same words are tested, meaning-recall is more difficult than meaning-recognition (Kremmel & Schmitt, 2016; Laufer & Goldstein, 2004; Stoeckel et al., 2019; Stoeckel & Sukigara, 2018). Related research has indicated that reasons for this difference include random guessing and use of not only construct-relevant but also construct-irrelevant test strategies on the meaning-recognition test (Gyllstad et al., 2015; McDonald, 2015; McLean, Kramer, & Stewart, 2015).

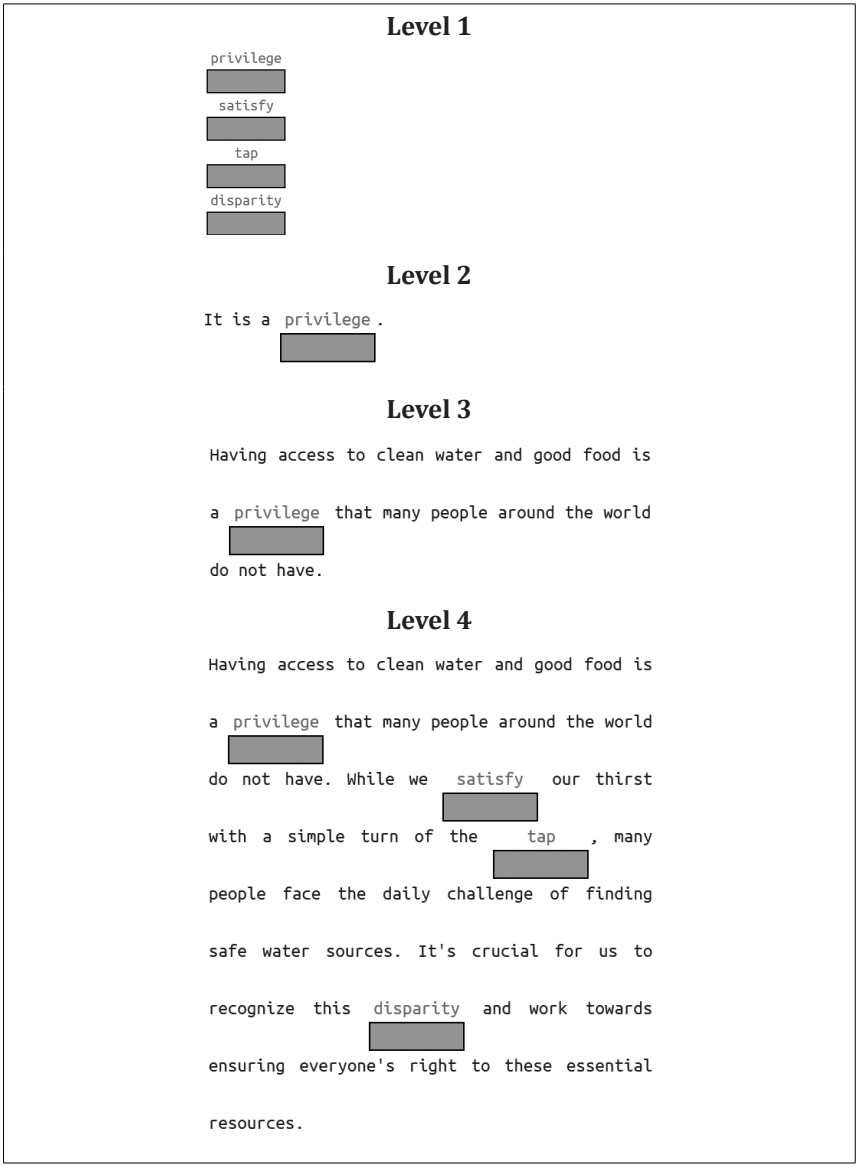
In sum, while meaning-recall tests may be somewhat less practical, they are more accurate, a better predictor of receptive language ability, and – we would argue – more likely to produce beneficial washback. In the final section, we provide a detailed description of meaning-recall assessment on the CMRT platform.

The Contextualized Meaning-Recall Testing Platform

The CMRT platform (https://cmrt.vocableveltest.org) can be used by L2 teachers and researchers to expeditiously create, administer, and mark contextualized meaning-recall vocabulary tests. For test creation, the platform is set up so that anyone with a Gmail account can create and administer tests. The test owner simply inputs a text and selects target words or phrases. This produces a test in which learners see discourse with boxes under the target items to input their responses. Although this format enables examinees to

Figure 3

Example Test Items at Four Levels of Contextualization on the CMRT Platform



consider broad context when discerning word meaning, the platform can be used to assess vocabulary at all four levels of contextualization, as shown in Figure 3. There is also a place for test creators to input instructions. This allows for the elicitation of different kinds of responses (e.g., L1 translations; L2 synonyms, definitions, explanations) depending on the learner group and testing purpose. To administer a test, the test creator needs only to share the test URL with test-takers. Examinees do not need to register as members of the site. To deter students from getting outside help, there is also an option to first warn test-takers and then automatically end the test if navigation away from the test app is detected. Regarding privacy, the platform is hosted on a secure cloud server, and if an added layer of protection is desired, students could be asked to use pseudonyms or examinee codes instead of their actual names.

After test administration, either the owner or one or more assigned raters mark the test. Raters access a list of distinct responses for each item and rate them (as correct, incorrect, or partially correct) without seeing the judgments of other raters (Figure 4). When marking is complete, all judgments can be viewed and final decisions recorded for discrepant ratings. Additionally, these decisions can be saved for future test use, reducing the burden of marking in subsequent test administrations. The test owner can also view and download tables of responses and points earned to each test item for every examinee (Figures 5 and 6).

Figure 4

Rater Interface for the CMRT Platform

Note. Each target word is displayed followed by (a) the context in which it appears in the passage (in small grey font) and (b) a list of each distinct response (in pink boxes), which can be rated as incorrect, partially correct, or correct. There is also a place for comments for ambiguous or otherwise noteworthy responses.

Figure 5
Online View of Test Responses

Responses

Answers Scores

Name	Start Time	Finish Time	Total Time	kind	heart	healthy	aerobics	hard	routine	powerful	judge	Participants	backgrounds	pair	spread	championship	International Olympic Committee	official
Mike M.	2022-08-20 10:30:26	2022-08-20 10:36:20	00:05:54	やさしい	心	健康	エアロビクス	激しく	ルーティーン	パワフル	審判	参加者	バックグラウンド	ペア	広がった	チャンピョン	国際オリンピック委員会	公式の
ky	2022-08-20 10:39:01	2022-08-20 10:43:05	00:04:04	種類	心臓	健康	エアロビクス	一生懸命	ルーティン	強い	審判	参加者	背景、経験	ペア、一足	広がった	大会	国際オリンピック委員会	公式の
Kodani	2022-08-20 12:24:29	2022-08-20 12:33:28	00:08:59	一種の	心臓	健康に	エアロビ	がんばる	ノルマ	力強く	裁判官	関心を持つ人	豊かな経験	*訳さない	ひろがった	選手権	国際オリンピック委員会	公式の
Urara	2022-08-20 14:15:19	2022-08-20 14:20:07	00:04:48	種類	心臓	健康	エアロビクス	がんばる	ルーチン	力強い	審査員	競技者	経験	一足	拡がる	大会	国際オリンピック協会	正式
tomoko	2022-08-20 15:45:53	2022-08-20 15:48:56	00:03:03	種類	心臓	健康な	エアロビクス	熱心に	ルーティーン	力強い	審判	参加者	背景	ひと揃い	広まる	競技会	国際オリンピック委員会	公式な

Note. Responses can be viewed online and downloaded as a csv file.

Figure 6

Online View of Ratings for Test Responses

Responses

Answers Scores

Name	Start Time	Finish Time	Total Time	kind	heart	healthy	aerobics	hard	routine	powerful	judge	Participants	backgrounds	pair	spread	championship	International Olympic Committee	official
Mike M.	2022-08-20 10:30:26	2022-08-20 10:36:20	00:05:54	0	0.5	0.5	1	1	1		1	1	0.5				1	
ky	2022-08-20 10:39:01	2022-08-20 10:43:05	00:04:04	1	1	0.5	1	1	1	1	1	1	1				1	
Kodani	2022-08-20 12:24:29	2022-08-20 12:33:28	00:08:59	1	1	1	1	1	0	1	1	0	1				1	
Urara	2022-08-20 14:15:19	2022-08-20 14:20:07	00:04:48	1	1	0.5	1	1	1	1	1	1	1					
tomoko	2022-08-20 15:45:53	2022-08-20 15:48:56	00:03:03	1	1	1	1	1	1	1	1	1	0.5				1	

Note. Blank cells depict unrated items; (0 = incorrect, 0.5 = partially correct, 1 = correct). Ratings can be viewed online or downloaded as a csv file.

Although the CMRT platform can substantially reduce the amount of time needed for meaning-recall assessment at several levels of contextualization, it has limitations. As a practical matter, because the platform was developed with grant funding for a particular research project, users should expect limited technical support. Also, tests can be rendered only in left-justified, plain text. Formatting options like bold or italicized font, underlining, centering, and auto-numbering are unavailable. Moreover, although test results are downloadable as csv files, the platform is not integrated into any existing learning management system. Concerning construct validity, it should not be assumed that tests developed and administered on the CMRT platform are valid for particular purposes. For low-stakes, classroom use, teachers can probably apply the same principles they use for other forms of assessment. For higher-stakes testing or research, however, validation evidence must be gathered to support the use of CMRTs for specific uses and score interpretations (Messick, 1995). Moreover, when multiple target words occur in close proximity in a single passage, the assumption of local independence would need to be checked (de Ayala, 2009).

Conclusions and Future Directions

The CMRT platform has potential uses in both research and pedagogy. In research, it could be employed to compare meaning-recall vocabulary assessment at different levels of contextualization, paralleling the above-mentioned studies on meaning-recognition formats. Additionally, inquiry comparing meaning-recognition and meaning-recall might be extended to more systematic investigation of the role of context. Coverage-comprehension studies could also be conducted with vocabulary measured at the meaning-recall level (see McLean, 2021). For teachers, the CMRT platform may be used for practicing and assessing lexical inferencing ability, where it is advantageous to provide learners with a continuous text with actual target words rather than blanks or pseudowords (Nation, 2013). Additionally, since learners tend to achieve higher scores when vocabulary is tested in the same context in which it is learned (Watanabe, 1997), the platform could be used as a sensitive measure of newly acquired vocabulary from class texts. Finally, its use in the classroom may promote positive washback if it encourages students to study and learn words more deeply than they would with meaning-recognition achievement tests.

In closing, we would like to address a reviewer's intriguing question regarding the possible use of AI to judge test responses. With the recent, rapid development of generative AI capabilities, it would be interesting to see

whether this is feasible. A foreseeable challenge is that there are multiple ways to express word meaning that go well beyond dictionary definitions and one-to-one translations. Humans can achieve high levels of inter-rater reliability for such responses, even when using rather nuanced marking criteria. Whether AI could match humans in this regard is an interesting question for researchers to explore.

Notes

1. Henning (1991) compared eight item types in total, but only five strictly assessed form-meaning knowledge. Of these, there was one item at each of contextualization levels 1, 2, and 4, and two items at level 3.

Acknowledgements

This study was partially funded by grant number JP 22K00793 from the Japan Society for the Promotion of Science.

Tim Stoeckel teaches English and applied linguistics at the University of Niigata Prefecture. His primary interests are in L2 reading, vocabulary learning and assessment, and the relationship between lexical knowledge and reading comprehension.

Stuart McLean holds doctorates in medicine and TESOL, and a PGCE. He publishes research related to vocabulary, assessment, and comprehension (https://scholar.google.com/citations?hl=en&user=yL_1NxsAAAAJ). His online self-marking meaning-recall (reading or listening) levels tests are at www.vocableveltest.org. Teachers can download automatically marked responses, typed responses, and the time taken to complete responses.

Paul Raine (MA TEFL/TESL) is an award-winning teacher, presenter, author, and developer. He has published numerous research articles on the teaching and learning of English as a second language, and is particularly interested in Computer Assisted Language Learning (CALL). He currently lives and teaches English in Osaka, Japan.

Hung Tan Ha is a PhD student in Applied Linguistics at Victoria University of Wellington. His research interests involve L2 reading comprehension and vocabulary assessment.

Nam Thi Phuong Ho teaches English Language at the School of Foreign Languages - University of Economics Ho Chi Minh City. Her research inter-

est is English for Specific Purposes, TESOL Methodologies, Quantitative and Qualitative research designs, and Vocabulary teaching and learning.

Young Ae Kim is an instructor at Kindai University and Kyoto Seika University and holds an MA from Kansai University. She is working on research related to test item format type, word counting units, listening and word difficulty, as well as making the items for the tests available at www.vocableveltest.org.

References

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography, 6*(4), 253–279. https://doi.org/10.1093/ijl/6.4.253

Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Pearson Longman.

Cameron, L. (2002). Measuring vocabulary size in English as an additional language. *Language Teaching Research, 6*(2), 145–173. https://doi.org/10.1191/1362168802lr103oa

de Ayala, R. J. (2009). *The theory and practice of item response theory.* The Guilford Press.

de Groot, A. M. (2006). Effects of stimulus characteristics and background music on foreign language vocabulary learning and forgetting. *Language Learning, 56*(3), 463–506. https://doi.org/10.1111/j.1467-9922.2006.00374.x

Eckes, T. (2017). Setting cut scores on an EFL placement test using the prototype group method: A receiver operating characteristic (ROC) analysis. *Language Testing, 34*(3), 383–411. https://doi.org/10.1177/0265532216672703

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning, 61*(2), 367–413. https://doi.org/10.1111/j.1467-9922.2010.00613.x

Grabe, W. (2009). *Reading in a second language*. Cambridge University Press.
Gyllstad, H., Vilkaitė, L., & Schmitt, N. (2015). Assessing vocabulary size through multiple-choice formats: Issues with guessing and sampling rates. *ITL - International Journal of Applied Linguistics, 166*, 278–306. https://doi.org/10.1075/itl.166.2.04gyl

Henning, G. (1991). *A study of the effects of contextualization and familiarization on responses to the TOEFL vocabulary test items.* Educational Testing Service. https://doi.org/10.1002/j.2333-8504.1991.tb01390.x

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning, 64*, 160–212. https://doi.org/10.1111/lang.12034

Jiang, N. (2002). Form–meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition, 24*(4), 617–637. https://doi.org/10.1017/S0272263102004047

Kremmel, B., & Schmitt, N. (2016). Interpreting vocabulary test scores: What do various item formats tell us about learners' ability to employ words? *Language Assessment Quarterly, 13*(4), 377–392. https://doi.org/10.1080/15434303.2016.1237516

Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316–323). Multilingual Matters.

Laufer, B. (2023). Understanding L2-derived words in context: Is complete receptive morphological knowledge necessary? *Studies in Second Language Acquisition*, 1-14. https://doi.org/10.1017/S0272263123000219

Laufer, B., & Aviad-Levitzky, T. (2017). What type of vocabulary knowledge predicts reading comprehension: Word meaning-recall or word meaning-recognition? *The Modern Language Journal, 101*, 729–741. https://doi.org/10.1111/modl.12431

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning, 54*(3), 399–436. https://doi.org/10.1111/j.0023-8333.2004.00260.x

Laufer, B., Webb, S., Kim, S. K., & Yohanan, B. (2021). How well do learners know derived words in a second language? The effect of proficiency, word frequency and type of affix. *ITL-International Journal of Applied Linguistics, 172*(2), 229–258. https://doi.org/10.1075/itl.20020.lau

McDonald, K. (2015). The potential impact of guessing on monolingual and bilingual versions of the vocabulary size test. *Osaka JALT Journal, 2*, 44–61. https://bit.ly/3PoRniq

McLean, S. (2021). The coverage comprehension model, its importance to pedagogy and research, and threats to the validity with which it is operationalized. *Reading in a Foreign Language, 33*, 126–140. https://nflrc.hawaii.edu/rfl/item/528

McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research, 19*(6), 741–760. https://doi.org/10.1177/1362168814567889

McLean, S., Kramer, B., & Stewart, J. (2015). An empirical examination of the effect of guessing on vocabulary size test scores. *Vocabulary Learning and Instruction, 4*, 26–35. https://doi.org/10.7820/vli.v04.1.mclean.et.al

McLean, S., Raine, P., Pinchbeck, G., Huston, L., Kim, Y. A., Nishiyama, S., & Ueno, S. (2021). The internal consistency and accuracy of automatically scored written receptive meaning-recall data: A preliminary study. *Vocabulary Learning and Instruction, 10*(2), 64–81. https://doi.org/10.7820/vli.v10.2.mclean

McLean, S., Stewart, J., & Batty, A. O. (2020). Predicting L2 reading proficiency with modalities of vocabulary knowledge: A bootstrapping approach. *Language Testing, 37*, 389–411. https://doi.org/10.1177/0265532219898380

Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749. https://doi.org/10.1037//0003-066X.50.9.741

Nation, I. S. P. (1983). Teaching and testing vocabulary. *Guidelines, 5*, 12-25.

Nation, I. S. P. (2008). *Teaching vocabulary: Strategies and techniques*. Heinle ELT.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139858656

Nation, P. (2015). Which words do you need? In J. R. Taylor (Ed.), *The Oxford handbook of the word* (pp. 568–581). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199641604.013.016

Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*, 9–13. https://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*(1), 86–99. https://doi.org/10.1177/0033688210390264

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, *52*(3), 513–536. https://doi.org/10.1111/1467-9922.00193

Qian, D. D. (2008). From single words to passages: Contextual effects on predictive power of vocabulary measures for assessing reading performance. *Language Assessment Quarterly, 5*(1), 1–19. https://doi.org/10.1080/15434300701776138

Read, J. (2000). *Assessing vocabulary*. Cambridge University Press. https://doi.org/10.1017/CBO9780511732942

Read, J., & Chapelle, C. A. (2001). A framework for second language vocabulary assessment. *Language Testing, 18*(1), 1–32. https://doi.org/10.1177/026553220101800101

Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice, 24*(2), 3–13. https://doi.org/10.1111/j.1745-3992.2005.00006.x

Sasao, Y., & Webb, S. (2018). The guessing from context test. *ITL-International Journal of Applied Linguistics, 169*(1), 115–141. https://doi.org/10.1075/itl.00009.sas

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

Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching, 52*(2), 261–274. https://doi.org/10.1017/S0261444819000053

Schmitt, N., Dunn, K., O'Sullivan, B., Anthony, L., & Kremmel, B. (2021). Introducing knowledge-based vocabulary lists (KVL). *TESOL Journal*, *12*(4), e622. https://doi.org/10.1002/tesj.622

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95*(1), 26–43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Schmitt, N., Nation, P., & Kremmel, B. (2020). Moving the field of vocabulary assessment forward: The need for more rigorous test development and validation. *Language Teaching, 53*(1), 109–120. https://doi.org/10.1017/S0261444819000326

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

Stewart, J., Gyllstad, H., Nicklin, C., & McLean, S. (2023). Establishing meaning recall and meaning recognition vocabulary knowledge as distinct psychometric constructs in relation to reading proficiency. *Language Testing.* Advance online publication. https://doi.org/10.1177/02655322231162853

Stewart, J., Stoeckel, T., McLean, S., Nation, P., & Pinchbeck, G. G. (2021). What the research shows about written receptive vocabulary testing: A reply to Webb. *Studies in Second Language Acquisition, 43*(2), 462–471. https://doi.org/10.1017/S0272263121000437

Stoeckel, T., Ishii, T., & Bennett, P. (2018). A Japanese-English bilingual version of the New General Service List Test. *JALT Journal, 40*(1), 5-21. https://doi.org/10.37546/JALTJJ40.1-1

Stoeckel, T., McLean, S., & Nation, P. (2021). Limitations of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition, 43*(1), 181–203. https://doi.org/10.1017/S027226312000025X

Stoeckel, T., Stewart, J., McLean, S., Ishii, T., Kramer, B., & Matsumoto, Y. (2019). The relationship of four variants of the Vocabulary Size Test to a criterion measure of meaning-recall vocabulary knowledge. *System, 87*, 102161. https://doi.org/10.1016/j.system.2019.102161

Stoeckel, T., & Sukigara, T. (2018). A serial multiple-choice format designed to reduce overestimation of meaning-recall knowledge on the Vocabulary Size Test. *TESOL Quarterly, 52*, 1050–1062. https://doi.org/10.1002/tesq.429

Ushiro, Y., Hirai, N., Hoshino, Y., Nakagawa, C., Kai, A., Ide, M., Oki, T., Suzuki, S., Terada, Y., Takaki, S., & Shimizu, H. (2009). Comparing effects of two types of vocabulary knowledge on six question types in reading tests among Japanese EFL learners. *JLTA Journal Kiyo, 12*, 104–115. https://doi.org/10.20622/jltaj.12.0_104

Ward, J., & Chuenjundaeng, J. (2009). Suffix knowledge: Acquisition and applications. *System, 37*, 461–469. https://doi.org/10.1016/j.system.2009.01.004

Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition, 19*, 287–307. https://doi.org/10.1017/S027226319700301X

Webb, S. (2021a). A different perspective on the limitations of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition, 43*(2), 454–461. https://doi.org/10.1017/S0272263121000449

Webb, S. (2021b). Lexical coverage and lexical profiling: What we know, what we don't know, and what needs to be examined. *Reading in a Foreign Language, 33*(2), 278–293. http://hdl.handle.net/10125/67407

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL-International Journal of Applied Linguistics, 168*(1), 33–69. https://doi.org/10.1075/itl.168.1.02web

Zhang, S., & Zhang, X. (2022). The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis. *Language Teaching Research, 26*(4), 696–725. https://doi.org/10.1177/1362168820913998

Expositions

Word Parts and Language Learning

Paul Nation
LALS, Victoria University of Wellington

Research with Japanese learners of English (Brown et al., 2020) has shown that Japanese learners of English typically have inadequate knowledge of the word parts of English. Research with other groups of learners also tend to show a wide range of degrees of knowledge of word parts (Laufer & Cobb, 2020; Sasao & Webb, 2017). This lack of knowledge has important implications for vocabulary size and reading in English because a very large proportion of Level 6 word families (Bauer & Nation, 1993) include many family members that contain prefixes and suffixes. Related to this, a large proportion of the tokens in any text are complex words (Nation, 2022), that is, words containing affixes. The study of word parts is called morphology, and occasionally I will use this term and its adjective, morphological, when referring to word parts, largely because first language research on word parts uses the terms morphological awareness and morphological knowledge. Word part knowledge is closely related to vocabulary size, and vocabulary size is closely related to proficiency level. This article looks at this relationship and tries to relate these three factors using the CEFR (Common European Framework of Reference for Languages) levels.

Brown et al. (2020)によると、日本人英語学習者は接辞や語根などの単語構成要素(word parts)に関する知識が不十分であるとされている。他の同分野における研究(Laufer & Cobb, 2020; Sasao & Webb, 2017)でも、単語構成要素に関する知識は学習者によって大きく異なることが示されている。こうした知識が欠けていることは、英語における語彙サイズや読む能力に大きな影響を与えることになる。なぜなら、Bauer and Nation (1993)の分類によるレベル6の単語の多くが、接頭辞や接尾辞を含んでいるからである。同様に、どのような文書でもかなりの部分が、接辞を含む複合語であるからだ(Nation, 2020)。こうした単語構成要素に関する学問を「形態論」という。単語構成要素について言及する場合、私はしばしばこの「形態論」やその形容詞形である「形態論的」という言葉を使う。母語における単語構成要素の研究では、「形態論的認識」

https://doi.org/10.37546/JALTJJ45.2-3

JALT Journal, Vol. 45, No. 2, November 2023

や「形態素の知識」という言葉を使うからである。単語構成要素に関する知識は語彙サイズと密接な関係があり、語彙サイズは熟達度と密接な関係がある。この論文ではこうした関係を吟味し、CEFR (Common European Framework of Reference for Language) レベルを用いて、この3つの要素を結び付けることを試みる。

Keywords: affixes; morphology; stems; word parts

The Importance of Knowing Word Parts

Research with native-speakers of English has shown that morphological knowledge has a strong relationship with vocabulary knowledge and, through vocabulary knowledge, a relationship with reading comprehension (Levesque et al., 2019). This causal chain exists because knowledge of word parts helps learners understand and remember complex words, and knowing words is an obvious and well-proven prerequisite for understanding a text.

Knowing word parts also affects the opportunities for learning through repeated meetings with words. Each time a word is met through reading or listening, the greater the chance that it will be remembered. If learners are able to see that word forms that share the same word stem are related to each other, then the occurrence of any one of those word forms becomes a repetition of the whole family. For example, the word family *adequate* contains the family members *adequate*, *adequately*, *inadequate*, *inadequately*. Knowledge of the very frequent affixes *in-* meaning 'not', and *-ly* making an adverb allows the occurrence of any one of these four family members to be effectively a meeting with the same word family.

Word parts can also be used for *morphological problem-solving* (Anglin, 1993) where learners work out the meaning of a new word. While this will work quite often, especially where the stem is also a free-standing word, it is a somewhat dangerous procedure in that this inferencing may lead to incorrect solutions. It is much safer to see word part knowledge as primarily a way of helping words stick in memory. If morphological problem-solving is done, then checking on the solution is a useful step in the procedure.

What Does Word Part Knowledge Involve?

There has been a large amount of research on morphological knowledge, including morphological awareness, with native speakers, and particularly native speakers of English. Morphological knowledge is seen as being made up of four related but distinct skills (Goodwin et al., 2021). These are morphological awareness (the ability to reflect upon and manipulate morphemes),

morphological-syntactic knowledge (knowing how morphemes switch part of speech), morphological-semantic knowledge (using the meanings of morphemes to work out the meanings of words), and morphological-orthographic/phonological knowledge (having the ability to spell complex words, and the ability to pronounce complex words). These skills each draw on particular kinds of morphological knowledge, as explained below:

1. Morphological awareness includes the general understanding that many English words have parts and that the meanings of these parts typically contribute to the meaning of the whole word. As a technical term, morphological awareness is not always clearly defined and in research is often measured by seeing if learners can see word parts in words. It may be best used as a blanket term for all the morphological skills, or used much more narrowly somewhat in parallel with phonological awareness to refer to the general understanding that words may have parts and that these parts can contribute to the meaning of the word. At present, its use in first language studies and how it is operationalised do not clearly distinguish it from other aspects of morphological knowledge.
2. Morphological-syntactic knowledge involves knowing how suffixes can affect part-of-speech. In research, it has been measured by getting the learners to choose from different given grammatical forms of the same word or adapt a given word to the right grammatical form to fit a given context. For a large proportion of the most useful suffixes, the part-of-speech marking is the main function of the suffixes, and only a few of the suffixes at Bauer and Nation Levels 3 and 4 have a strong lexical meaning, for example *-able, -less, -like, -ess, -ful, -ism,* and *-ist*.
3. Morphological-semantic knowledge involves being able to use the meanings and functions of word parts to work out the meanings of words. In research, it has been measured by choosing the meaning for a complex word in context. The first six levels of the Bauer and Nation scale involve stems that are free forms, that is, the stem can be a word in its own right. At Level 7 of the Bauer and Nation scale, stems may be bound forms which cannot stand as words in their own right. So, at Level 6 of Bauer and Nation, *appropriacy* is not a member of the *appropriate* word family because it does not obviously contain the free stem *appropriate*.
4. Morphological-orthographic/phonological knowledge involves having the ability to spell complex words, and the ability to pronounce complex words. In research, it has been measured by getting learners to write

complex words that they hear and to pronounce given complex words.

While these distinctions may be useful when looking at what makes up morphological knowledge, learning word parts and learning how to use them in reading and listening requires a different kind of focus.

How Are Word Parts Learned?

Word part learning, like vocabulary learning and most other learning, requires three basic closely related conditions, namely (1) a focus on what is to be learned, (2) quantity of attention including repetition, and (3) quality of attention. If we just focus on deliberately learning the meaning of word parts, for example that *un-* means 'not', then the technique of word card (also known as flash card) learning is a very suitable way of quickly learning a relatively small number of very useful affixes. In this technique, an affix is written on one side of the card, and its meaning and grammatical function (either in English or the L1) is written on the other side. A flash card app may also be used. Having a known word containing the affix on the same side of the card as the affix may help learning in that it provides a quality element to the learning in relating new knowledge to old knowledge. This is worth investigating.

Word part learning occurs not only through deliberate learning, but also incidental learning while reading and listening. Such incidental learning probably accounts for how the high frequency, regular, productive affixes are learned by native speakers. Such incidental learning is likely to be rather slow and gradual because affixes are much less obvious than words. Nonetheless, it is an important source of learning and would be supported and boosted by the deliberate study of affixes.

Surveys of the attention given to word parts in EFL textbooks and by EFL teachers (Dang & Li, 2022) show that the small amount of attention given is neither systematic nor well principled. That is, the selection of word parts tends to be opportunistic rather than based on criteria such as frequency, regularity, and productivity. What attention is given tends to be occasional rather than repeated and systematic. Research with teachers of native-speakers shows a similar lack of attention to word parts in teacher training (Gellert et al, 2020; Mulder et al, 2022), and teachers of native speakers tend to have poor knowledge of word parts (Washburn & Mulcahy, 2019).

Word part learning should not involve a large investment of time and effort, because at any level of proficiency only a small number of affixes are

involved (typically around 12 affixes at any one time), and if the words in the word cutting practice activities to apply them are well chosen, then some very useful learning can occur in a short time (see the Word Parts section of Paul Nation's web resources site or Nation & Bauer, 2023 for ready-made material with answers). It would be well worthwhile conducting a simple feasibility study with learners using Level 3 or Level 4 affixes in the Bauer and Nation (1993) scale to see how much time is involved in reaching a reasonable degree of success in learning the affixes and being able to recognise them in words. It is likely that less than an hour in total spread over four or five lessons may be enough, but this remains to be seen.

Word Parts, Vocabulary Size, and Proficiency Level

There has been considerable debate over the appropriate size of word family to use when making words lists to guide teaching, to make vocabulary size tests, and to use in text coverage studies (McLean, 2018, 2021). Much of one issue of the journal *SSLA* (Volume 43, Issue 5, 2021) consisted of papers debating the issue. This debate centres on the relationship between learners' knowledge of word parts and vocabulary size. As this paper shows, the amount of learning of word parts required to get learners to operate with reasonably sized word families is not great and would not require much classroom time. In Appendix 1, I have indicated what I consider to be useful word part and vocabulary size goals at a range of language proficiency levels. The CEFR levels were chosen because they are widely used in Europe and by European publishers and have influence outside Europe. My suggestions of levels of vocabulary size and word part knowledge are based purely on my own informed judgement, using text coverage figures, the boundary between high and mid-frequency words, and the vocabulary sizes of native-speaking adults. Other researchers who have tried to relate vocabulary size to proficiency levels (Milton & Alexiou, 2009, 2020) have suggested different sizes, although there are some points of agreement.

While there is debate about appropriate word family sizes, I think there is largely agreement about some aspects of word part knowledge. These points of agreement are as follows:

1. As proficiency develops, vocabulary size and word part knowledge (which are part of proficiency) also grow.
2. Vocabulary size and word part knowledge grow in roughly predictable sequences, with frequency of occurrence being the major influencing factor.

3. Receptive knowledge develops before productive knowledge.
4. A small amount of deliberate study of word parts can have a marked effect on knowledge of word families.

The vocabulary size levels and the word part knowledge levels in Appendix 1 represent goals. They are not necessarily a description of learners' current achievement.

What Word Parts Should be Learned at What Level of Proficiency?

Learners need to work on the inflectional affixes at the beginner level of proficiency. Inflections do not change the part of speech of a word, but they indicate time, person, plurality, possession, comparative, and superlative. English has eight inflections (each followed by an example): *-s* [plural] (books), *-s* [third person *-s*] (laughs), *-ed* [past tense] (watched), *-ed* [past participle] (discussed, spoken), *-ing* [present participle] (walking), *-er* [comparative] (bigger), *-est* [superlative] (smallest), and *-'s* [possessive] (Fred's). For more detailed information on the affixes of English, see the book by Bauer and Nation (2020).

Word part learning and analysis involving derivational affixes need not begin until learners know the most frequent 1000 words of English and their inflections. The first group of derivational affixes to learn are the most frequent (that is, they occur in the greatest number of different words), the most regular in form and meaning, and the most productive (they are still used to create new words; Bauer & Nation, 1993, Level 3). These are: -able, -er (as in *singer*), -ish, -less, -like, -ly, -ness, -th (as in the ordinal numbers *fourth, fifth, sixth*, etc.), -y, non-, and un-, which is a total of eleven affixes. Learners should work on learning these while they are learning the second 1000 words of English. Some of these could be learned earlier, but mostly the study of derivational word parts is most usefully begun when learners are at the low intermediate level (Ur, 2022). As well as deliberately learning the meaning and grammatical function of word parts, learners should practice recognizing them in words. Here is a group of words taken from Paul Nation's resources site that learners could use to gain practice in cutting words into parts. In this group, only Level 3 affixes from Bauer and Nation (1993) are used. There are many other such lists on the web site at various levels that can be used for practice.

un/reason/able, disposable, doubtless, driver, easily, eighth, eleventh, employer, emptiness, endless, enjoyable, especially, essentially, exactly, expressionless, fairness, farmer, fearless, feverish, finally, fitness, freezer, frequently, friendliness, funny, generally, girlish, nonstandard, goodness, grassy, wealthy, greatness, greenish, nonprofit, greyish, hairy, happiness, unfortunately, headless, healthy, heartless, heater, hellish, helpless, homeless, nonsense, nonspecific, honorable, hopeless, humorless, hundredth, hungry, nonsmoking

Learners should work on learning the eleven Level 4 affixes (-al, -ation, -ess, -ful, -ism, -ist, -ity, -ize, -ment, -ous, in-) when they know around 2000 to 3000 words. Level 5 affixes and beyond should be learned when learners are at the high intermediate and advanced levels of proficiency. That is, they are well into learning the mid-frequency words (Nation, 2022, Chapter 1), with at least a knowledge of the 4000 to 5000 most frequent word families. Teachers can measure their learners' vocabulary sizes by using one of the recent Vocabulary Levels Tests on Paul Nation's resources site.

Learning to Deal with Word Parts

Morphological knowledge can be receptive (as used in listening and reading) and productive (as used in speaking and writing). Receptive morphological knowledge involves being able to (1) recognize the forms of word parts and recall their meaning and/or grammatical function, (2) recognize word parts in words (at an advanced level this involves dealing with irregularities of spoken and written form and function), and (3) work out how the meanings of the parts of a word contribute to the meaning of the whole word. It is worth learners deliberately learning the meanings and functions of the small number of high frequency word parts (Levels 3 and 4) at the appropriate stage in their vocabulary development. It is also worthwhile doing several small amounts of spaced practice of cutting of words into parts, both as a focused word cutting activity and as a part of intensive reading. Similarly, there should be several spaced sessions of applying the word part strategy which involves cutting a word with familiar parts into parts and explaining the meaning of the word using the meanings of the parts. Learners should look up the meaning or be given the meaning before cutting the word into parts. They also have to adapt the meaning to include the meanings of the parts. So, the strategy is not a guessing strategy but is

instead a mnemonic strategy. Readers may consult Nation (2022) for more details on the word part strategy. The below words are taken from the lists of words for practice in Nation and Bauer (2023).

Immeasurable: Something which is immeasurable cannot be measured because it is too large.
Expectation: Expectation is a noun made from expect.
Nonexistent: Something which is nonexistent (adjective) does not exist.

Notice that each explanation of the word contains a deliberate mention of the meaning and/or the grammatical function of the affixes, as well as the stem.

For many learners of English as a foreign language, receptive knowledge of most derivational word parts is sufficient. Productive knowledge involves being able to (1) recall the forms of word parts to express a meaning or grammatical function and (2) use word parts to make words.

The learning activities that are most suitable for receptive knowledge include learning the form, meaning, and function of isolated word parts through the teacher giving them some deliberate teaching and through using flash cards, recognizing known parts in words by using the word cutting activity, and learning and using the word part strategy by practising explaining the meaning of a word using the meanings of its parts.

Productive knowledge of word parts is a skill suited to advanced learners and is probably best left to incidental learning although this kind of knowledge can be tested through activities where a blank in a context sentence needs to be completed with a given stem that requires an affix to fit -- *He had a feeling of great _____ (happy)*.

Using Useful Word Stems

So far, we have looked at derivational affixes. If we move from morphology to etymology, it is also worthwhile for advanced learners to learn a few useful word stems as a way of helping words stick in memory. This memory trick of relating a part of the word form and its meaning to the whole word is similar to the keyword technique, which has been very thoroughly researched and shown to have very positive effects for vocabulary learning (see Nation 2022 pp. 423-429 for a review of keyword research). Wei & Nation (2013) provide a list of 25 stems that each occur with roughly the same meaning in at least nine other different word families, and another list of words that

occur in at least five other families. The most frequent stems include *spec*, *pos*, *vers*, *vent*, *ceive* as in *inspect*, *composition*, *reverse*, *prevent*, and *deceive*. The study of word stems is interesting and helpful, but it is clearly an activity for advanced learners with vocabulary sizes of well over 4000 words.

Investigating the Learning of Word Parts

There has been no case study research on the effectiveness of the deliberate learning of word parts on the analysis of transparent complex words. There has also been no case study research or experimental research on the effectiveness of the word cutting activity for learning to recognize affixes in words.

For native-speakers of English, the most easily analysed complex words are those where the word stem is much more frequent than the derived form. Highly frequent derived forms, like *computer*, *government*, *dirty*, and *probably*, which occur more frequently than their stem, are often not seen as derived forms but as independent words (Hay, 2001). In some notable cases, such as *business*, these frequent derived forms take on a life of their own with their meaning becoming less accessible through their parts. This suggests that when EFL learners work on analyzing words, it may be more effective to work with lower frequency family members of high frequency word families, rather than with very frequent derived forms.

Paul Nation is Emeritus Professor of Applied Linguistics in the School of Linguistics and Applied Language Studies at Victoria University of Wellington, New Zealand. His latest books include *Teaching Extensive Reading in Another Language* (2020) (with Rob Waring) and *Measuring Native-speaker Vocabulary Size* (with Averil Coxhead), and second editions of *Language Curriculum Design* (2020), *Teaching ESL/EFL Listening and Speaking* (2020), *Teaching ESL/EFL Reading and Writing* (2020), and the 3rd edition of *Learning Vocabulary in Another Language*. His web resources site contains many free resources for teachers and researchers, including books, word lists, vocabulary tests, articles and resources for speed reading and extensive reading.

References

Anglin, J. M. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development, 58*(10) 1–186. https://doi.org/10.2307/1166112

Bauer, L., & Nation, I. S. P. (2020). *English morphology for the language teaching profession*. Routledge. https://doi.org/10.4324/9780367855222

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography, 6*(4), 253–279. https://doi.org/10.1093/ijl/6.4.253

Brown, D., Stoeckel, T., McLean, S., & Stewart, J. (2020). The most appropriate lexical unit for L2 vocabulary research and pedagogy: A brief review of the evidence. *Applied Linguistics, 43*(3), 596–602. https://doi.org/10.1093/applin/amaa061

Dang, Y. & Li, W. (2022). *Word part instruction in pre-university EFL contexts*. [Manuscript submitted for publication].

Gellert, A. S., Arnbak, E., Wischmann, S., & Elbro, C. (2020). Morphological intervention for students with limited vocabulary knowledge: Short-and long-term transfer effects. *Reading Research Quarterly, 56*(3), 1–19. https://doi.org/10.1002/rrq.325

Goodwin, A., Petscher, Y., & Tock, J. (2021). Multidimensional morphological assessment for middle school students. *Journal of Research in Reading, 44*(1), 70–89. https://doi.org/10.1111/1467-9817.12335

Hay J. (2001). Lexical frequency in morphology: Is everything relative? Linguistics: An Interdisciplinary Journal of the Language Sciences 39(6), 1041–1070. https://doi.org/10.1515/ling.2001.041

Laufer, B., & Cobb, T. (2020). How much knowledge of derived words is needed for reading? *Applied Linguistics, 41*(6), 971–998. https://doi.org/10.1093/applin/amz051

Levesque, K. C., Kieffer, M. J., & Deacon, S. H. (2019). Inferring meaning from meaningful parts: The contributions of morphological skills to the development of children's reading comprehension. *Reading Research Quarterly, 54*(1), 63–80. https://doi.org/10.1002/rrq.219

McLean, S. (2021). The coverage comprehension model, its importance to pedagogy and research, and threats to the validity with which it is operationalized. *Reading in a Foreign Language, 33*(1), 126–140. https://doi.org/10.1093/applin/amw050

McLean, S. (2018). Evidence for the adoption of the flemma as an appropriate word counting unit. *Applied Linguistics, 39*(6), 823–845. https://doi.org/10.1093/applin/amw050

Milton, J., & Alexiou, T. (2009). Vocabulary size and the Common European Framework of Reference for languages. In B. Richards, M. H. Daller, D. D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition* (pp. 194–211). Palgrave. https://doi.org/10.1057/9780230242258_12

Milton, J., & Alexiou, T. (2020). Vocabulary size assessment: Assessing the vocabulary needs of learners in relation to their CEFR goals. In M. Dodigovic & M. P. Agustín-Llach (Eds.), *Vocabulary in curriculum planning: Needs, strategies and tools* (pp. 9–27). Springer. https://doi.org/10.1007/978-3-030-48663-1_2

Mulder, E., van der Graaf, J., Verhagen, V., Moss, M., & Droop, M. (2022). Developing a morphological awareness intervention through inquiry-based learning. *Journal of Research in Reading, 45*(3), 468–487. https://doi.org/10.1111/1467-9817.12405

Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed). Cambridge University Press. https://doi.org/10.1017/9781009093873

Nation, I. S. P., & Bauer, L. (2023). What is morphological awareness and how can you develop it? *Language Teaching Research Quarterly, 33*, 81–99. https://doi.org/10.32038/ltrq.2023.33.04

Ur, P. (2022). How useful is it to teach affixes in intermediate classes? *ELT Journal, 76*(3), 330–337. https://doi.org/10.1093/elt/ccac018

Sasao, Y., & Webb, S. (2017). The word part levels test. *Language Teaching Research, 21*(1), 12–30. https://doi.org/10.1177/1362168815586083

Washburn, E., & Mulcahy, C. (2019). Morphology matters, but what do teacher candidates know about it? *Teacher Education and Special Education, 42*(3), 246–262. https://doi.org/10.1177/0888406418806649

Wei, Z., & Nation, P. (2013). The word part technique: A very useful vocabulary teaching technique. *Modern English Teacher, 22*(1), 12–16.

Appendix

The following appendix is related to the Expositions article by Paul Nation (November 2023 issue, pp. 235–245). The *JALT Journal* editorial team sincerely apologizes for this omission.

Appendix 1: The CEFR levels, vocabulary size, and word family size

The following table attempts to relate the CEFR descriptors, proficiency levels, vocabulary size and the Bauer & Nation (1993) levels

Level	CEFR descriptors	Suggested vocabulary size	Suggested word family size (Bauer & Nation, 1993)
C2	Has a good command of a very broad lexical repertoire including idiomatic expressions and colloquialisms; shows awareness of connotative levels of meaning.	7000-9000 words	Level 6 and beyond
C1	Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions; little obvious searching for expressions or avoidance strategies. Good command of idiomatic expressions and colloquialisms.	5000-6000 words Oxford 5000 (B2-C1) See introduction to OALD 10th edition p.x.	Level 5
B2	Has a good range of vocabulary for matters connected to his/her field and most general topics. Can vary formulation to avoid frequent repetition, but lexical gaps can still cause hesitation and circumlocution.	4000 words (2000-3000 high frequency words plus 1000-2000 relevant technical vocabulary) Oxford 3000 (B2)	Level 4
B1	Has a sufficient vocabulary to express him/herself with some circumlocutions on most topics pertinent to his/her everyday life such as family, hobbies and interests, work, travel, and current events. Has sufficient vocabulary to conduct routine, everyday transactions involving familiar situations and topics.	2000-3000 most frequent high frequency words	Level 3
A2	Has a sufficient vocabulary for the expression of basic communicative needs. Has a sufficient vocabulary for coping with simple survival needs.	The 1000 most frequent word families	Level 3 partial
A1	Has a basic vocabulary repertoire	120 words and phrases from the Crabbe & Nation (1991) survival vocabulary.	Level 2 Flemma + -ly

You could classify A as Elementary, B as Intermediate, and C as Advanced.

Corpus Linguistics in EFL Language Teaching: Insights From Research and Practice

Shelley Staples
University of Arizona

Laurence Anthony
Waseda University

Corpus linguistics can provide curriculum developers and teachers with theoretical foundations and guidance when deciding learning objectives, identifying materials and methods, and evaluating learner outputs. In this Exposition piece, we discuss how corpus linguistics can inform EFL language teaching in the areas of materials creation, skill development (reading, writing, speaking, and listening), and evaluation, with special attention given to lexico-grammar and vocabulary. We also provide examples from courses delivered at Waseda University to illustrate these approaches.

コーパス言語学は、カリキュラム開発者と教師に学習目標の決定、教材と教育方法の特定、学習者の成果の評価を行う際の理論的基礎とガイダンスを提供します。この解説記事では、語彙文法に注意を払いながらコーパス言語学が資料作成、スキル開発（読み書き、スピーキング、リスニング）および評価の分野におけるEFL言語教育にどのように役立つかを解説します。また、これらのアプローチを解説するために、早稲田大学で実施されているコースの例も紹介します。

Keywords: corpus linguistics; data-driven learning; lexico-grammar; vocabulary

https://doi.org/10.37546/JALTJJ45.2-4

JALT Journal, Vol. 45, No. 2, November 2023

Corpus linguistics (CL) is a research methodology that helps researchers, curriculum developers, teachers, and even learners to understand language use in different domains (e.g., journal articles, health care communication, conversation) through the analysis of a large, principled set of authentic "texts", called a corpus, which is sampled to represent the target language. A corpus (or 'corpora' in the plural form) is usually comprised of written or spoken texts from the target domain, but it can also be comprised of a mixture of different language modes including video and audio files. In this case, it would be called a multimodal corpus. Researchers, teachers and even learners of a foreign language can interact with corpora using special corpus software tools and gain a deep understanding of how language works in the real world. In some cases, this new knowledge may complement their existing knowledge, but in many cases, it may challenge their pre-existing ideas. Importantly, in an EFL context, the corpus linguistics methodology can empower researchers, teachers, and learners who are L2 speakers of English by providing them with data and tools that deliver insights which are not readily known even by L1 speakers.

In this paper, we will discuss how corpus linguistics can inform EFL language teaching in the areas of materials creation, skill development (reading, writing, speaking, and listening), and evaluation, with special attention given to lexico-grammar and vocabulary. To illustrate some of these ideas, we will provide real-world examples taken from courses delivered as part of the Center for English Language Education in Science and Engineering (CELESE) program at Waseda University (https://celese.jp/about). We will conclude the paper with some suggestions for important areas of future research that might inform EFL instruction.

Corpus Linguistics in EFL Materials Creation and In-Class Teaching

Corpus linguistics has profoundly changed the way in which EFL language teaching materials are created. One of the earliest examples of this trend was the use of corpora in the creation of the *Collins COBUILD English Language Dictionary,* a project headed by John Sinclair at The University of Birmingham (Sinclair, 1987). The COBUILD dictionary was unique for its time as it included frequency of use information on the words included, as well as examples taken directly from the COBUILD corpus, which allowed learners to see how words were used in authentic contexts. Today, almost all learner's dictionaries are created with the help of corpus linguistics methods, from the design of the underlying corpus and the selection of entries and examples to the inclusion of supplementary notes on grammar and usage patterns. We

can see a similar trend in the creation of corpus-based reference grammars, such as the *Grammar of Spoken and Written English* (Biber et al., 2021) and the creation of corpus-informed textbooks for General English and English for Specific Purposes (ESP). A good example of the latter is the *Touchstone* series of textbooks created by McCarthy et al. (2014) that takes into account the differences between spoken and written discourse and offer students examples of the "messiness" of spoken interactions.

Teachers interested in incorporating CL into their own courses can, as a first step, evaluate textbooks for their use (or not) of corpora to inform their design, and assess the approaches taken to introduce topics to students. For example, does the textbook include authentic texts/dialogues that allow students to explore language in use (see McCarthy & O'Keefe, 2014 for a discussion of this issue), or, when introducing passive voice, is the higher frequency of the agentless passive mentioned (see Meunier & Reppen, 2015 for a discussion of passive voice presentation in corpus vs. non-corpus informed textbooks)? Going one step further, insights from corpus linguistics can also be used directly by teachers as they prepare materials for class. For example, instead of adopting the traditional approach of covering all tense-aspect pairs in a grammar class, starting with present progressive, teachers can take a more frequency-based approach which would start with simple present and not introduce progressive until later (see Biber and Reppen, 2002 for a discussion of this phenomenon in learner textbooks).

Another way that teachers can use corpus linguistics in their materials creation is with a (learner) corpus of their students' own assignments. Learner corpora have traditionally been used by teachers (and researchers) to identify errors in learner output, but there is growing interest in using learner corpora to identify examples of both positive and negative language use. With a learner corpus at hand, teachers can apply corpus methods such as KWIC (Key-Word-In-Context) concordancing and cluster analysis to identify common patterns in learner writing and use some of the authentic examples in class to illustrate both effective and ineffective language patterns. They can also use learner corpora to introduce students to genres that are more commonly used in the classroom and in more familiar contexts than general corpora (Seidlhofer, 2002; Tribble, 2001), presenting student texts as models for discussion of language choice and effectiveness. Finally, teachers can create a corpus of the textbook materials they are asked to use and apply corpus methods such as word list and keyword list generation to identify the most frequent words to teach.

We have just noted that corpus texts can be used directly in the class-

room as models of language use. Expanding on this concept, some notable scholars in the late 1980s and early 1990s (e.g., Johns, 1990) proposed introducing the principles of corpus linguistics directly to students in the form of data-driven learning (DDL). DDL is an inductive approach to language learning whereby students are provided with data from a corpus and asked to analyze and reach conclusions about common patterns of language used in the target domain. Initially, Johns adopted a 'soft' approach to DDL that relied on printed handouts of KWIC concordance outputs. Today, the dramatic increase in the power of computers has allowed for a 'hard' approach to be possible, where students query a target corpus directly either through a web-based corpus analysis tool (e.g., SketchEngine[1], CQPWeb[2]) or a desktop tool (e.g., AntConc[3], WordSmith Tools[4]). In fact, numerous large scale meta-analyses of results of 'hard' DDL have shown it to produce moderate to large gains (effect sizes) in learning, particularly in the areas of vocabulary, lexico-grammar, and writing (Boulton & Cobb, 2017; Boulton & Vyatkina, 2021). Also, while the 'soft' DDL approach showed lower effect sizes in Boulton and Cobb's (2017) meta-analysis, the reported gains are still larger than for many other types of computer assisted language learning (CALL) (c.f. Plonsky & Ziegler, 2016).

Several models have been introduced for conceptualizing the lesson arc for DDL. The "4 Is" approach proposed by Lynne Flowerdew (2009) builds off an earlier "3 Is" model (which excluded step 3) by Carter and McCarthy (1995) and includes the following steps:

1. Illustration: looking at data
2. Interaction: discussion and sharing observations and opinions
3. Intervention: optional step to provide learners with hints or clearer guides for induction
4. Induction: making one's own rule for a particular feature

Ma et al. (2021) proposes a broader model for DDL that comprises the following four steps:

1. testing students' knowledge
2. hands-on corpus search by students
3. inductive discovery by students
4. output activities

Steps 2 and 3 from Ma et al. (2021) clearly fit within steps 1-4 from Flowerdew's model, and thus the two can be usefully combined. Notably, both models can be used with the 'soft' and 'hard' approaches to DDL.

In the CELESE program at Waseda University, corpus linguistics methods are used as part of materials creation across the entire curriculum, from required courses in communication strategies (CS), academic listening comprehension (ALC), concept building and discussion (CBD), and academic reading (AR) to elective courses, such as technical writing (TW) and technical presentation (TP). All materials for the program are developed in-house, which allows for target vocabulary, grammar patterns, illustrative, dialogues, and examples to be informed by corpora from the target domain of science and engineering. An extreme version of this happens in the second part of the TW course, where students are encouraged to create their own materials in the form of a corpus of target research articles from their own specific disciplines, such as physics or mathematics. Then, in class, students are guided on how to use these materials in combination with corpus tools to inform their own writing practices.

Corpus Linguistics in EFL Skills Development

Vocabulary and Extensive Reading

Corpus linguistics provides insights on language use in authentic settings, whether those be real-world conversational settings or the fictional worlds of novels and plays. As a result, many of the findings from corpus linguistics can be applied directly to reading instruction.

Perhaps the most influential corpus work on reading has been in the area of vocabulary. Early pioneering researchers, such as West (1953) and later Nation (2001) and others took large corpora of general and specialized English and profiled the vocabulary used in the texts to generate lists of the most productive vocabulary in terms of frequency and dispersion. Today, many such lists exist across a huge range of target domains, such as general English, academic English, TOEIC, law, politics, sports, and many more (see https://www.newgeneralservicelist.com/ for many such lists). These lists can be used not only to evaluate the vocabulary knowledge of learners, but also used in combination with a vocabulary profiling tool, such as AntWordProfiler[5], to gauge the difficulty (and suitability) of texts for a target learner audience. In preparation for a reading class, a teacher can profile the target reading and then decide whether or not to gloss any potentially difficult vocabulary or perhaps even simplify the target text if the reading

goal is fluency (see Donley & Reppen, 2001 and Huang & Liou, 2007 for example implementations of this approach). Another obvious application of such profiling is in the creation of reading materials for high-stakes entrance examinations.

The systematic profiling, glossing, and simplification of reading materials based on corpus-informed frequency lists has led to the development of modern graded reader book series, such as Oxford Bookworms (Bladon, 2014), Cambridge Young Readers (Prowse, n.d.) and many others. We also see the approach used in the creation of more specialized academic reading materials such as the *Longman Academic Reading* (Bottcher et al., 2014) and the *College Reading* series (Byrd et al., 2006), many of which are based on the Academic Word List (AWL) of Coxhead (2000). These readers have been shown to be effective as part of an extensive reading program (e.g., Huang & Liou, 2007), where the learners aim to read a large number of books over a set period with gradually increasing difficulty. Importantly, the books should always be at a level that is below the learner's current reading level so that the books are relatively easy for them to understand and can be read fluently and for enjoyment.

In the first year of the CELESE program at Waseda University, students are given a vocabulary goal of mastering the first 2000 word families of the West (1953) general service list. To support this goal, the students are provided with the complete word list that includes a pronunciation guide and authentic example sentences from the British National Corpus (BNC). In addition, all the course materials are designed to illustrate the use of these words, and other science, technology, engineering, and mathematics (STEM) related target words, in context. In the second year, the vocabulary goal switches to the Academic Word List (AWL) of Coxhead (2000). Again, all the words are provided in the form of word lists, the materials are designed to highlight these words, and the students are evaluated on the use of these words in their writing. Although CELESE does not run a formalized extensive reading program, students are encouraged to develop their reading fluency using science news articles, which are evaluated in terms of their vocabulary load.

Lexico-Grammar and Writing

Corpus tools are easily able to generate the most frequent words in a target corpus and show examples of how these words are used in context through KWIC concordances. Writing instruction, however, must go beyond vocabulary and guide learners on how to combine vocabulary with syntactic patterns to create phrases, clauses, sentences, paragraphs, and whole sec-

tions of discourse that adhere to the conventions of a particular register (Biber & Conrad, 2019) and discourse community (Swales, 1990).

The COBUILD project (Sinclair, 1987) mentioned earlier was initially designed as a lexicography project, but results soon emerged that blurred the lines between vocabulary and grammar and led to new insights on the connections between the two. This area of work was later termed lexico-grammar, but it also relates to 'pattern grammar', a term coined by Hunston and Francis (2000). One of the most notable works in lexico-grammar that is relevant to EFL language teaching is that of Willis (2003), who introduces numerous patterns that are useful for learners to know, such as the "FORGET + WH clause" that appears in the "*I forgot what I said*" and "*They always forget where the car keys are*". Hunston (2022) argues against presenting lexico-grammatical patterns to students in the form of a list, and instead recommends using awareness raising activities, such as re-writing activities and the hands-on analysis of corpus data by the learners through the data-driven learning (DDL) approach discussed earlier.

When it comes to writing, the DDL approach has been shown to be particularly effective at the tertiary level in the teaching of academic research paper writing, as discussed in detail by Anthony (2016, 2019), Charles (2007, 2014, 2018), and others. Charles, for example, describes how students collect high-quality research papers in their own discipline, convert the papers into a text-based form, and then load these papers as a corpus into the AntConc[3] corpus analysis toolkit. Once the corpus is loaded, the students can directly query the existence of common words, multi-word units, and phrases, as well as lexico-grammatical structures and discourse markers. Anthony (2016) reports on the many strengths of this approach, especially in a STEM context. Students are not only empowered to find answers to their individual language questions, but the language insights they gain are directly relevant to their learning goals, i.e., research article writing for publication in high impact journals.

Learner corpora can also be used effectively in the language classroom. Staples (2022) shows how learner corpora are used in the teaching of writing to promote asset-based approaches to language learning, with students examining lexico-grammatical patterns in a corpus of student papers from the same course context (Staples & Dilger, 2018-). Here, student papers are used as models for "allowable contributions to the genre" (Tribble, 2001, p. 381) and the students are asked to engage in questions around language choices that create more or less effective versions of a given assignment.

In the CELESE program at Waseda University, both traditional process-

writing methods and DDL are employed, with process writing being predominantly used in the first two years of undergraduate study, and DDL being the core methodology used in the rest of the program. As an example, students are exposed to a 'hard' form of DDL in the second half of the technical writing (TW) course that they take in the third year of their undergraduate studies. The DDL approach adopted at CELESE mirrors that described by Anthony (2016) and Charles (2007), with students analyzing corpora that they build themselves using the AntCorGen[6] discipline-specific corpus creation tool. Notably, the students all major in STEM subjects, which tends to reduce issues and challenges related to computer literacy that are often discussed in the literature on DDL (e.g., Adel, 2010).

Speaking and Listening Instruction

In our modern world, there is an abundance of easily available written text data that can be obtained from the Internet and used to create general and discipline specific corpora. It is perhaps no surprise, therefore, that much of today's research in corpus linguistics is focused on written language. However, corpus-based research on spoken language has been a feature of the field from the earliest days, and that interest appears to be growing with the availability of new general and specialized spoken corpora, such as the Spoken BNC (Love et al., 2017) and the British Academic Spoken Corpus (BASE) (Thompson & Nesi, 2001).

Research using spoken corpora shows a number of key features of spoken language compared with writing. These features include the use of incomplete clauses and sentences (and the related phenomenon of ellipsis), much more frequent use of ready-made chunks (i.e., lexical bundles, formulaic language), use of vague language (e.g., *thing, stuff*), use of high-frequency vocabulary, use of hesitation markers (e.g., *um*, *uh*), use of discourse markers (e.g., *well, so*), and repetition of vocabulary. For listeners, spoken corpora of conversation and other interactive discourse shows us that backchannelling and response tokens (e.g., *yeah, right*) are important cues for speakers to know their interlocutors are listening and understanding what they are saying. Within spoken discourse, we also tend to see more language associated with stance (e.g., *really, very*) due to the strong emphasis on interpersonal and pragmatic functions (Biber et al., 2021; McCarthy & McCarten, 2022; McCarthy & O'Keefe, 2014; Staples, 2015).

One way that teachers can incorporate these important features of spoken discourse into their classrooms is through the selection of corpus-informed textbooks or engagement with ready-made online materials. The *Touchstone*

series (McCarthy et al., 2014) is the most prominent example of a corpus-informed textbook for conversational English. It utilizes research from the Cambridge English Corpus (https://www.cambridge.es/en/about-us/cambridge-english-corpus) and is distinctive in its inclusion of the types of spoken features discussed above. *Real Grammar* (Biber & Conrad, 2009) is a textbook that includes several units focused on spoken characteristics (e.g., discourse markers, incomplete sentences) based on the *Longman Grammar of Spoken and Written English* (Biber et al., 1999). For more academic spoken language, teachers might choose to use a corpus-informed textbook such as *Academic Interactions* (Feak et al., 2009), which is based on the Michigan Corpus of Spoken English (MICASE). This textbook provides audio samples and transcripts from the corpus for speech events like office hours and classroom discussions.

Others have developed stand-alone ready-made materials for instructors to use in classrooms. Gablasova and Brezina (2017) and Gablasova et al. (2019) describe sample materials from the Trinity Lancaster Corpus (TLC) on disagreement and active listenership. Importantly, the TLC is a learner corpus. The materials can be accessed at https://www.trinitycollege.com/about-us/research/Trinity-corpus/corpus-resources. The MICASE Handbook (Simpson-Vlach & Leicher, 2006) also contains activities based on the MICASE corpus and ideas for using MICASE for pedagogical purposes.

Data-driven learning is also possible in the speaking and listening classroom through the use of spoken corpus interfaces, particularly those that provide multimodal search results. As an example, Youglish (https://youglish.com/) searches 100 million spoken tracks to give users samples of words pronounced in context. It also allows users to search varieties of English (e.g., US, UK, Australia). The TED Corpus Search Engine (TCSE, https://yohasebe.com/tcse/) provides users with the ability to retrieve audio and transcripts from TED talks in context (Hasebe, 2015). In addition, there are various commercial learning platforms that allow learners to query multimodal corpora and view examples phrases and sentences aligned with their associated video clips.

Speaking and listening are essential components of the CELESE program at Waseda University. In the first year of the program, for example, the Communication Strategies (CS) course aims to develop the students' ability to speak in various academic settings, such as research labs and conferences. Similarly, the Academic Lecture Comprehension (ALC) course aims to develop the students' ability to listen and comprehend academic lectures, as well as take notes on those lectures, and summarize the main points in the

form of a written or oral report. In the second year, the Concept Building and Discussion (CBD) course is designed to develop these speaking and listening skill further so that the students can confidently present and discuss the findings of mini projects that they conduct in groups or individually. Then, in the third and fourth year of the program, these skills are extended further in the Technical Presentation (TP) course, which aims to help students deliver a conference-level oral presentation about their research and respond to questions and comments about the work. Corpus-based research has been a key factor in the development of all these courses. For example, a major corpus project was initiated to understand the language used by experienced lecturers and presenters in different STEM fields (Kunioshi et al., 2016). Similarly, materials for developing successful Q&A strategies used in the TP course are based on a corpus of Q&A interactions recorded at a real conference and later transcribed.

Corpus Linguistics in EFL Evaluation

Corpus linguistics as a field is primarily concerned with describing how language is used in the real world. In the EFL classroom, however, one of the most important jobs of the teacher is evaluating the language output of the learner and assigning a grade. This raises an interesting question: What insights do corpus linguistics provide in terms of learner assessment?

In fact, corpora are the foundation of almost all automated evaluation tools used in EFL. At the most basic level, corpora of existing public domain language and local student submissions are used by plagiarism detection tools, such as Turnitin (https://www.turnitin.com/), to measure the degree to which a newly submitted student paper overlaps with existing work. The algorithms used to measure the degree of overlap are also founded on principles developed through corpus linguistics research, such as n-gram analyses (https://en.wikipedia.org/wiki/N-gram).

In the area of error detection, corpora of manually error-corrected learner writing samples are used by many automatic error detection tools to flag potential errors in writing and offer suggestions for improvement (Callies & Götz, 2015). Similarly, corpora of writing samples at different quality levels are used by testing services, such as the Educational Testing Service (ETS) (https://www.ets.org/), to automatically grade writing submitted as part of tests such as TOEIC and TOEFL. In these cases, the algorithms that compare the submitted writing with the corpus samples and assign grades can vary from simple rule-based error counting algorithms to highly complex algorithms that involve large-language models (LLMs) and deep learning.

People in the field are currently debating what aspects of LLMs (if any) are developed out of ideas from corpus linguistics. However, it is clear that some of the most important underlying principles of LLMs match ideas that were discussed in the very early days of corpus linguistics (see Firth, 1957).

In the CELESE program at Waseda University, all student reports are graded manually by teachers. However, research is in progress to determine the effectiveness of automated corpus-based grading approaches (Wang, 2022). In addition, corpus-based methods are used to support some aspects of teacher evaluation. For example, Turnitin is used to check for potential cases of plagiarism across all courses. Also, in the Concept Building and Discussion (CBD) course, students are required to highlight the use of at least three words from the Academic Word List (Coxhead 2000) in their writing, so they are encouraged to use the AntQuickTools[7] to quickly profile their work and highlight all words from the Academic Word List automatically. Teachers are also recommended to use this tool to check that students' have completed the task correctly.

Possibilities for Future Research

With the growth of technology and access to data on the Internet, a great number of corpora are now available for teachers to choose from, including general corpora (e.g., Corpus of Contemporary American English), national corpora (e.g., British National Corpus), and specialized corpora (e.g., British Academic Written English corpus, a corpus of written work by students in British universities). These may be useful for teachers who find the existing materials (e.g., textbooks) less relevant for their contexts or who want to introduce students to corpus consultation to enhance their own learning. However, as mentioned earlier, the number of large-scale spoken corpora is still relatively small. Therefore, one important area of corpus research is determining how to effectively collect, transcribe, and annotate spoken data. A related question is how to develop corpus tools that align and visualize multimodal data in an intuitive way.

Vocabulary lists have been another major outgrowth of CL research, providing students and instructors with frequently used words in specialized areas such as engineering (e.g., Basic Engineering Word List; Ward, 2009) and medicine (e.g., Medical Academic Word List; Wang, Liang, & Ge, 2008). However, the words in these lists are almost universally ranked by either their frequency of occurrence in the corpus or their dispersion across files in the corpus. Several methods have been proposed to rank words using other measures. For example, Savický and Hlavácová (2002) have proposed

an average reduced frequency (ARF) measure that combines frequency and dispersion into a single number. However, the meaning of this number is effectively impossible to interpret without knowing the values on which it is based. Schmitt et al. (2021) have released knowledge-based vocabulary lists (KVL) that are based on a measure of learners' ability to produce words. However, these lists are extremely time consuming to generate and are currently only available for Chinese, German, and Spanish learner contexts. Clearly, there is a need for more research on effective word ranking measures for different purposes. In fact, the unit of analysis in all these works (i.e., the definition of a word) is also open to challenge.

In view of the availability of written corpora and the scarcity of large-scale spoken corpora, it is not surprising that there is relatively more research that looks at corpus-based lexico-grammar and writing instruction, and less that focuses on reading and speaking instruction. While important inroads have been made, the research on how to effectively use corpora for reading and speaking purposes is limited. In addition, and in some ways related to these limitations, most of the empirical research on the effectiveness of corpus-based instruction has focused on the use of concordance lines to inductively highlight patterns of lexico-grammar. More work to show broader contextual use of language is needed, as well as alternatives that might be more relevant for instruction beyond lexico-grammar (including genre-based instruction of writing and features of dialogic spoken discourse that rely on the unfolding of meaning over several turns). The use of corpora for spoken instruction necessarily relies on the development of multimodal corpora, which are limited and almost never found in studies of corpus-based instruction. Such developments would also align with multiliteracies frameworks for language learning, which are being adopted more broadly (see New London Group, 1996 for details of this framework).

While Chujo and Nishigaki (2004), Crosthwaite (2020), Kakiba et al. (2021), and a few others provide important first looks at what corpus-based instruction can look like in primary and secondary schools, much more work is needed to understand how corpus methods can be incorporated with other approaches commonly found in these contexts, including content-based instruction. Similarly, research on the use of corpora in post-tertiary adult learning courses is scarce, although interesting work is beginning to emerge from the teaching of teaching of Welsh to adult learners as part of the CorCenCC project (Knight et al., 2020).

Summary

In this paper, we have focused on key areas of corpus linguistics that are relevant to EFL teaching. Firstly, we introduced definitions for corpora and corpus linguistics and then discussed how corpora can be used in EFL materials creation and in-class teaching. Next, we discussed how corpus linguistics principles can be used in the teaching of vocabulary and reading, lexicogrammar and writing, and speaking/listening. We ended with thoughts on the use of corpora for evaluation/assessment and future research. In each section, we provided examples to contextualize the various approaches for EFL teaching in Japan. We hope this paper will provide teachers with the background to get started in using corpora, as well as references that they can use to gain a deeper understanding of corpus linguistics in EFL.

Notes

1. SketchEngine. https://www.sketchengine.eu/.
2. CQPWeb. https://cqpweb.lancs.ac.uk/.
3. AntConc. https://www.laurenceanthony.net/software/antconc/.
4. WordSmith Tools. https://www.lexically.net/wordsmith/.
5. AntWordProfiler. https://www.laurenceanthony.net/software/antwordprofiler/.
6. AntCorGen. https://www.laurenceanthony.net/software/antcorgen/.
7. AntQuickTools. https://www.laurenceanthony.net/software/antquicktools

Shelley Staples is Associate Professor of English Applied Linguistics/Second Language Acquisition and Teaching in the English Department at University of Arizona, United States. Her research focuses on corpus-based analyses and instruction of academic writing and speaking, with a particular emphasis on second language writing.

Laurence Anthony is Professor of Applied Linguistics and founding member of the Center for English Language Education in the Faculty of Science and Engineering, Waseda University, Japan. His main research interests are in corpus linguistics, educational technology, language data science, and English for Specific Purposes (ESP) program design and teaching methodologies.

References

Adel, A. (2010). Using corpora to teach academic writing: Challenges for the direct approach. In M. C. Campoy-Cubillo, B. Belles-Fortuño, & M. L. Gea-Valor (Eds.), *Corpus-based approaches to English language teaching* (pp. 18–35). Continuum.

Anthony, L. (2016). Introducing corpora and corpus tools into the technical writing classroom through Data-Driven Learning (DDL). In J. Flowerdew & T. Costley (Eds.) *Discipline specific writing* (pp. 162–180). Routledge.

Anthony, L. (2019). Tools and strategies for Data-Driven Learning (DDL). In K. Hyland & L. Wong (Eds.) *Specialised English: New directions in ESP and EAP research and practice* (pp. 179–194). Routledge. https://doi.org/10.4324/9780429492082-14

Biber, D., & Conrad, S. (2009). *Real grammar: A corpus-based approach to English*. Pearson Longman.

Biber, D., & Conrad, S. (2019). *Register, genre, and style* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/9781108686136

Biber, D., Johansson, S., Leech, G. N., Conrad, S., & Finegan, E. (2021). *Grammar of spoken and written English*. John Benjamins. https://doi.org/10.1075/z.232

Biber, D., & Reppen, R. (2002). What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition, 24*(2), 199–208. https://doi.org/10.1017/S0272263102002048

Bladon, R. (Ed.) (2014). *Oxford bookworms library*. Oxford University Press.

Bottcher, E., Sanabria, K., Miller, J. L., Cohen, R. F., & Smith, L. C. (2014). *Longman academic reading series.* Longman.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*, 348–393. https://doi.org/10.1111/lang.12224

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology, 25*(3), 66–89.

Byrd, P., Schuemann, C., Reid, J., Benz, C., & Folse, K. (2006). *College reading series*. Cengage.

Callies, M., & Götz, S. (2015). *Learner corpora in language testing and assessment*. John Benjamins. https://doi.org/10.1075/scl.70

Carter, R., & McCarthy, M. (1995). Grammar and the spoken language. *Applied Linguistics, 16*(2), 141–158. https://doi.org/10.1093/applin/16.2.141

Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes, 6*(4), 289–302. https://doi.org/10.1016/j.jeap.2007.09.009

Charles, M. (2014). Getting the corpus habit: EAP students' long-term use of personal corpora. *English for Specific Purposes, 35*, 30–40. https://doi.org/10.1016/j.esp.2013.11.004

Charles, M. (2018). Corpus-assisted editing for doctoral students: More than just concordancing. *Journal of English for Academic Purposes, 36*, 15–25. https://doi.org/10.1016/j.jeap.2018.08.003

Chujo, K., & Nishigaki, C. (2004, December). Creating e-learning material to teach essential vocabulary for young EFL learners. In *Proc. of An Interactive Workshop on Language e-Learning* (pp. 35–44). Waseda University.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213–238. https://doi.org/10.2307/3587951

Crosthwaite, P. (2020). *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners*. Routledge. https://doi.org/10.4324/9780429425899

Donley, K., & Reppen, R. (2001). Using corpus tools to highlight academic vocabulary in SCLT. *TESOL Journal, 10*(2–3), 7–12.

Feak, C. B., Reinhard, S. M., & Rohlck, T. N. (2009). *Academic interactions: Communicating on campus*. Michigan ELT.

Firth, J. R. (1968). A synopsis of linguistic theory, 1930-55. In F. Palmer (Ed.), *Selected papers of J. R. Firth, 1952-59* (pp. 168–205). Indiana University Press.

Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics, 14*(3), 393–417. https://doi.org/10.1075/ijcl.14.3.05flo

Gablasova, D., & Brezina, V. (2017). Disagreement in L2 spoken English: From learner corpus research to corpus-based teaching materials. In V. Brezina & L. Flowerdew (Eds.), *Learner corpus research: New perspectives and applications* (pp. 69–89). Bloomsbury Publishing.

Gablasova, D., Brezina, V., & McEnery, T. (2019a). The Trinity Lancaster corpus: Applications in language teaching and materials development. In S. Götz & J. Mukherjee (Eds.), *Learner corpora and language teaching* (pp. 7–28). John Benjamins. https://doi.org/10.1075/scl.92.02gab

Hasebe, Y. (2015). Design and implementation of an online corpus of presentation transcripts of TED Talks. *Procedia – Social and Behavioral Sciences, 198,* 174–182. https://doi.org/10.1016/j.sbspro.2015.07.434

Huang, H-T., & Liou, H-C. (2007). Vocabulary learning in an automated graded reading program. *Language Learning & Technology, 11*(3), 64–82.

Hunston, S. (2022). *Corpora in applied linguistics* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/9781108616218

Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. John Benjamins. https://doi.org/10.1075/scl.4

Johns, T. (1990). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria, 10*, 14–34.

Kakiba, A., Nishigaki, C., & Oghigian, K. (2021). DDL applications to the seventh grade EFL classroom in Japan. *Bulletin of the Faculty of Education, Chiba University, 69*, 167–179.

Knight, D., Morris, S., Fitzpatrick, T., Rayson, P., Spasić, I., Thomas, E-M., Lovell, A., Morris, J., Evas, J., Stonelake, M., Arman, L., Davies, J., Ezeani, I., Neale, S., Needs, J., Piao, S., Rees, M., Watkins, G., Williams, L., Muralidaran, V., Tovey-Walsh, B., Anthony, L., Cobb, T., Deuchar, M., Donnelly, K., McCarthy, M., & Scannell, K. (2020). *CorCenCC: Corpws Cenedlaethol Cymraeg Cyfoes – the National Corpus of Contemporary Welsh*. Cardiff University. http://doi.org/10.17035/d.2020.0119878310

Kunioshi, N., Noguchi, J., Tojo, K., & Hayashi, H. (2016). Supporting English-medium pedagogy through an online corpus of science and engineering lectures. *European Journal of Engineering Education, 41*(3), 293–303. https://doi.org/10.1080/03043797.2015.1056104

Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics, 22*(3), 319–344. https://doi.org/10.1075/ijcl.22.3.02lov

Ma, Q., Tang, J., & Lin, S. (2021). The development of corpus-based language pedagogy for TESOL teachers: A two-step training approach facilitated by online collaboration. *Computer Assisted Language Learning, 35*(9), 2731–2760. https://doi.org/10.1080/09588221.2021.1895225

McCarthy, M., & McCarten, J. (2022). Corpora for teaching social conversation. In R. R. Jablonkai & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (pp. 102–115). Routledge. https://doi.org/10.4324/9781003002901-9

McCarthy, M., McCarten, J., & Sandiford, H. (2014). *Touchstone* (2nd ed.). Cambridge University Press.

McCarthy, M., & O'Keefe, A. (2014). Spoken grammar. In M. Celce-Murcia, D. M. Brinton, & M. A. Snow (Eds.), *Teaching English as a second or foreign language* (4th ed., pp. 271–287). National Geographic/Cengage.

Meunier, F., & Reppen, R. (2015). Corpus vs. non-corpus informed pedagogical materials: Grammar as the focus. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 498–514). Cambridge University Press. https://doi.org/10.1017/CBO9781139764377.028

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review, 66*(1), 60–92. https://doi.org/10.17763/haer.66.1.17370n67v22j160u

Plonsky, L., & Ziegler, N. (2016). The CALL-SLA interface: Insights from a second-order synthesis. *Language, Learning and Technology, 20*(2), 17–37.

Prowse, P. (Ed.) (n.d.). *Cambridge young English readers*. Cambridge University Press.

Savický, P., & Hlavácová, J. (2002). Measures of word commonness. *Journal of Quantitative Linguistics, 9*(3), 215–231. https://doi.org/10.1076/jqul.9.3.215.14124

Schmitt, N., Dun, K., O'Sullivan, B., Anthony, L., & Kremmel, B. (2021). Introducing knowledge-based vocabulary lists (KVL). *TESOL Journal, 12*(4). https://doi.org/10.1002/tesj.622

Seidlhofer, B. (2002). Pedagogy and local learner corpora: Working with learning-driven data. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 213–234). John Benjamins. https://doi.org/10.1075/lllt.6.14sei

Simpson-Vlach, R., & Leicher, C. (2006). *The MICASE handbook: A resource for users of the Michigan corpus of academic spoken English*. Michigan ELT. https://doi.org/10.3998/mpub.101203

Sinclair, J. (Ed.). (1987). *COBUILD dictionary* (1st ed.). Collins ELT.

Staples, S. (2015). *The discourse of nurse-patient interactions: Contrasting the communicative styles of U.S. and international nurses*. John Benjamins. https://doi.org/10.1075/scl.72

Staples, S. (2022, July 15). *Learner corpora and data-driven learning: moving toward an asset-based approach* [Plenary session]. 15th Teaching and Language Corpora (TaLC) Conference, Limerick, Ireland.

Staples, S., & Dilger, B. (2018-). *Corpus and repository of writing* [Learner Corpus Articulated With Repository]. https://crow.corporaproject.org.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

Thompson, P., & Nesi, H. (2001). Research in progress, the British Academic Spoken English (BASE) corpus project. *Language Teaching Research, 5*(3), 263. https://doi.org/10.1191/136216801680223443

Tribble, C. (2001). Small corpora and teaching writing. In M. Ghadessy, A. Henry, & R. L. Roseberry (Eds.), *Small corpus studies and ELT: Theory and practice* (pp. 381–408). John Benjamins. https://doi.org/10.1075/scl.5.22tri

Wang, J., Liang, S-I., & Ge, G-C. (2008). Establishment of a medical academic word list. *English for Specific Purposes, 27*(4), 442–458. https://doi.org/10.1016/j.esp.2008.05.003

Wang, Q. (2022). The use of semantic similarity tools in automated content scoring of fact-based essays written by EFL learners. *Education and Information Technologies, 27*(9), 13021–13049. https://doi.org/10.1007/s10639-022-11179-1

Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes, 28*(3), 170–182. https://doi.org/10.1016/j.esp.2009.04.001

West, M. (1953). *A general service list of English words*. Longman, Green and Co.

Willis, D. (2003). *Rules, patterns and words: Grammar and lexis in English language teaching*. Cambridge University Press. https://doi.org/10.1017/CBO9780511733000

A Change for the Times: Rethinking Book Review Writing

Melodie Cook
University of Niigata Prefecture

In this exposition, Melodie Cook, the incoming Book Reviews Editor, outlines how book reviews have been evaluated and written in the past and how *JALT Journal* is currently making an effort to provide reviewers a platform through which to exercise their critical skills, as well as provide a more interactive experience for readers to engage with book reviews. At the end of the exposition, she provides guidelines for future book reviews to consider when writing book reviews for JALT Journal as we move forward.

この解説で著者は、書評の歴史を簡単に説明し、なぜ書評がこれまでそれほど真剣に受け止められてこなかったのか、そして書評を知的対話の場としてどのように活用できるのかを述べている。書評をより批評的で双方向的なものにするための根拠を示した後、*JALT Journal*における今後の書評のガイドラインを提示している。

Keywords: book reviews, critical reviews, critical book review guidelines

In a piece called "Confessions of a Book Reviewer" (1946), George Orwell wrote of the trials and tribulations, in his early career, writing short reviews, under shorter deadlines, for books the contents of most of which he had little or no familiarity with. He opined that his work did questionable service to authors; and that other book reviewers were in the same boat as he. Orwell advocated for closer and better reading and reviewing of fewer books with longer word-counts and that books about specific topics be read by experts.

https://doi.org/10.37546/JALTJJ45.2-5

JALT Journal, Vol. 45, No. 2, November 2023

In 1980 the second issue of *JALT Journal* was published, and it included its first book review written by Kevin Gregg (1980). Since then, book reviews have been an essential feature of *JALT Journal*, providing valuable insight into new publications for its readership. Generally, book reviews have been useful in not only giving authors and publishers a space for readers to learn about a publication but have also been helpful for providing new researchers an outlet to begin their academic-publishing journeys. In light of these purposes, the JALT Publications Board wants to provide guidance on the future of book reviews in *JALT Journal* as a space for presenting a rigorous discussion informed by new conceptual and empirical developments in the literature of our field and provides readers with a grounded understanding and balanced interpretation of a given work. In this *Expositions* article, the incoming Book Reviews Editor, Melodie Cook, writes about how book reviews have come to be, how they are presently viewed, and how they might be critically expanded. In doing so, she, as well as all involved in JALT, hope that a new vision of book reviews will emerge and that book reviews can continue to be seen as an essential and worthwhile part of *JALT Journal*.

A Brief History of Book Reviews

Book reviews have been a part of journals since the *Journal des Scavans* began publishing them in Paris in 1665 (Orteza y Miranda, 1996). Book reviews covered all fields of knowledge at that time and were not expected to include the reviewer's opinions nor any discussion of the content of books. In short, book reviews "had a conservative function, namely, to record publication and to inform scholars and the reading public" (p. 192). Because the scope of book reviews was so broad and all new books needed to be recorded, reviewers were generally inundated and, at times, reviews were of questionable quality. I was reminded of Lady Carbury in Trollope's *The Way We Live Now*; a sloppy researcher who wrote what were deemed to be bad biographies, but nevertheless sent letters to various male newspaper editors begging them to give her favorable reviews. Although Trollope was writing about newspaper book reviews, I felt the same was applicable when I read the following quotation (Roper, 1978, as cited in Orteza y Miranda, 1996):

> Most of the critical journals of the time were either what amounted to publishers' organs, written by hacks who sneered or rhapsodized at their employers' bidding, or unscrupulous

> instruments of party politics, buttering or slashing up a book in accordance with its author's political affiliations. (p. 43).

This manner of "comprehensive reviewing" (Orteza y Miranda, 1996) changed with the publication of *The Edinburgh* in 1802; the editors began to show more selectivity and agreed that quality, not quantity, should be the main concern of book reviews. In addition, the function of reviews began to be more seriously considered – in the past, reviewers simply made observations on the books, and perhaps quoted from them at length. Now, reviewers were required to show knowledge of the subject and be able to provide useful critiques and analyses.

The audience for book reviews, in the past, "was for a restricted and limited audience of educated people, presumed to be capable of making their own judgments regarding quality of books" (Orteza y Miranda, 1996, p. 193), but as more people received a formal education, thus increasing literacy rates, more books needed to be published. This resulted in pressure on publishers to produce more books and which generated a need for authors of fiction and academics to publish more works. Orteza y Miranda notes that the judges of such reviewers became "a new breed of professionals" (p. 194). Reviewers often struggled with the following questions:

> How partial or impartial could I or should I be to the author's work? Will I use the author's work as an opportunity to advance my views regarding the author's subject disregarding the latter's intents and purposes? Does the fact that the author of a work being reviewed is alive or is a close associate of mine, make a difference to me? In short, the question is: how ought I to conduct the review observing intellectual honesty, fairness, and objectivity? Who, in any way, is supposed to benefit from book reviews? What are reviews for? (Orteza y Miranda, 1996, p. 194).

According to Orteza y Miranda (1996), book reviews are often written using a descriptive style of writing that can be characterized as a "simple enumeration of … contents and a description of how these are laid out chapter by chapter" (p. 194). Although a few criticisms may have been leveled at a book under review, the reviews also may have been quite slight, likely because of strict word limits.

Book Reviews: Not Getting the Respect They Deserve

In the field of language learning, teaching, and research, Sealey (2015) explains why book reviews have generally not enjoyed the status they truly deserve. Citing Lindholm-Romantschuk's (1998) work, she lists the following reasons why journals' book reviews may be taken less seriously than the articles they accompany: they are shorter, because of a lack of space for developing arguments; they are largely subjective due to a lack of standardization; they are derivative in nature, because they are deemed to be "not a form of original scholarship" (p. 478); and, they tend to lack enough academic contextualization, as shown by the few citations used in a book review. Book reviews, too, tend to be counted for less than monographs; they are not awarded as many "points" by academic institutions or databases and are often not considered by tenure and promotion committees. Furthermore, with the rise of citation metrics, fewer academics want to spend their time writing them as book reviews tend to be cited less often than research studies, for example. In the end, some researchers have characterized book reviews in a very unfavorable light, such as Hoge and West (1979, as cited in East, 2011) in saying that they are "frequently brief, impressionistic, formulaic, bland, badly written or… nothing more than sales pitches" (p. 35).

Stilwell (2003, as cited in Obeng-Odoom, 2014) refers to academic capitalism as a potential reason for the relatively poor public perception of book reviews. Book reviews may not be perceived as a legitimate part of academia because they generate nothing for institutions, such as a grant for authors to write a book review. One result of these realities is book reviewing is put into the hands of budding academics, such as graduate students; who are often encouraged to write book reviews as a first post-graduate academic endeavour. I, too, was encouraged to do so after receiving my freshly-minted PhD and my first book review was published in Linguist List (Cook, 2011) (by the way, I don't believe I ever even thought to list it on my curriculum vitae – I am referencing it for the first time in this *Exposition*).

On the other hand, some academics are more positive, and see the potential for book reviews to become another space for strengthening and furthering academic rigor. Obeng-Odoom (2014) argues that book reviews "… are evaluative commentaries in which reviewers demonstrate their knowledge of the books, where they stand in the scholarly literature, and what contribution they make" (p. 79). He lists several journals in which stated aims for book reviews are clear: *International Sociology Reviews*, which presents book reviews as a "vehicle for considering, examining, appraising, assessing, and evaluating books by sociologists all over the world" (p. 80) and *The*

Journal of Political Economy which combines book notes and book reviews; the former provides summaries of books, and the latter provides evaluation. Thus, it appears that academic journals are already reimaging traditional notions of book reviews and giving them the space, attention, and respect they are due.

The Benefits of Book Reviews

For book reviewers, having to critically read a whole book and evaluate it is an excellent exercise in itself. Obeng-Odoom (2014) shares his own experiences of being able to get more out of a book by reviewing it; he creates, for himself, "documentation of the salient points raised in the book" which is "useful for future study and reference" (p. 81). Not only this, he can judge the quality of the writing which enhances his own research. In short, Obeng-Odoom claims that "[b]ook reviews can help sharpen our own writing and develop our ideas" (p. 81). He also mentions that the skill of evaluating "substantial amounts of research" (p. 81) can be developed. This is a transferable skill that can be useful to not only PhD students embarking on their academic journeys, but also to seasoned researchers keeping them abreast of the state of the art in their fields. Another benefit of writing reviews, according to Obeng-Odoom, is that the reviewer becomes known as a specialist in the field. "The benefits of being known as an expert are numerous; they give one visibility, impact and attention, among other things" (p. 81). Finally, reviewers, if given good feedback from review editors, help them get ideas. Obeng-Odoom lists several journals which have a dialogic form of feedback between reviewers and book review editors: *Review of Radical Political Economics, Agriculture and Human Values, and African Review of Economics and Finance.*

By bringing a publication to the attention of an academic community, book reviewers can help authors spread knowledge of their work, which in turn can help academics with their careers. On the other hand, book reviewers can warn readers about books that should not have been published in the first place; this forces academics to work more rigorously and keep standards in the field high. Another benefit, as we well know, is that textbook reviews can help teachers choose course materials.

Book reviews are also beneficial for the academic community. They can become a forum for professional discussions. In that sense they can offer readers an intermediary space between authors and reviewers and open up room for dialogue between the author, the reviewer, and the reader (Sealey, 2015). Because technology allows readers to locate and peruse books

themselves from online catalogs and form their own opinions about them, book reviews need to provide something that readers can't simply get from a search engine. Sealey (2015), herself a former Book Reviews Editor for *Applied Linguistics*, recommends that a critical book review should include answers to the following questions:

- What are the main themes and issues covered by this book?
- What approach do(es) the author(s) take to their material?
- Where does the book sit in relation to other books in its field and subfield? What new contribution does it make – or perhaps fail to make?
- Who is the audience for this book and is it written in a way that suggests it is appropriate for that audience?
- Is anything distinctive about the way the book is produced that would be helpful for readers to know (e.g., layout, design, images, tables, index)? (p. 482).

Although she also suggests that the "informed critical commentary" (p. 482) might more easily be provided by established researchers who not only have current knowledge of the subject, but also that appropriate and extensive historical knowledge of the subject should bring a wider perspective to the review beyond what is afforded by the questions listed above. She also recommends that postgraduate students, who may be "immersed in the literature about their topic more thoroughly than full-time faculty are able to be" (p. 484) are also valuable and knowledgeable book-review writers. A very well-researched and articulated book review could reveal several skills held by the author: the ability to summarize, the ability to write well, and the ability to problematize. Also, it can represent a thorough understanding of a given field.

JALT Journal, as the flagship research publication of JALT, is now seeking to increase its significance to readers partly by publishing book reviews which not only describe new publications, but also provide critical reviews of them. Such evaluations can help readers determine whether the new publication is of high academic quality and integrity, offers something new to the field, "discloses hitherto important but undetected and untreated problems in a study" (Orteza y Miranda, 1996, p. 194), or argues that the book should not have been published at all (e.g., Gregg, 1980). The following is a summary of Orteza y Miranda's recommendations for a critical book review:

- Determining whether or not, or how far the author has succeeded in putting forth their arguments clearly, convincingly and compellingly.

- Capturing the essence of the author's work by examining how arguments are made to support the main thesis.
- Providing critical comments in an academic tone.
- Determining that the author's expression shows logical coherence and flow.
- Setting the book in its broader disciplinary base or in relation to other works of the same genre.
- Writing in a forceful, vigorous, forthright, non-evasive and discerning manner without being unnecessarily savage.
- Engaging the reader of the review in a discussion about the author's intentions for writing the book.

To this end, and based on the opinions and suggestions of the incoming Book Review Editor enumerated in this *Expositions* article, we suggest new guidelines for *JALT Journal* book reviewers to follow along with more detailed questions that will help academics new to this genre understand exactly what is being asked of them. We hope that future reviewers will join us on this journey and help our book reviews evolve.

Who Should Write Reviews?

In the past, book reviews were written by solo authors, often graduate students, embarking on their academic careers. We would like to continue this tradition but would also like to offer the task of reviewing books to experts in their fields, their graduate students, perhaps both writing together. Although the expert researcher brings a seasoned and well-grounded view to the task, the graduate student, who is likely well-versed in state-of-the-art research, can bring a contemporary perspective. We would also like to invite pairs of researchers to co-author reviews of the same book. This would bring an active, dialogic perspective and offer an in-depth discussion of a book from multiple perspectives.

What Should Reviews Include?

Of course, presenting the fundamental information about authors, publishers, prices, and the main contents of books to readers remains an important aspect of book reviews. However, following the advice of Zabin (2003) we recommend that book summaries comprise no more than one-third of the total review. The remainder of the review should contain an evaluation of the book. According to Monash University's (2007) recommendations,

such an evaluation, among other things, should at least attempt to answer the following questions:

- Is the question the text tries to answer relevant, interesting, new, and useful?
- Who will find the text useful?
- Does the text give new answers to an old question?
- Is the text detailed or brief? Simple or complex?
- Is evidence presented to support the answer extensive? Strong? Weak? Contradictory?
- Are the conclusions reached final or preliminary?

How Long Should Book Reviews Be?

In order to help book-review authors be more evaluative and critical, we have decided to extend our original 1000-word limit to 2000 words. This will make it easier for reviewers to provide more comprehensive reviews than previously.

Guidance for Reviewers

The following section contains a sample guideline for prospective book reviewers to consider while writing their reviews. Of course, not all questions may be applicable, but we hope this outline provides a useful guideline for authors when structuring their book reviews and helps them provide a thorough evaluation of the book. As the guidelines show, most questions which we hope reviewers will answer can be found in the "Critique" category.

Proposed Guidelines for a *JALT Journal* Critical Review

Introduction:

- Give general information about the book's title, author(s), publisher, date of publication, number of pages, and cost in yen.
- Specify the type of book, outline the theme, and target audience.

Summary (about 650-700 words for a 2000-word review):

- What does the book promise to cover or argue?

- How is the book organized?
- What are the main points of each section of the book?

Critique:

- How is the writing style? Is it academic? Can a novice to the topic understand it or is expert knowledge of the subject needed?
- Did the author identify an audience for the work, and, if so, how well do they address that audience? If not, who do you think will get the most out of the book?
- Is the material factually accurate and contemporary?
- Does the author show an in-depth knowledge of the topic and situate the book among others of its kind?
- How well does the book fulfill its stated aims? How effective is the methodology if empirical research was conducted? Is the argument persuasive? Why or why not?
- How well does the author describe and use the presented evidence?
- How do you relate to the author's arguments? Do you agree or disagree with them and why?
- What possibilities does the book suggest or leave out? Explain, in detail, why this is a strength or weakness of the book.
- How does the book contribute to the field? What new or valuable information is given?

Conclusion:

- Provide a summary comment on the book that ties together the issues raised in the review.
- Make recommendations to readers – who would benefit from reading/purchasing the book?

Conclusion

We at *JALT Journal* are hoping to publish book reviews that provide not only summaries of new publications, but also more rigorous critical discussions which help situate each book within our field. We hope ultimately to publish book reviews which can be deemed as insightful, balanced, original, interesting, well-written, and informative. We are looking forward to your (solo or pair) contributions!

Melodie Cook is a Professor at the University of Niigata Prefecture. She has worked in a number of roles for JALT: Membership Chair, Program Chair, Chapter President, Associate Editor for *JALT Journal*, Editor for *JALT Journal*, and most recently on the Board of Directors as Director of Membership. She is happy to continue her service to JALT as *JALT Journal* Book Reviews Editor.

Acknowledgments

Melodie would like to personally thank Greg Rouault for all his years of service and hard work as Book Reviews Editor for *JALT Journal* and for his kind offer of help to her as she makes the transition to this important position. *Yoroshiku*, Greg and I will be in touch! That's a promise AND a threat!

References

Cook, M. (2011). [Review of the book *Second and foreign language education: Encyclopedia of language and education,* by N. Van Deusen-Scholl & N. H. Hornberger (Eds)]. *Linguist List*, https://old.linguistlist.org/issues/22/22-746.html

East, J. W. (2011). The scholarly book review in the humanities: An academic Cinderella? *Journal of Scholarly Publishing, 43*(1), 52–67.

Gregg, K. R. (1980). [Review of the book *English and Japanese in Contrast,* by H. M. Taylor]. *JALT Journal, 2*(1), 111–119. https://jalt-publications.org/sites/default/files/pdf-article/jj-2.1-review.pdf

Monash University Library. (2007). *Writing a critical review.* https://www.scribd.com/document/376951575/How-to-Write-a-Critical-Review-MONASH#

Obeng-Odoom, F. (2014). Why write book reviews? *Australian Universities Review, 56*(1), 78–82.

Orteza y Miranda, E. (1996). On book reviewing. *The Journal of Educational Thought (JET), 30* (2), 191–202. https://doi.org/10.11575/jet.v30i2.52434

Orwell, G. (1946, May 3). Confessions of a book reviewer. *Tribune*.

Sealey, A. (2015). Book reviews and forum contributions in applied linguistics—Continuity and change. *Applied Linguistics, 36*(4), 478–487. https://doi.org/10.1093/applin/amv036

Shaban, R. (2006). A guide to writing book reviews. *Journal of Emergency Primary Health Care, 4*(3), 1–4. https://doi.org/10.33151/ajp.4.3.379

Trollope, A. (1875). *The way we live now.* Wordsworth Classics.

Zabin, S. (2003). *How to write a critical book review*. https://www.carleton.edu/history/resources/history-study-guides/criticalbookreview/

Reviews

Language Learning Motivation: An Ethical Agenda for Research. Ema Ushioda. Oxford University Press, 2020. xi + 167 pp.

Reviewed by
Ian Allensworth
Kindai University

Author Ema Ushioda brings her considerable experience and depth of knowledge to bear in outlining a new ethical agenda for research into language learning motivation with this addition to the Oxford Applied Linguistics series. This new ethical agenda calls for those in the field to critically examine their motivations for conducting research and to re-think the objectives for the range of inquiries. Ushioda advocates for a shift away from an orientation to robust theoretical models and instead a move towards a more socially-conscious framework that is focused on providing benefit and agency to the communities and people who are being studied. She outlines the history of the field and the theoretical concept of the language learner before discussing the social contexts and ideologies that underpin language learning motivation studies. She continues to build the argument by citing anecdotal situations from her career and then presents the new agenda and concludes with a few suggestions for implementation.

Chapter 1 begins by making the case for the need to set a new ethical agenda in language learning motivation research in preparation for deeper discussions later in the work. Ushioda examines the need for a specific ethical agenda for language learning motivation studies and questions researchers' motivations for conducting such investigations. She goes on to address the question of "for what and for whom is our research?" by looking at both the psychological and pedagogical complexities of work in this area. The chapter concludes with a discussion of the ethical complexities of language learning motivation studies and an overview of the structure of the remainder of the book.

Chapter 2 gives an overview of the history of language learning motivation research. It begins with a description of the early focus on language learning and the language learner rather than language teaching and the language teacher. A discussion follows on whether the researchers' or the participants' needs are better served by this research. Next, Ushioda details the emergence of a more practitioner-validated model of research that appeared in the 1990s with an improved focus on the needs of language teachers. Finally, the chapter contains a description of the latest developments in language learning motivation research in the era of global English.

In Chapter 3, the discussion turns to the evolution of the theoretical concept of the language learner. The starting point of this evolution is characterized as a learner of a second language in a bilingual environment. Over time the conception changes into the learner as a learner of a foreign language in an institutionalized classroom. The progression continues and comes to encompass the language learner as a learner of English as a globally important language and finally, to a learner of languages other than English as a personal choice. Additionally, the theorization of motivation as a separate construct, the ethical issues associated with such a viewpoint, and the deficiencies of viewing motivation in isolation are also covered.

Chapter 4 continues the discussion of the history and development of language learning motivation studies by shifting the focus to the contexts in which the research has historically been carried out. Specifically highlighted, is the dominance of English language classrooms in research in recent years as privileging the learning of English over other contexts and other forms of multilingualism. Apart from what language is being learned, it is also argued that L2 classrooms are favored as a setting, and thus disfavor other settings in which language learning occurs. The chapter concludes with Ushioda arguing that by favoring the aforementioned contexts in which research has been conducted, an elitist bias has pervaded the practice of investigating language learning motivation throughout its history.

Chapter 5 in turn illustrates the ideologies and social forces that shape the environment of the language classroom. In particular, the ethical and moral implications of the use of motivational strategies in the classroom are discussed. Ushioda asserts that the use of motivational strategies such as personalization, ideation of an L2 self, and the use of L2 near-peer role models puts the control over students' motivation in the hands of the teacher, She then suggests this could lead the learners to become dependent on the teacher for sources of motivation. The chapter summary states that as teachers we must be careful not to decontextualize research findings and

appropriate them into motivational techniques to be used in the classroom because of the ethical and moral complexities of doing so.

Chapter 6 highlights the relational and ethical complexities and difficulties that researchers face when investigating language learning motivation. The discussion centers around the tension between the concepts of procedural ethics and ethics in practice and the potentially transformative effects that research can have on the participants' motivation. At the end of the chapter, Ushioda reminds us that any ethical agenda for research must call for a critical analysis of the relationships between the researcher and research subjects before, during, and after the research project is concluded.

In Chapter 7, the new ethical framework for language learning motivation research is outlined in detail. Ushioda argues that an ethical shift has already occurred in the larger ethical context of research, that being away from an emphasis on personal autonomy towards a greater emphasis on engaging in collective sociopolitical action. She proposes that a similar shift should happen in the field of language learning motivation. Ushioda advocates that research should be guided by philosophical pragmatism and should be value driven, not purpose driven. Finally, she argues that the underlying claims on the neutrality of practical utilitarianism amount to an implicit adoption of an ends justify the means mentality and must be rejected.

In Chapter 8, Ushioda proceeds to discuss and illustrate a framework for applying the new ethical agenda that has been introduced in the previous chapters. The discussion begins by showing how to frame inquiry in the field in a socially responsive way. Ushioda also looks at the value of teacher-researcher collaborations and teacher-led research. Additionally, a section is devoted to drawing a distinction between investigating and influencing student motivation during research procedures. Next, there is attention given to the challenges of reorienting the field to better serve marginalized and underserved groups. The chapter wraps up with a note on possible forms of ethical training and awareness for language learning motivation researchers.

As this volume primarily concerns itself with an ethical agenda tailored to the needs of the language learning motivation research community, it will primarily be of interest to researchers of motivation in language learning. However, the title of the book may mislead readers on the primary theme of the work. It is not a book about the fundamentals of language learner motivation or a complete history of the previous research conducted in the field, though some attention is given to these topics. This is a book about critical language theory, or critical theory more broadly, and the author's views on

how these should be applied to language learning motivation studies. Despite this, language teaching practitioners may also be interested in parts of the discussion pertaining to classroom practices surrounding student motivation. Personally, I found the issues around the ethical and moral implications of the use of motivational strategies in the classroom in Chapter 5 (p. 68) to be very insightful, and other teachers and language educators may as well.

Social Networks in Language Learning and Language Teaching. Avary Carhill-Poza and Naomi Kurata (Eds.). Bloomsbury Publishing, 2020. ix + 256 pp. https://doi.org/10.5040/9781350114289

Reviewed by
John Bankier
Kanagawa University

This volume brings together nine studies which address the role of social networks in language learning and teaching. As the editors describe in the first chapter, "Social Network Analysis and its Application in Applied Linguistics," the studies take a primarily sociocultural or ecological approach, seeing L2 learning as "constructed in and through interaction in situated activities" (p. 5). Although social network analysis originated as a quantitative method, the chapters draw on qualitative, quantitative, and mixed methods approaches to consider how language learners' social relationships provide opportunities for language learning and identity development. In addition to illuminating how social relationships shape language learning, most chapters include insightful and innovative pedagogical recommendations.

The volume is divided into three sections, with the first part titled "Immigrant Children and Adolescents' Social Networks in School Settings." Two chapters address the classroom networks of immigrant language learners in the United States. In Avary Carhill-Poza's chapter, "The Social Networks of Adolescent Emergent Bilinguals in High School," she found that higher academic L2-English proficiency was correlated with networks comprised of many bilingual Spanish-English speaking peers. However, Spanish-English bilingual practices were actively discouraged by classroom teachers. In their chapter "Social Networks and Patterns of Participation in Linguistically

Heterogeneous Classrooms," Amanda Kibler and her co-authors consider degrees of *linguistic integration* in multilingual classrooms, or the extent to which speakers of different L1s interact with each other. They found that active and confident participation was associated with high-integration classrooms. Pedagogical recommendations include increased training of teachers in awareness of the peer relationships, and classroom activities such as peer learning. In the third chapter, Kaya Oriyama's "Social Networks with Purpose: Heritage Language Networks of Practice Among Transnational and Transcultural Japanese Youth in Sydney," she investigated the networks and identities of Japanese youth in families which have relocated overseas for work. She found that institutional heritage language support such as Japanese language tutoring was crucial for maintaining language proficiency. She also demonstrated how the identities of these youth, moving between Japan, Australia, and other contexts, were not fixed but *transcultural*, existing on a continuum and highly context/network specific.

As the editors observe, there is a lack of research on social networks in home country language learning contexts, such as English learning in Japan. Despite this, fully one third of the volume is dedicated to such contexts, with Part II titled "Out-of-Class Social Networks of University Students in Home-Country Settings." The studies demonstrate that existing networks of ties to peers, family members, and others provide opportunities for language learning and motivation outside of the classroom contexts. In her four-year case study, "The Effects of Social Networks on L2 Experiences and Motivation: A Longitudinal Case Study of a University Student of Japanese in Australia," Naomi Kurata investigated networks and L2 motivation. The learner's motivation became increasingly integrative as she took part in festivals and cultural events, sought out ties to Japanese study abroad students, and finally developed ties to the Japanese community in Australia partly through her Japanese boyfriend. Kurata argues for greater awareness among teachers about the potential of network opportunities outside the classroom. In "Changing Informal Language Learning Networks in a Gulf Arab Community," David Palfreyman used quantitative methods to compare social network surveys of female Emirati students conducted in 2003 and 2018. He reveals how family relationships provided opportunities for English learning but suggests that societal changes in the UAE led learners to rely less on brothers and husbands and more on sisters and online sources for English support. In "How Do Social Networks Facilitate Out-of-class L2 Learning Activities?," Miho Inaba mapped the networks of learners of Japanese in Australia and Sweden through qualitative interviews and diary entries. Despite some ties

to L1 Japanese speaking acquaintances, interactions with classmates who held similar interests (e.g., anime) had a greater impact on the participants' language learning activities and motivation outside the classroom.

The final section is entitled "Social Networks in Study Abroad Contexts." Hannah Trimble-Brown and co-authors' chapter, "Implementing Mental Contrasting to Improve English Language Learning Social Networks" considers how social networks during study abroad are affected by classroom training in a self-regulation strategy. Their findings suggest that, at the start of a SA program, training students to visualize goals for L2 social interaction can encourage the formation of larger and denser social networks. The most comprehensive study in scope is "Developing Friendships or Practicing Japanese?: Differential Impacts of Language Pledge on Study Abroad Students." Atsushi Hasegawa maps the networks of Japanese learners in Japan to consider whether the program's "Japanese only" pledge and system of placing international students with Japanese roommates facilitated networks for L2 usage. He concludes that networks are more likely to develop when fewer constraints are placed on learners, encouraging translingual Japanese-English interactions and social ties which do not conform to host/guest or learner/supporter binaries. Perhaps the most interesting study from a Japanese perspective is Levi Durbridge's chapter, "Social Network Development and Language Learning in Multilingual Study Abroad Contexts: Case Studies of Japanese Adolescents." Drawn from a pool of 100 survey respondents, he investigated four Japanese learners taking part in a high school exchange program in non-English majority countries (e.g., Brazil, Finland). While the students' proficiency in the majority language did shape their social networks, key individuals (host siblings, schoolmates with an interest in Japanese pop culture) were important in facilitating interactions and social ties in the host community. He argues for greater attention in pre-departure language programs to necessary practical communication skills.

As the editors state in their "Concluding Discussion," the studies in this volume demonstrate that "social networks that support language learning do not occur easily or incidentally" (p. 238). Learners' agency to form and leverage networks is necessarily constrained by learning context, resources available, language proficiency, and positioning as "non-native speaker", among other aspects. However, the editors emphasize that language teachers have a role: teachers may raise awareness of the potential of varied networks beyond the "native-speaking friend myth" (p. 238) or arrange activities, mentoring opportunities, and online learning communities. In particular, teachers can help learners set realistic expectations for networks both overseas and in their home countries.

As well as the importance of social networks in language learning and teaching, the chapters in this volume showcase the wide diversity of approaches to networks in research. At times, this diversity is warranted, as different contexts, timeframes, and phenomena necessitate different networks. However, as someone using social networks in my own research, I would argue for greater consistency in how networks are represented in future research. While many studies in the volume draw on related frameworks, no two studies represent networks in the same way, and indeed many authors choose not to represent networks visually at all. In future studies, greater use of existing and established frameworks (e.g., *individual networks of practice* [Zappa-Hollman & Duff, 2015], *classroom networks of practice* [Bernstein, 2018]) would benefit readers who are less familiar with social network analysis and aid in comparison across contexts.

References

Bernstein, K. A. (2018). The perks of being peripheral: English learning and participation in a preschool classroom network of practice. *TESOL Quarterly, 52*(4), 798–844. https://doi.org/10.1002/tesq.428

Zappa-Hollman, S., & Duff, P. (2015). Academic English socialization through individual networks of practice. *TESOL Quarterly, 49*(2), 333–368. https://doi.org/10.1002/tesq.188

Political Economy and Sociolinguistics: Neoliberalism, Inequality, and Social Class. David Block. Bloomsbury, 2018. xiii + 227 pp. https://doi.org/10.5040/9781474281478.ch-004

Reviewed by
Lachlan Jackson
Ritsumeikan University

Professor David Block has been contributing highly influential work to the fields of sociolinguistics, second language acquisition (SLA), and applied linguistics for more than three decades. Block's (2003) work, for example, *The Social Turn in Second Language Acquisition*, is considered groundbreaking by many for its insistence that SLA researchers incorporate wider social considerations into accounts of language learning. Similarly, his *Social Class*

in Applied Linguistics (2014) was also influential in that it helped flag the previously under-researched construct of social class as an area worthy of greater attention from applied linguists. So too it is with this work, *Political Economy and Sociolinguistics*, where Block laments "the seeming unwillingness of far too many researchers to situate political economy in general, and social class in particular as central to their efforts" (p. 5). Block plants a flag in the introduction, stating that the book, for some, "will be a case of taking the social science angle to the extreme" (p. ix). The central argument of this work is that Marxist political economy provides sociolinguists with a superior alternative to more conventional post-structuralist critiques of a modern, unjust, neoliberal world. The Marxist influenced Block is unapologetically vociferous in this view when he declares "one does not take a water pistol to combat a housefire and so one cannot take on … predatory capitalism…with actions that are merely palliative" (pp. 46-47).

Structurally, this book strikes a balance between on the one hand, chapters that can be described as theoretical, and on the other, examples of detailed practical applications of this theory in real world sociolinguistic inquiry. For instance, Block provides accessible yet comprehensive theoretical explanations of political economy (Chapter 2), neoliberalism (Chapter 3), and social class (Chapter 4) before moving on to discussions of the neoliberal citizen (Chapter 5), and discursive class warfare as seen through representations of housing evictions occurring across Spain over the last decade.

Chapter 1 is titled "A Short History of Political Economy in Sociolinguistics." Here, Block makes reference to the pioneering work of Judith Irvine and Susan Gal. While sociolinguists have generally been slow to incorporate political economy into their writing, Block offers an initial summary of the political economy literature as it relates specifically to sociolinguistics. He categorizes this work as having engaged with one of five key strands: (1) the English Divide and how access to English language learning is intertwined with class position; (2) the commodification of language in the workplace; (3) the economics of language based on the classic notion of "optimal resource allocation" (p. 17); (4) language and tourism; and (5) Critical Discourse Studies/Analysis which examines the way language discursively constructs ideology and "takes on issues of inequality" (p. 24).

Then, in Chapter 2, "Political Economy: Background and Approach," Block takes a much deeper dive into the ever-evolving history of political economy, which he defines, in the widest sense, as "the science of the laws governing the production and exchange of the material means of subsistence in human society" (p. 32). Block discusses the relationship between humanism and

political economy, and introduces the reader to Global Political Economy (GPE) and International Political Economy (IPE) – both of which having emerged "as loosely organized fields of inquiry at the confluence of economics and developmental studies from the 1970s onwards" (p. 42). Finally, in this theoretically dense chapter, Block considers how sociolinguists working within a political economy framework should think about ontological (reality) and epistemological (knowledge) considerations.

"Neoliberalism: Historical and Conceptual Considerations" is the title of Chapter 3. Here, Block points out that the term neoliberalism, as the latest incarnation of a particularly pernicious, predatory capitalism, has become a "default epithet for all things that are despised as unjust" (p. 51). Neoliberalism is shown to have emerged in two stages, the first being a "roll-back stage" that saw the obliteration of the Keynesian welfare state, followed by the "roll-out stage" with the privatization of public assets and the heralding in of low-cost, non-state, service providers (pp. 53-54). A detailed account of the most influential neoliberal thinkers is provided in this chapter. Block argues that the media has been complicit in convincing the public that neoliberalism is somehow "commonsensical" (p. 56) and that this has resulted in dramatic changes to our social institutions. Schools now compete for clients, not students, for example, and domestic labour markets have suffered from first the "out-sourcing" and, later, "off-shorization" of jobs (pp. 72-74).

In Chapter 4, "Stratification, Inequality, and Social Class," some of the unsavory byproducts of neoliberalism are examined. The stratification that Block describes is a term that captures "how societies are based on different forms of differentiation" (p. 102). As with the previous chapters, Block offers a chronological perspective on our emerging understanding of social class, providing the reader with a comprehensive and contrastive panoramic survey of the major theorists, which, despite the aforementioned warning about "taking the social science angle to the extreme", will prove, for some, rather dense ground.

Chapter 5, "The Neoliberal Citizen: Conceptualizations and Contexts," is where the theoretical rubber starts to meet the sociolinguistic road. Here, Block is dealing with the "human consequences" (p. 103) of neoliberalism, namely the promulgation of the "neoliberal citizen" – the individualistic, self-branded, self-sufficient entrepreneur whose "political and social rights and duties revolve around a conformity with the 'choices' that neoliberal regimes offer – increasing precaritized jobs, flexibility imposed from above, being a good consumer to keep the economy growing, voting in elections for marketized candidates and so on" (p. 104). Block offers an interesting analy-

sis of the neoliberal citizen as represented by the characters in a French as a foreign language textbook. Also in this chapter is a discursive analysis of the 2013 Spanish law supporting entrepreneurs and a call for proposals for funding grants to show just how deeply "discourses of competitiveness and entrepreneurialism have extended into realms where they would likely not have been welcome or accepted in the past" (p. 125).

The sixth and final chapter is titled "Inequality, Class and Class Warfare: Discourse, Ideology, and Truth." Block focusses primarily on the way various stakeholders (e.g., home owners, bankers, politicians) discursively represent the increasing practice of evicting defaulting Spanish mortgagees from their homes. Drawing from Critical Discourse Studies on the notion of semiosis (i.e., the making of meaning using semiotic resources) (p. 141), and the idea of "classtalk" (Turgeon et al., 2014), Block shows how "public policy talk and the media (specifically reality shows) construct class, class relations and class warfare" (p. 143). Block also introduces us to the notion of corrupt discourses, which involve the deliberate misuse of language to win symbolic battles (pp. 157-163), and the emerging dangers for both society and democracy in a "post-truth" era (pp. 163-164).

Regardless of one's politics, readers will find David Block's *Political Economy and Sociolinguistics* a thought-provoking, impeccably-researched, and skillfully written book. Block knowledgeably guides his reader through what for some will be unfamiliar and rugged terrain, but the reward will certainly be worth the effort. Block's colleague and sometimes co-author John Gray is reported in the Epilogue to have remarked that what Block "has tried to do here is recalibrate sociolinguistics in the direction of a more thoroughgoing sociological orientation" (p. 169). Without a doubt, David Block has succeeded.

References

Block, D. (2003). *The social turn in second language acquisition*. Edinburgh University Press.

Block, D. (2014). *Social class in applied linguistics.* Routledge. https://doi.org/10.4324/9781315871141

Turgeon, B., Taylor, T., & Niehaus, L. (2014). Contrasts and classtalk: A critical discourse analysis of welfare-to-work program managers. *Discourse & Society, 25*(5), 1–16. https://doi.org/10.1177/0957926514536827

JALT Journal Aims and Scope

JALT Journal is a bi-annual, Scopus-approved research journal of the Japan Association for Language Teaching (全国語学教育学会). JALT's larger mission is to support the research programs and professional development of JALT members, promote excellence in language learning, teaching, and research, and provide opportunities for those involved in language education. In line with this mission, *JALT Journal* publishes high-quality English- and Japanese-language, quantitative and qualitative, theoretically-informed and empirically-grounded studies of relevance to second/foreign language education in Japan. Although emphasis is placed on the Japanese context, *JALT Journal* values contributions which also transcend geographical boundaries to illuminate the complex interaction between language, language use, people, education, and society across cultural and socio-political contexts.

When possible, submissions to *JALT Journal* should aim to be both descriptive (*What is my data?*) and explanatory (*Why is my data like this and not otherwise?*) in purpose, and further stimulate scholarly debate, to hopefully improve existing applied linguistic scholarship around the world. Areas of interest include but are not limited to the following:

- Bilingualism and multilingualism
- Classroom-based language education
- Cognitive linguistics
- Contrastive linguistics
- Conversation/discourse/critical discourse analysis
- Critical language pedagogy
- Curriculum design and teaching methods
- Intercultural communicative competence
- Language acquisition/learning
- Language policy and planning
- Language testing/evaluation
- Phonetics
- Pragmatics
- Psycholinguistics
- Semantics
- Sociolinguistics
- Syntax
- Teacher training
- Translation and interpretation
- Vocabulary

Guidelines

Style

Authors are encouraged to submit manuscripts in five categories: (1) full-length articles, (2) short research reports (*Research Forums*), (3) theory-grounded essays which may include analysis of primary or secondary data (*Perspectives*), (4) comments on previously published *JALT Journal* articles (*Point-to-Point*), and (5) book/media reviews (*Reviews*) either requested by the Book Reviews editor or suggested by the author. Articles should be written for a general audience of language educators; therefore, statistical techniques and specialized terminologies must be clearly explained and their use clearly justified.

Authors are responsible for obtaining permissions for any copyrighted material included in their manuscript submission. When submitting a manuscript based on a thesis or dissertation published in an institutional repository, authors are responsible for checking with the copyright holder for permission to publish portions of the original text.

JALT Journal follows the *Publication Manual of the American Psychological Association, 7th edition* (available from <https://apastyle.apa.org/products/publication-manual-7th-edition>). A downloadable copy of the *JALT Journal* style sheet is also available on our website at <https://jalt-publications.org/jj/>. Instructions for online submissions can also be found at this website.

Format

Full-length articles must not be more than 8,000 words, including references, notes, tables, and figures. *Research Forum* submissions should not be more than 4,000 words. *Perspectives* submissions should not be more than 5,000 words. *Point-to-Point* comments on previously published articles should not be more than 1,000 words in length, and *Reviews* should generally be around 2,000 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 and for obtaining any permissions for copyrighted material contained in the manuscript.

Submission Procedure

Please submit the following two documents 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 the corresponding author.
2. A blinded manuscript, including title, abstract, and keywords, with no citations referring to the author. Do not use running heads and do not include your biographical information. 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 them 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. The first round of review 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. For further information, see the JALT Publications Statement of Ethics and Malpractice at: https://jalt-publications.org/jalt-publications-statement-ethics-and-malpractice

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

Please upload submissions in these categories to the following website:

https://jalt-publications.org/content/index.php/jj

Manuscripts should follow the **American Psychological Association (APA) 7th Edition style**. Please indicate if your submission is a 1) Full-length article; 2) *Perspectives* article; 3) *Research Forum* article or 4) *Point-to-Point* submission.

For any general inquiries about English-language submissions, please contact:

Dennis Koyama, *JALT Journal* Editor
jaltpubs.jj.ed@jalt.org

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:

Kiwamu Kasahara, *JALT Journal* Japanese-Language Editor
jaltpubs.jj.ed.j@jalt.org

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* and can be found online in each issue at <https://jalt-publications.org/tlt/>. Review authors receive one copy of the *Journal*. *JALT Journal*'s latest *Reviews* guidelines can be found in Melodie Cook's *Expositions* piece published in the November 2023 issue. Please send submissions, queries, or requests for books, materials, and review guidelines to:

Melodie Cook, Incoming *JALT Journal* Reviews Editor
jaltpubs.jj.reviews@jalt.org

Special Issues

Special issues often make an important contribution to the development of academic discourse in a specific field, because they allow researchers and practitioners to (a) identify an issue or topic of particular relevance to the context in which the journal is read, (b) summarize the key concepts and debates shaping that issue, (c) bring further sophistication to existing academic discourse and identify new research possibilities, and (d) identify key readings for the journal readership. Special issues can also attract new authors and readers to an academic journal, and can be an effective means of finding new editors for that journal.

We strongly encourage *JALT Journal* readers to submit special-issues proposals. When submitting such proposals, please make sure that they adhere to the aims and scope of *JALT Journal*. Proposals should include: (1) a title which clearly captures the special issue topic, (2) a brief description of the special issue, (3) an account of the motivation behind the special issue and its importance to the field at large, (4) a list of no more than three guest editors with short biographical information, including editorial work experience, and (5) a list of article contributors, with a short description of each article contribution.

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

JALT Central Office
Marunouchi Trust Tower Main Building 20F
1-8-3 Marunouchi, Chiyoda-ku, Tokyo 100-0005 JAPAN
Tel.: (+81) 3-5288-5443
Email: jco@jalt.org URL: https://jalt.org

日本語論文投稿要領

*JALT Journal*では日本語で執筆された（a）論文、（b）研究報告、（c）展望論文、（d）*JALT Journal*に掲載された著作物へのコメント・考察、（e）書評を募集しています。（a）論文と（b）研究報告の違いは、以下の通り字数制限による違いです。（c）展望論文は、言語教育研究に関する課題に焦点をあてた短い論文で、先行研究の検証、理論や1次2次データに基づく議論などを含むものです。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*（7th edition）の定める方式に合わせて下さい。*JALT Journal*書式シート（日本語原稿用）を以下からダウンロードできます<https://jalt-publications.org/jj/>。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、（a）論文は25,000字、（b）研究報告は13,000字、（c）望論文は16,000字、（d）*JALT Journal*に掲載された著作物へのコメント・考察は2,000字、（e）書評は1,500~3,000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:
*JALT Journal*書式シート（日本語原稿用）を参考に作成の上、電子メールの添付書類でお送りください。　なお、上記（a）論文~（e）書評のどのカテゴリーへの投稿かを明記ください。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願います。

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:

070-8621 北海道旭川市北門町9丁目　北海道教育大学旭川校
JALT Journal 日本語編集者　笠原　究
電話: 0166-59-1273
jaltpubs.jj.ed.j@jalt.org

JALT Journal 第 45 巻　第 2号

2023年 10月20日　印刷
2023年 11月1日　発行
編集人　小山デニス
発行人　クレア・カーネーコー
発行所　全国語学教育学会事務局
〒100-0005 東京都千代田区丸の内1-8-3 丸の内トラストタワー本館20階
TEL（03) 5288 5443
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

JALT2023
GROWTH
MINDSET
in language education

JALT2023

全国語学教育学会
第49回年次国際大会教材展示会

2023年11月24日～2023年11月27日

茨城県つくば市 つくば国際会議場

49th Annual International Conference on
Language Teaching and Learning & Educational
Materials Exhibition

November 24 - 27, 2023

Tsukuba International Congress Center
(Epochal Tsukuba), Tsukuba, Ibaraki, Japan