jalt journal

The research journal of the Japan Association for Language Teaching

Volume 38 • No. 1 • May 2016

JALT

全国語学教育学会

Japan Association for Language Teaching

¥1,900 ISSN 0287-2420

JALT Journal

Volume 38 • No. 1

May, 2016

Contents

May 2016
Volume 38 • No. 1

Japan Association for Language Teaching

A Nonprofit Organization

The Japan Association for Language Teaching (JALT) is a nonprofit professional organization dedicated to the improvement of language teaching and learning in Japan. It provides a forum for the exchange of new ideas and techniques and 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. There are 34 JALT chapters and 26 special interest groups (SIGs). JALT is a founder of PAC (Pan-Asian Consortium), which is an association of language teacher organizations in Pacific Asia. PAC holds annual regional conferences and exchanges information among its member organizations. JALT is the Japan affiliate of International TESOL (Teachers of English to Speakers of Other Languages) and a branch of IATEFL (International Association of Teachers of English as a Foreign Language). JALT is also affiliated with many other international and domestic organizations.

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

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

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

JALT National Officers, 2016

President: . . . . *Caroline Lloyd*
Vice President: . . . . *Richmond Stroupe*
Auditor: . . . . *Aleda Krause*
Director of Treasury: . . . . *Kevin Ryan*
Director of Records: . . . . *Nathaniel French*
Director of Program: . . . . *Steve Cornwell*
Director of Membership: . . . . *Fred Carruth*
Director of Public Relations: . . . . . . *Ted O'Neill*

Chapters

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

Special Interest Groups

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

JALT Central Office

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

In This Issue

Articles

Assessment is a subject that education professionals in Japan are always talking about, and we are pleased to present two articles on formative assessment in this issue. In our feature article, **Christopher Weaver** looks at the formative role of the TOEIC IP test as a placement test. In a Perspectives piece, **Paul Wicking** examines the role that formative assessment can play in the formation of global human resources.

Reviews

The May 2016 issue brings reviews of titles from five different publishers on a range of topics from motivation and autonomy to critical thinking, EAP, curriculum design, and even one on Japanese linguistics. To open, **Matthew T. Apple** covers Ema Ushioda's edited volume on motivation in a variety of international contexts. In the second review, **James Crocker** looks at how critical thinking is understood, taught, and constructed in disciplinary discourses. In collaboration with fellow JALT Learner Development SIG member Jim Ronald, **Huw Davies** examines how autonomy is viewed in context from a social dimension. **Rick Derrah** contributes a review of a title of interest to those involved in curriculum design. **Mark Rebuck** then tackles a comprehensive overview of Japanese linguistics with supplementary multimedia resources and online exercises provided by Cambridge University Press. **Gregory Strong** critiques de Chazal's *English for Academic Purposes*. **Blake Turnbull** wraps up the book reviews with a look into translanguaging and the notion of a shift from language as a *product* to language as a *practice* in language education.

From the Editor

Every issue of *JALT Journal* is completed with the help, support, and forbearance of a large number of people, but this issue was truly exceptional for the group effort expended by the authors, reviewers, and members of the Pubs team. I'm extremely grateful as always to the many reviewers who take the time to write thoughtful reviews. I also want to thank the authors who stick it out through the very long review process. Special and particular thanks to Greg Scholdt, the Consulting Editor who did extra work for this issue, and as ever to Aleda Krause, the Production Editor, and her wonderful team.

James A. Elwood joins the Editorial Advisory Board from this issue. Jim has served as a capable, reliable, and good-humored additional reader for several years. With this issue we also welcome Eric Hauser as our Associate Editor. Eric has only been on the job a few months but has already proved himself willing to jump in and help with editorial business. I am anticipating relying on him more and more as time goes on.

Anne McLellan Howard

Articles

The TOEIC IP Test as a Placement Test: Its Potential Formative Value

Christopher Weaver
Toyo University

The use of the Test of English for International Communication Institutional Program (TOEIC IP) at postsecondary institutions in Japan has been increasing over the past decade. The revised TOEIC test provides test-takers and administrators with 3 levels of information: total score, listening and reading sectional scores, as well as the percentage of correct responses achieved on 4 listening subskills and 5 reading skills. Using a hierarchical cluster analysis, this study found significant differences between placement decisions for 1,524 university business students in an English for Specific Purposes program based upon their TOEIC IP scores and their 9 TOEIC subskill scores. The author also discusses some of the diagnostic shortcomings of using a standardized proficiency examination for placement purposes.

日本の高等教育機関におけるTOEIC IPの英語テストの使用は、ここ過去10年増加し続けている。改訂版TOEICテストでは受験者とテスト使用者に合計スコア、リスニング、リーディングという3つのスコアに加え、4つのリスニングサブスキルと5つのリーディングスキルの正答率が提供される。本研究では、階層的クラスター分析を用いて、1,524名の経営学を選考する大学生を対象に調査を行い、TOEIC IPテストのスコアに基づいた英語のクラス分けと、TOEICの９つのサブスコアに基づいた英語のクラス分けに有意な差があることを確認した。また、クラス分けの目的で標準化された能力試験を利用する際の診断的な問題点を指摘する。

The use of the Test of English for International Communication Institutional Program (TOEIC IP) continues to increase in education institutions across Japan. The number of TOEIC IP test-takers reached 1,287,456 in 2014 compared to 698,000 test-takers in 2001 (Institute for International Business Communication [IIBC], 2015b). Educational institutions such as high schools, junior colleges, universities, and vocational schools accounted for 45.8% of the TOEIC IP test-takers (589,191 students) in 2014. As a result, teachers and administrators are increasingly facing the need to understand what a TOEIC score means for their students and how it may inform English language programs. Previous empirical investigations have highlighted the importance of curriculum (Hisatsune, 2007), textbook design (Uchibori, Chujo, & Hasegawa, 2006), student study habits (Falout, 2006; Mizumoto & Takeuchi, 2009), and teacher training (Boldt & Ross, 2005) to help facilitate score gains on the TOEIC test. There is, however, a gap in the literature concerning how TOEIC scores can be used to inform placement decisions in a postsecondary EFL program. Of special interest in this paper is the potential for TOEIC IP scores to provide formative feedback to test-takers and administrators.

Placement Examinations

The primary purpose of a placement examination is to create student groups of relatively homogeneous language abilities (Brown, 1996). Typically there are two types of placement examination used to fulfill that purpose (Wall, Clapham, & Alderson, 1994). One type has a preachievement orientation such as the English Placement Test at the University of Illinois at Urbana-Champaign that reflects the academic demands of the courses offered at the university. The other type is standardized tests such as TOEFL, TOEIC, and IELTS that have a general proficiency orientation with no direct relationship to the content of the courses in which test-takers are placed. The use of standardized proficiency examinations has become more common at postsecondary institutions because of the need to evaluate an increasing number of students applying from overseas (Kokhan, 2012; Mullen, 2009). Standardized proficiency examinations are also typically a convenient, cost effective, and rapid means of placing test-takers on the same score scale, which in turn allows relatively easy comparisons and placement of students. Despite these advantages, an increasing number of researchers have become critical of the disconnect between the communicative competence required in different academic settings and the scores achieved on standardized proficiency tests (Chapman & Newfields, 2008; Fox, 2009; Kokhan, 2013).

Central to the argument against the use of standardized proficiency tests for placement purposes is that tests are usually designed to assess a range of knowledge, skills, and abilities within set target domains of language use by a target population (Fulcher & Davidson, 2009). In the case of the TOEIC test, the intention of the test is to measure the everyday English skills of people working in an international environment (Educational Testing Service, 2013). In the case of using the TOEIC test in postsecondary institutions, test-takers may have limited work experience and thus be unfamiliar with the communicative situations featured in this test. Moreover, students' academic interests and ultimate career goals might be quite distant from the world of international business.

The development of the TOEIC test is another cause for concern. The validation process used to evaluate the design of the TOEIC test and the interpretation of its test scores has largely been done in the context of developing a measure of general language proficiency (In'nami & Koizumi, 2012; Tannenbaum & Wylie, 2008) and not a placement examination. As a result, the use of the TOEIC test for a purpose that was not intended is problematic. However, in certain contexts such as in a faculty of business administration or in other postsecondary programs, a test may not only be a practical placement examination but also a valid means of placing university students into an English for Specific Purposes program (IIBC, 2015a). It must be remembered that validity is not a property of the test, but rather concerns the meaning of test scores and the implications for action based upon the interpretation of test scores (Cronbach, 1971; Messick, 1996). From this perspective, placement examination scores should provide test-takers and administrators with diagnostic information about test-takers' strengths and weaknesses in order to group students of similar ability together so that they may receive appropriate materials and instruction (Green & Weir, 2004). This approach to placement examinations, however, relies upon a clear understanding of what a test score means.

Test Scores and Interpretations of Test Scores

A test score describes the interaction between test-takers' performance on the items on a test and the types of knowledge that these items are thought to assess. A test score can be norm-referenced by specifying the relative rankings of the test-takers; it can be criterion-referenced by providing a description of the tasks that test-takers can or cannot perform; or it can be both. The TOEIC test is an attempt to provide test-takers and administrators with information from both of these perspectives. It should be remembered, however, that any test score includes measurement error depending upon the quality of the test. For example, the standard error of measurement for the TOEIC listening and reading sections is 25 scaled points. Thus if a test-taker has a scaled score of 300 on the listening section of the TOEIC test, 68% of the time this individuals' true score will vary between 275 and 325 (Educational Testing Service, 2013). The standard error of measurement of the TOEIC test may also differ between the listening and reading sections of the test as well as between different populations of test-takers (Zhang, 2006).

Norm-Referenced Information From TOEIC Scores

The official score certificate of the TOEIC test provides test-takers with a total score ranging from 10 to 990, which is the summation of two scaled subscores ranging from 5 to 495 for the listening and reading sections of the test (Educational Testing Service, 2013). Takers of the TOEIC Secure Program (TOEIC SP) test also receive a percentile ranking indicating the percentage of global TOEIC SP test-takers who have a lower scaled score than theirs on the listening and reading sections of the TOEIC test (Educational Testing Service, 2012). This information, however, is not available for TOEIC Bridge and TOEIC IP test-takers (IIBC, 2015b).

Test-takers and administrators need to consult the TOEIC data and analysis report issued annually by the IIBC to understand the relative ranking of test-takers, apart from percentile rankings. For example, in 2014, university students in Japan scored an average of 564 on the TOEIC SP test and 440 on the TOEIC IP test (IIBC, 2015b). Test-takers and administrators can also compare scores against others in the same field of study. In 2014, university students studying commerce, economics, or finance had an average TOEIC IP score of 410 in their first year, 425 in their second year, 476 in their third year, and 516 in their fourth year of study. This seemingly upward trend of TOEIC IP scores, however, may be an artifact of test-taker selection. In 2014, 33,337 first-year students, 18,749 second-year students, 8,643 third-year students, and 2,343 fourth-year students took the TOEIC IP test. The decrease in the number of test-takers each year means that making norm-referenced inferences using TOEIC scores should be carried out with due caution.

Criterion-Referenced Information From TOEIC Scores

In an attempt to help test-takers and administrators interpret TOEIC scores, the Educational Testing Service, the Chauncey Group, and the IIBC undertook a number of initiatives to provide criterion-referenced information for TOEIC scores. TOEIC representatives in Japan developed a 5-level proficiency letter-scale (ranging from A to E) based upon anecdotal information and informal observations on the relationship between TOEIC scores and general English language proficiency. Test-takers and administrators can also consult the score descriptors for the listening and the reading sections to learn about the typical strengths and weaknesses test-takers have. For example, a test-taker with a TOEIC listening score of 250 can typically make simple inferences based on a limited amount of text but cannot make inferences requiring paraphrasing or connecting information (Educational Testing Service, 2007).

To further clarify the relationship between TOEIC scores and the likelihood of being able to perform specific tasks in English, the Educational Testing Service (2000) introduced the *TOEIC Can-Do Guide*. Based upon a correlation analysis of TOEIC IP scores and self-reports of 8,601 Japanese employees (Tannenbaum, Rosenfeld, & Breyer, 1997), the *TOEIC Can-Do Guide* provides a list of different daily life and basic job activities that can be performed at three levels of proficiency: can do, can do with some difficulty, and cannot do. For example, test-takers with a TOEIC reading score ranging from 230 to 350 are thought to be able to understand the type of store or the type of service offered by reading the storefront and to read and understand an agenda for a meeting with some difficulty, but they are thought not to be able to identify differing opinions that opposing party politicians give in two newspaper interviews. Following a similar methodological approach, Ito, Kawaguchi, and Ota (2005) developed a can-do list focusing upon 65 job-related tasks occurring in seven different communicative situations. Based upon the responses of 8,386 Japanese company employees, TOEIC scores ranging from 400 to 495 were reported as the point where people began feeling comfortable about using English to do job-related tasks such as reading a manual written in English about office equipment or sending an email to a company to complain about a product.

The Redesigned TOEIC Test and Scores

One of the main difficulties in interpreting TOEIC scores is that test-takers and administrators need to consult outside resources that attempt to link TOEIC scores to certain communicative competencies, the Language Proficiency Interview test (Wilson, 1989), or the Common European Framework of Reference for languages (Tannenbaum & Wylie, 2008). As a response to this difficulty, one of the goals for the redesigned TOEIC test was to provide more specific information about test-takers' abilities based upon their test performance (Liao, 2010). To do so, the designers of the revised TOEIC test used the Evidence-Centered Design (ECD) method. ECD is an evidentiary reasoning approach to assessment design that (a) identifies the important domains of knowledge to be assessed and ascertains how this knowledge is acquired and used; (b) establishes a chain of reasoning between what individuals say and do in assessments and the inferences that can be made about what test scores mean in terms of the abilities that individuals currently possess and what should be done next; and (c) ensures that test design reflects the purpose of the assessment while taking into consideration the constraints, resources, and conditions of use (Mislevy, Almond, & Lukas, 2003). These three premises guided the new item prototypes and pilot testing of the revised TOEIC test to ensure a sufficient distribution of items that could provide reliable support for claims about test-taker abilities (Schedl, 2010). The abilities measured in the TOEIC test (shown in Table 1) are articulated in four listening subskills and five reading subskills.

Table 1. TOEIC Test Subskills

Listening subskills	
L1	Can infer gist, purpose, and basic context based on information that is explicitly stated in short spoken texts.
L2	Can infer gist, purpose, and basic context based on information that is explicitly stated in extended texts.
L3	Can understand details in short spoken passages.
L4	Can understand details in extended spoken texts.
Reading subskills	
R1	Can make inferences based on information in written texts.
R2	Can locate and understand specific information in written texts.
R3	Can connect information across multiple sentences in a single written text and across texts.
R4	Can understand vocabulary in written texts.
R5	Can understand grammar in written texts.

(Educational Testing Service, 2008)

These ECD-based claims about test-takers' listening and reading abilities give rise to a situation in which the TOEIC subskill scores may potentially provide test administrators with more detailed information that could enable them to make more nuanced placement decisions in a postsecondary EFL education setting. From the perspective of formative assessment, teachers and program administrators may be able to identify specific gaps in communicative competence that are exposed in test-takers' performance on the test (Long & Norris, 2000). This possibility warrants an investigation into the use of TOEIC scores and TOEIC subskill scores for placement purposes.

The following research questions guided this investigation:

RQ1. To what extent do test-takers' TOEIC IP scores relate to the average percentage of correct responses they achieved on the nine TOEIC subskills?

RQ2. To what extent can test-takers be grouped together according to the percentage of correct answers they achieved on the nine TOEIC subskills?

RQ3. To what extent do test-taker groupings based upon TOEIC IP scores differ from test-taker groupings based upon the percentage of correct answers achieved on the nine TOEIC IP subskills?

The first research question examines the degree to which the nine TOEIC subskills provide unique test-taker information compared to the total TOEIC IP score. If the subskills are highly related to TOEIC IP score or the sectional scores, then there is little reason to spend extra time and resources, which are typically in short supply in placement situations, to examine the test-takers' performances on the nine TOEIC subskills. However if the nine TOEIC subskills provide unique information about the test-takers, is it then possible to create student groups more sensitive to test-takers' abilities? This is the focus of the second research question. The final research question examines the practical effect of student placements based upon TOEIC scores and the test-takers' performance on the nine TOEIC subskills. In other words, does more refined test-taker information really make a difference when it comes to actually placing them into a postsecondary EFL program?

Participants

The test-takers in this study were 1,524 university students (883 men and 641 women) studying in the faculty of business administration at a private university located in Tokyo during the 2012-13 academic year. This population represents a fraction of the TOEIC IP test-takers in Japan and this needs to be kept in mind when considering the generalizability of the findings of this study. The test data originated from a TOEIC IP test administered at the university by a group of trained proctors from the IIBC in the fall semester. The TOEIC IP test was a requirement for the students to receive credit for a required English course. The results of the TOEIC IP test were also used to place the students into their subsequent English courses.

Analysis

To answer the first research question, Pearson correlation coefficients between the TOEIC IP score, the listening section test scores, the reading section test scores, and the percentage of correct responses achieved on the nine TOEIC subskills were examined. The magnitude of these correlations was also evaluated using the reduction in uncertainty (RiU) index (Dorans, 2004). This statistic measures the amount of certainty that the total test score and the scores from the listening and the reading sections are similar to the percentage of correct responses achieved on the nine TOEIC subskills. Dorans suggested that a 50% RiU in one observed score from another observed score is needed to confirm a linkage between the test scores.

To answer the second research question, a hierarchical cluster analysis was used to classify the test-takers into groups based upon the percentage of correct answers they achieved on the nine subskills. Similar to a factor analysis, a cluster analysis examines the interrelationships between the variables; however, there are no preconceived ideas about the composition of the groups in a cluster analysis. The formation of the clusters is informed by the analysis of the data (Everitt, Landau, & Leese, 2001) and thus can help identify individual differences that exist between language learners (Skehan, 1989). Once the clusters were identified, a series of univariate one-way ANOVAs were used to determine which subskills significantly differed between the groups of test-takers. To offset the chances of a Type I error (Simes, 1986), a Bonferroni adjustment was used to set a p value of 0.005 for this study.

To answer the third research question, Cramer's V was used to determine the degree of agreement between the test-taker groupings based upon their total TOEIC IP scores and the test-taker groupings based upon the cluster analysis of the nine subskills. This nonparametric statistic measures the degree of strength between categorical variables that have more than two categories (Sheskin, 2007). A Cramer's V of 0 reflects complete independence between the categorical variables, whereas a Cramer's V of 1 indicates a complete association or dependence between the variables.

Results

Table 2 shows that the mean TOEIC IP score for the test-takers in this study was 352.32 with a standard deviation of 91.91. These test-takers had a higher level of listening proficiency (*M* = 205.89, *SD* = 53.59) compared to their reading proficiency (*M* = 146, *SD* = 48.24). These test-takers had lower test scores compared to other university students who took the TOEIC IP test during the same time period in Japan (i.e., a mean score of 433 for the TOEIC IP test, 245 for the listening section, and 188 for the reading section), according to the 2012 TOEIC analysis report (IIBC, 2013).

Table 2. Test-takers' TOEIC IP Scores, Listening and Reading Section Scores, and the Percentage of Correct Responses Achieved on the Nine TOEIC Subskills (*N* = 1,524)

	M	*SD*	Min	Max	Kurtosis	*SEM*
TOEIC IP	352.32	91.91	150	765	0.53	0.13
Listening section	205.89	53.59	60	430	0.29	0.13
Reading section	146.43	48.24	30	365	0.71	0.13
Listening subskill L1	58.05	12.59	5	95	0.08	0.13
Listening subskill L2	46.46	14.11	8	92	-0.22	0.13
Listening subskill L3	60.35	14.93	5	95	0.05	0.13
Listening subskill L4	42.91	12.31	11	89	0.04	0.13
Reading subskill R1	31.98	13.78	0	88	0.32	0.13
Reading subskill R2	40.19	14.61	0	90	-0.21	0.13
Reading subskill R3	30.60	12.10	0	88	0.73	0.13
Reading subskill R4	34.74	10.59	7	72	-0.02	0.13
Reading subskill R5	48.70	13.95	4	92	-0.23	0.13

Note. Test scores for the TOEIC IP test, the listening section, and the reading section are reported as scaled points, whereas scores for the TOEIC subskills are reported as a percentage of correct answers. *SEM* = standard error of measurement.

In terms of the TOEIC listening subskills, these test-takers had the highest percentage of correct answers (60.35%) on items designed to assess the understanding of details in a short passage (L3) and the lowest percentage of correct answers (42.91%) on test items designed to assess the understanding of details in extended spoken texts (L4). For the TOEIC reading subskills, the test-takers had the highest percentage of correct answers (48.7%) on items designed to assess the understanding of grammar in written texts (R5) and the lowest percentage of correct answers (30.6%) on test items designed to assess the ability to connect information across multiple sentences in a single written text and across texts (R3).

Tables 3 and 4 show the Pearson correlation coefficients and magnitudes for the TOEIC IP score, the listening and reading section scores, and the percentage of correct responses achieved on the nine TOEIC subskills. The correlations between the total test score and the sectional scores are relatively high, with the highest correlation being between the total test score and the listening section score (.91). The correlations between the percentage of correct responses achieved on the nine TOEIC subskills, the TOEIC IP score, and the sectional scores for listening and reading are small to moderate. The listening subskill showing the highest correlation with the total test score is L4 (can understand details in extended spoken texts) at .77 with an RiU of 36% between these two scores. The reading subskills showing the highest correlations with the total test score are R5 (can understand grammar in written texts) at .68 with an RiU of 26% and R2 (can locate and understand specific information in written texts) at .67 with an RiU of 26%.

A hierarchical cluster analysis using Ward's minimum variance method with the squared Euclidean distance technique (Szekely & Rizzo, 2005) was run on the percentage of correct answers achieved by the test-takers on the nine TOEIC subskills. In the case of this analysis, a four-cluster solution was the point at which subsequent clusters did not significantly add to the process of distinguishing the test-takers from each other (Burns & Burns, 2008). Figure 1 shows how the four clusters of test-takers performed on the nine subskills.

Table 3. Correlations Among TOEIC IP (T) Score, Listening (L) and Reading (R) Section Scores, and the Percentage of Correct Responses Achieved on the Nine TOEIC Subskills

	L	R	L1	L2	L3	L4	R1	R2	R3	R4	R5
T	.91	.89	.63	.73	.67	.77	.52	.67	.53	.59	.68
L		.63	.68	.80	.74	.85	.35	.49	.35	.38	.50
R			.45	.51	.45	.52	.59	.73	.63	.69	.74
L1				.39	.45	.43	.26	.32	.24	.31	.37
L2					.46	.58	.28	.42	.30	.31	.39
L3						.47	.23	.34	.22	.26	.40
L4							.30	.42	.30	.31	.39
R1								.42	.47	.21	.23
R2									.40	.28	.36
R3										.33	.36
R4											.44

Table 4. Reduction in Uncertainty (RiU) Among the Total TOEIC IP (T) Score, the Listening (L) and Reading (R) Section Scores, and the Percentage of Correct Responses Achieved on the Nine TOEIC Subskills

	T	L	R
L1	22%	26%	11%
L2	32%	39%	14%
L3	25%	33%	10%
L4	36%	47%	15%
R1	14%	6%	20%
R2	26%	13%	32%
R3	15%	6%	22%
R4	19%	8%	28%
R5	26%	13%	33%

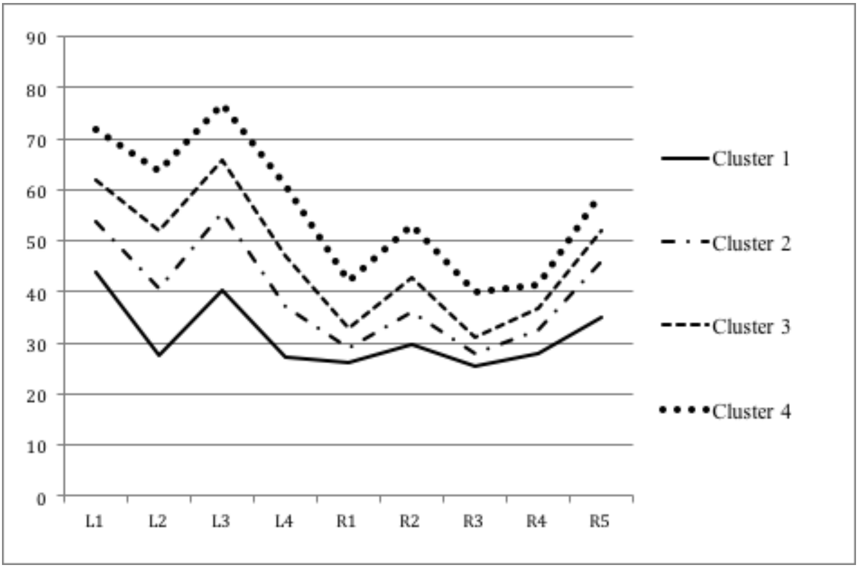

***Figure 1*. The percentage of correct responses on the nine TOEIC subskills for the four test-taker clusters.**

In regards to the TOEIC listening subskills, the four clusters of test-takers followed a similar pattern. They had a higher percentage of correct responses on items designed to assess their ability to infer gist, purpose, and basic context based on information that is explicitly stated in short spoken texts (L1) and their ability to understand details in short spoken passages (L3). In contrast, test items featuring extended spoken texts (subskills L2 and L4) were more challenging.

For the five TOEIC reading subskills, the four clusters of test-takers had some distinctive characteristics. The proficiency level of Cluster 1 was relatively flat across the five reading subskills compared to the other clusters of test-takers. Cluster 2 shared Cluster 1's difficulty on items designed to assess the ability to make inferences based on information in written texts (R1) and to assess the ability to connect information across multiple sentences in a single written text and across texts (R3). Clusters 3 and 4 followed a similar pattern with higher rates of success on items designed to assess the ability to locate and understand specific information in written texts (R2) and items assessing the ability to understand grammar in written texts (R5).

Nine univariate one-way ANOVAs found that the four clusters significantly differed from each other in terms of the percentage of correct responses they achieved on the nine subskills. The right column of Table 5 shows that only two out of the 45 possible post hoc pairwise comparisons were not significantly different (i.e., Clusters 1 and 2 are not significantly different for the TOEIC reading subskills R1 and R3). The results of the Tukey post hoc test comparisons were reconfirmed with the Games-Howell procedure (Wilcox, 1987) because there was an unequal number of test-takers in each cluster.

Table 5. The Percentage of Correct Answers Achieved on the TOEIC Subskills for the Four Clusters

TOEIC subskills	Cluster 1 (*n* = 187)	Cluster 2 (*n* = 589)	Cluster 3 (*n* = 520)	Cluster 4 (*n* = 228)	Post hoc comparison of clusters
L1	44.65 (11.51)	54.67 (10.36)	61.43 (9.96)	70.04 (10.08)	1-2, 1-3, 1-4, 2-3, 2-4, 3-4
L2	30.14 (10.37)	42.71 (11.42)	49.77 (11.07)	62.02 (10.21)	1-2, 1-3, 1-4, 2-3, 2-4, 3-4
L3	39.49 (12.34)	57.32 (11.78)	64.89 (10.38)	74.96 (11.01)	1-2, 1-3, 1-4, 2-3, 2-4, 3-4
L4	30.26 (8.09)	38.87 (9.62)	45.53 (9.31)	57.7 (10.75)	1-2, 1-3, 1-4, 2-3, 2-4, 3-4
R1	25.61 (8.81)	25.49 (11.14)	34.60 (10.89)	47.97 (14.18)	1-3, 1-4, 2-3, 2-4, 3-4
R2	28.24 (10.01)	33.52 (11.26)	45.28 (12.12)	55.57 (12.46)	1-2, 1-3, 1-4, 2-3, 2-4, 3-4
R3	24.66 (11.25)	25.31 (9.03)	33.20 (9.88)	43.22 (12.91)	1-3, 1-4, 2-3, 2-4, 3-4
R4	26.57 (8.04)	32.33 (8.86)	36.60 (10.10)	43.42 (10.43)	1-2, 1-3, 1-4, 2-3, 2-4, 3-4
R5	31.17 (10.27)	45.22 (10.65)	53.18 (11.27)	61.48 (11.40)	1-2, 1-3, 1-4, 2-3, 2-4, 3-4

Note. *M* (*SD*); all post hoc tests assessing differences between student clusters were set to $p < .005$.

The degree of agreement between test-taker groupings based upon their TOEIC IP scores and the percentage of correct answers they achieved on the nine TOEIC subskills was large, a Cramer's V of .63 (Cohen, 1988). Table 6 shows the crosstabulation table for these two categorical groupings. Test-takers with the lowest test scores (in the 100s) and the highest scores (in the 500s and 600s) generally grouped together in Cluster 1 and Cluster 4 respectively. The majority of test-takers with test scores ranging from the 200s to the 400s belonged to Clusters 1, 2, or 3.

Table 6. Crosstabulation Table of TOEIC IP Scores and TOEIC Subskill Clusters

TOEIC IP score	TOEIC subskill clusters				
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total
600s	0	0	0	20	20
500s	0	**0**	**2**	73	75
400s	0	9	187	133	329
300s	**1**	345	323	**2**	671
200s	136	233	**8**	0	377
100s	50	**2**	0	0	52
Total	187	589	520	228	1,524

Note. Bold numbers denote test-takers who were outliers in their cluster membership.

To check that clusters did not simply reflect a score difference of 50 scaled TOEIC points, the cluster memberships of test-takers with scores ranging from 300 to 345 and of those of test-takers with scores ranging from 350 to 395 were examined. In Cluster 2, there were 249 test-takers with scores between 300 and 345 and 96 test-takers with scores between 350 and 395. A similar split occurred in Cluster 3: 88 test-takers had scores between 300 and 345; 235 test-takers had scores between 350 and 395. In sum, about 28% of the test-takers belonged to a different cluster than did other test-takers with a similar TOEIC IP score.

Table 6 also shows that 15 test-takers were outliers in their cluster memberships (highlighted in bold). In Cluster 1, there was one test-taker with a score in the 300s who correctly answered 30% of the items for all five reading subskills. Cluster 2 had two test-takers who had a higher percentage of correct answers on items designed to assess the ability to understand vocabulary (R4) and grammar (R5) in written texts. In Cluster 3, there were eight test-takers with a score in the 200s who had significantly higher scores and two test-takers with a score in the 500s who had significantly lower scores on the five reading subskills. Cluster 4 had one test-taker with a score in the 300s who had significantly higher scores on the four listening subskills and another test-taker who had significantly higher scores on the five reading subskills.

Discussion

Understanding the meaning of test scores is central to well-informed placement decisions. In the case of the TOEIC IP test, interpretations of the total test score and the listening and reading section test scores can be clarified with the percentage of correct answers test-takers achieved on the four listening and five reading subskills. In this study, the correlations between the test-takers' TOEIC IP scores and the percentage of correct responses they achieved on the nine TOEIC subskills were all moderately correlated. Yet, the magnitude of these correlations is below Dorans's 50% recommendation for the RiU index. These low percentages are understandable considering that the nine subskills are used together to calculate the test-takers' subsection and total test scores. However, the subskills that have larger percentages on the uncertainty index highlight competencies that may benefit from subsequent EFL instruction. For example, the stronger link between the listening section test scores, listening subskill L2 (the ability to infer gist, purpose, and basic context based on information that is explicitly stated in extended texts), and listening subskill L4 (the ability to understand details in extended spoken texts) suggests the importance of providing students with the opportunity to hear and act upon extended spoken texts within the classroom.

Beyond correlation analysis, a hierarchical cluster analysis of the nine TOEIC subskills provides test administrators with a graphical representation of the test-takers' strengths and weaknesses (see Figure 1). In this study, test-takers had the greatest difficulty with items that required them to comprehend extended spoken texts and make inferences or connect information in written texts. This type of information can in turn help define cut-scores and allow for more nuanced placement decisions (Powers & Powers, 2014). Table 6 shows two possible ways in which the test-takers can be divided into their classes. The first way is to simply use the total TOEIC IP score or the listening and reading subsection test scores. The result would be a continuum of classes that range from test-takers who have a score in the 100s to test-takers with score in the 600s. This vertical approach, however, ignores that there are different ways to reach the same TOEIC IP score. Administrators and teachers as a result have limited information that could inform curriculum design and implementation.

The second way is take into account the listening and reading subskills that comprise the TOEIC test. Reading Table 6 horizontally reveals the group memberships that exist within the vertical continuum of TOEIC IP scores. The test-takers with the lowest scores (i.e., in the 100s) and the highest scores (i.e., in the 500s and 600s) predominately belonged to Cluster 1 and Cluster 4 respectively. In contrast, test-takers with scores ranging from the 200s to the 500s were split between two or three different clusters of students. Using this information, initial placement decisions can be made according to the total TOEIC IP scores and final placements can take into consideration the test-takers' level of success on the nine subskills. Administrators and teachers would then have the needed information to select materials to address specific listening and reading skills. There is a meaningful difference between (a) informing a teacher that a class has a mean score of 187.17 for the listening section and 121.83 for the reading section of the TOEIC IP test and (b) telling that teacher the class also belongs to Cluster 2, which generally has difficulty understanding details in extended spoken texts (L4), making inferences based on information in written texts (R1), and connecting information across multiple sentences in a single written text and across texts (R4). In short, the criterion-referenced information included in the nine TOEIC subskills can help test-takers and administrators gain a better understanding of what a TOEIC score actually means.

The analysis of the nine TOEIC subskills has the additional benefit of potentially identifying test-takers who have unique strengths and weaknesses. In this study, there were 15 students who had significantly different skill sets when compared to test-takers with similar test scores. Although this group of outlying students is only 1% of the test-takers, their needs should not be overlooked. Ideally placement decisions should be responsive to individual needs and not group norms.

There are, however, limits to the diagnostic information that the TOEIC IP test can provide. The nine TOEIC subskills are not as fine-grained as the subskills used in other standardized proficiency tests (see, e.g., Kim, 2014). The TOEFL test, for example, divides the TOEIC reading subskill R4 (understanding vocabulary in a written text) into deducing word meaning from context or without context (Jang, 2005). These more refined subskills can in turn help teachers and materials designers to develop lesson plans to address specific test-taker strengths and weaknesses. The TOEIC IP test also does not assess test-takers' level of productive English competence. As a result, score interpretation and placement decisions into four-skills language programs can be problematic when based on TOEIC IP scores (Mullen, 2009). A possible remedy might be administering the speaking and writing sections of the TOEIC test. However, the additional cost and resources required to test large groups of students may threaten the practicality of using the TOEIC IP test for placement purposes.

Although placement examinations are not necessarily high-stakes tests, they can have significant consequences on test-takers' chances of success in a program after they have been placed. As a result, there is a need for future investigations into the trustworthiness of inferences made from standardized proficiency test scores and placement decisions made based upon these scores. In addition, placement examinations that not only assess linguistic ability but also take into consideration test-takers' domains of language use (see Thompson, 2015) or their willingness to use their L2 require further investigation. Future research might also examine what type of information stakeholders such as teachers consider informative in the placement decision process as well as in the subsequent implementation of the language program (see Fox, 2009).

Despite a number of concerns surrounding the use of standardized proficiency examinations for placement purposes, the ever-increasing use of the TOEIC IP test at educational institutions in Japan gives rise to the need to carefully consider the potential formative value of TOEIC IP scores. This test provides test-takers and administrators with three levels of information: the total score, the listening and the reading section scores, and the percentage of correct responses for four listening and five reading TOEIC subskills. These scores are interrelated, but they provide unique vantage points that can be used to identify groups of test-takers based upon their strengths and weaknesses, which in turn have the potential of facilitating nuanced placements and more targeted language instruction.

Christopher Weaver is an associate professor in the Faculty of Business Administration at Toyo University, Tokyo, Japan. His research focuses upon assessment, task-based instruction, and willingness to communicate, with the purpose of facilitating opportunities for second language development.

References

Boldt, R., & Ross, S. (2005). *Language proficiency gain on the Test of English for International Communication: Meta-analysis of Japanese and Korean corporate language programs.* TOEIC Research Report. Retrieved from http://www.toeic.or.jp/library/toeic_data/toeic/data/pdf/boldt_ross2005.pdf

Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice-Hall Regents.

Burns, R., & Burns, R. (2008). *Business research methods and statistics using SPSS*. London, UK: Sage.

Chapman, M., & Newfields, T. (2008). The "new" TOEIC. *Shiken: JALT Testing & Evaluation SIG Newsletter, 12*(2), 32-37.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cronbach, L. (1971). Test validation. In R. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, DC: American Council on Education.

Dorans, N. (2004). Equating, concordance, and expectation. *Applied Psychological Measurement, 28,* 227-246. http://dx.doi.org/10.1177/0146621604265031

Educational Testing Service. (2000). *TOEIC can-do guide: Linking TOEIC scores to activities performed using English*. Retrieved from http://www.ets.org/Media/Research/pdf/TOEIC_CAN_DO.pdf

Educational Testing Service. (2007). *TOEIC listening score descriptors. TOEIC reading score descriptors.* Retrieved from https://www.ets.org/Media/Tests/TOEIC/pdf/TOEIC_LR_Score_Desc.pdf

Educational Testing Service. (2008). *TOEIC listening and reading: Official score report*. Retrieved from https://www.etsglobal.org/Pl/Pol/content/download/1461/25305/version/1/file/TOEIC-score-report.pdf

Educational Testing Service. (2012). *TOEIC listening & reading percentile rank*. Retrieved from https://www.ets.org/s/toeic/pdf/toeic_listening_and_reading_percentile_rank.pdf

Educational Testing Service. (2013). *TOEIC user guide: Listening and reading*. Retrieved from http://www.ets.org/Media/Tests/Test_of_English_for_International_Communication/TOEIC_User_Gd.pdf

Everitt, B. S., Landau, S., & Leese, M. (2001). *Cluster analysis* (4th ed.). London, UK: Arnold.

Falout, J. (2006). Japanese college laboratory of science majors preparing for the TOEIC. In K. Bradford-Watts (Ed.), *JALT2005 Conference Proceedings* (pp. 1140-1151). Tokyo: JALT.

Fox, J. (2009). Moderating top-down policy impact and supporting EAP curricular renewal: Exploring the potential of diagnostic assessment. *Journal of English for Academic Purposes, 8,* 26-42. http://dx.doi.org/10.1016/j.jeap.2008.12.004

Fulcher, G., & Davidson, F. (2009). Test architure, test retrofit. *Language Testing, 26,* 123-144. http://dx.doi.org/10.1177/0265532208097339

Green, A., & Weir, C. (2004). Can placement tests inform instructional decisions? *Language Testing, 21,* 467-494. http://dx.doi.org/10.1191/0265532204lt293oa

Hisatsune, A. (2007). Meet the challenges: Empowering TOEIC students. In K. Bradford-Watts (Ed.), *JALT2006 Conference Proceedings* (pp. 315-331). Tokyo: JALT.

In'nami, Y., & Koizumi, R. (2012). Factor structure of the revised TOEIC test: A multiple-sample analysis. *Language Testing, 29,* 131-152. http://dx.doi.org/10.1177/0265532211413444

Institute for International Business Communication. (2013). *TOEIC test data & analysis 2012*. Retrieved from http://www.toeic.or.jp/library/toeic_data/toeic_en/pdf/data/TOEIC_Program_DAA2012.pdf

Institute for International Business Communication. (2015a). *TOEIC Newsletter No. 125*. Retrieved from http://www.toeic.or.jp/library/toeic_data/toeic_en/pdf/newsletter/newsletterdigest125.pdf

Institute for International Business Communication. (2015b). *TOEIC test data & analysis 2014*. Retrieved from http://www.toeic.or.jp/library/toeic_data/toeic_en/pdf/data/TOEIC_Program_DAA.pdf

Ito, T., Kawaguchi, K., & Ohta, R. (2005). *A study of the relationship between TOEIC scores and functional job performance: Self-assessment of foreign language proficiency.* Retrieved from http://www.toeic.or.jp/library/toeic_data/toeic_en/pdf/newsletter/1_E.pdf

Jang, E. (2005). *A validity narrative: Effects of reading skills diagnosis on teaching and learning in the context of NG TOEFL* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (AAT 3182288)

Kim, A. (2014). Exploring ways to provide diagnostic feedback with an ESL placement test: Cognitive diagnostic assessment of L2 reading ability. *Language Testing*, *32,* 227-258. http://dx.doi.org/10.1177/0265532214558457

Kokhan, K. (2012). Investigating the possibility of using TOEFL scores for university ESL decision-making: Placement trends and effect of time lag. *Language Testing, 29,* 291-308. http://dx.doi.org/10.1177/0265532211429403

Kokhan, K. (2013). An argument against using standardized test scores for placement of international undergraduate students in English as a second language (ESL) courses. *Language Testing, 30*, 467-489. http://dx.doi.org/10.1177/0265532213475782

Liao, C. (2010). *TOEIC listening and reading test scale anchoring study*. Princeton, NJ: Educational Testing Service. Retrieved from http://www.ets.org/Media/Research/pdf/TC-10-05.pdf

Long, M., & Norris, J. (2000). Task-based language teaching and assessment. In M. Byram (Ed.), *Encyclopoedia of language teaching* (pp. 597-603). London, UK: Routledge.

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13,* 241-256. http://dx.doi.org/10.1177/026553229601300302

Mislevy, R., Almond, J., & Lukas, J. (2003). *A brief introduction to evidence-centered design.* Princeton, NJ: Educational Testing Service. Retrieved from https://www.ets.org/Media/Research/pdf/RR-03-16.pdf

Mizumoto, A., & Takeuchi, O. (2009). A closer look at the relationship between vocabulary learning strategies and the TOEIC scores*. TOEIC Research Report, 4,* 1-34.

Mullen, A. (2009). *The impact of using a proficiency test as a placement tool: The case of the Test of English for International Communication (TOEIC)* (Doctoral dissertation). Retrieved from www.theses.ulaval.ca/2009/26672/26672.pdf

Powers, D., & Powers, A. (2014). The incremental contribution of TOEIC Listening, Reading, Speaking, and Writing tests to predicting performance on real-life English language tasks. *Language Testing, 32*, 151-167. http://dx.doi.org/10.1177/0265532214551855

Schedl, M. (2010). *Background and goals of the TOEIC Listening and Reading test redesign project*. Princeton, NJ: Educational Testing Service. Retrieved from https://www.ets.org/Media/Research/pdf/TC-10-02.pdf

Sheskin, D. (2007). *Handbook of parametric and nonparametric statistical procedures* (4th ed.). Boca Raton, FL: Chapman & Hall/CRC.

Simes, R. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika, 73,* 751-754. http://dx.doi.org/10.1093/biomet/73.3.751

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

Szekely, G., & Rizzo, M. (2005). Hierarchical clustering via joint between-within distances: Extending Ward's minimum variance method. *Journal of Classification, 22,* 151-183. http://dx.doi.org/10.1007/s00357-005-0012-9

Tannenbaum, R., Rosenfeld, M., & Breyer, F. (1997). *Linking TOEIC score to self-assessments of English-language abilities: A study of score interpretation*. Unpublished manuscript.

Tannenbaum, R., & Wylie, E. (2008). *Linking English-language scores onto the Common European Framework of Reference: An application of standard-setting methodology.* Princeton, NJ: Educational Testing Service. Retrieved from http://www.ets.org/Media/Research/pdf/RR-08-34.pdf

Thompson, G. (2015). Understanding the heritage language student: Proficiency and placement. *Journal of Hispanic Higher Education, 14,* 82-96. http://dx.doi.org/10.1177/1538192714551277

Uchibori, A., Chujo, K., & Hasegawa, S. (2006). Towards better grammar instruction: Bridging the gap between high school textbooks and TOEIC. *The Asian EFL Journal Quarterly, 8,* 228-253. Retrieved from http://asian-efl-journal.com/June_2006_EBook_editions.pdf

Wall, D. , Clapham, C., & Alderson, J. C. (1994). Evaluating a placement test. *Language Testing, 11,* 321-344. http://dx.doi.org/10.1177/026553229401100305

Wilcox, R. (1987). *New statistical procedures for the social sciences: Modern solutions to basic problems*. Hillsdale, NJ: Erlbaum.

Wilson, K. (1989). *Enhancing the interpretation of a norm-referenced second-language test through criterion referencing: A research assessment of experience in the TOEIC testing context*. Princeton, NJ: Educational Testing Service. http://dx.doi.org/10.1002/j.2330-8516.1989.tb00153.x

Zhang, S. (2006). Investigating the relative effects of persons, items, sections, and languages on TOEIC score dependability. *Language Testing, 23,* 351-369. http://dx.doi.org/10.1191/0265532206lt332oa

Perspectives

The Role of Formative Assessment in Global Human Resource Development

Paul Wicking
Nagoya University

Over the last few years, the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) has been advocating the development of *guroubaru-jinzai* [global human resources]. These global human resources are to be top global performers who are intellectually savvy and internationally minded young men and women able to communicate across cultural divides and promote Japanese interests on the world stage. The success or failure of this policy will in no small part be determined by how well the various components of the educational system are aligned towards the goal. In particular, the ways in which assessment is conceptualized and practiced is vitally important. The continuation of a testing and examination system that consistently rewards memorization and repetition of learned facts will stifle the independence, creativity, and collaboration skills that MEXT hopes to develop. Rather, the development of these skills is best facilitated through formative assessment processes, which have until now been lacking within secondary and higher education in Japan. Instead there has been an inordinate focus on summative assessment and high stakes testing within the education system. In this paper I argue for the importance of formative assessment in shaping learners into the top global performers desired by MEXT and suggest some steps to be taken towards a contextually-based formative assessment practice.

ここ数年文部科学省はグローバル人材の育成を推奨している。グローバル人材とは異文化コミュニケーション能力と国際理解を備え、日本の国益を視野に世界で活躍できる知的エリートと言える。この政策が成功するには、教育制度を構成する様々な要素の良

好な連携が不可欠である。中でも教育評価の適切な概念化と実行は特に重要である。丸暗記や機械的な学習を奨励する試験制度は、自立、創造力および協調力など文部科学省が人材に求める能力の育成を妨害していると思われる。学生がこの能力を習得するためには、形成的評価が有効であるが、従来日本の高校や大学ではあまり行われてこなかった。実際には、総括的評価と一発試験ばかりが重要視されてきたと言ってよいだろう。本稿では、グローバル人材育成における形成的評価の重要性を論じ、教育環境に合わせた実践方法を提案する。

The importance of English education for Japan's national interest has been officially acknowledged for many years now and only seems to increase with time. The Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) in a recent policy document stated,

> グローバル化の進展の中で、国際共通語である英語力の向上は日本の将来にとって極めて重要である。 (Guroubaruka no shinten no naka de, kokusai kyoutsuugo de aru eigoryoku no koujou wa nihon no shourai ni totte kiwamete juuyou de aru.) [*With the progress of globalization, English has become a common language internationally and so improvement of English ability is vitally important for Japan's future*.] (MEXT, 2014)

As part of the recognition of the forces of globalization and the role of English within it, MEXT has advocated the development of *guroubaru-jinzai* [global human resources]. These are Japanese men and women equipped to succeed on the world stage by way of three general characteristics. Firstly, they have excellent language and communication skills. Secondly, they have a disposition that is proactive, ambitious, flexible, and responsible. Thirdly, they have a deep understanding of foreign cultures while maintaining a steadfast Japanese identity (MEXT, 2011).

The way in which classroom assessment is conceptualized and practiced will have a great impact in either helping or hindering the development of these *guroubaru-jinzai*. This is because of the pivotal role that assessment plays in the education system. "Every model of the teaching-learning process requires that teachers base their decisions—instructional, grading, and reporting—on some knowledge of student attainment of and progress towards desired learning outcomes" (Cheng, Rogers, & Hu, 2004, p. 361). It has been observed that much of assessment practice the world over is grounded in behaviourism, even though learning theory itself has moved on. According to James (2012),

> The dominance of psychometric models [of assessment] must, in large measure, be attributable to the fact that parents, em-

> ployers, policy makers, the media and the general public do not really understand what goes on in classrooms. Therefore they are wedded to proxy measures of learning and achievement that have doubtful validity. (p. 203)

Presently, little is known about how individual teachers working in the Japanese context conceive of and practice assessment and what effect this is having on student learning. It does seem, though, that in general teachers in Japan are less concerned with formative assessment practices, or assessment *for* learning, and instead prefer practices that promote student metacognition, or assessment *as* learning (Gonzales & Aliponga, 2012). This is despite the fact that there is a rich literature from Japan advocating essential components of formative assessment such as learning-oriented assessment tasks, development of evaluative expertise, and student engagement with feedback. In particular, portfolio creation as a learning-oriented assessment task has been used successfully to aid reflection (Howrey & Tanner, 2009), increase student engagement (Howrey, 2011), and boost motivation (Apple & Shimo, 2004). The development of evaluative expertise has received some attention through studies in self- and peer assessment. Matsuno (2009) found that although students did not assess themselves objectively, they were much more reliable in assessing their peers. Peer assessment has thus been recommended as one means of awarding grades, as peer scores correlate highly with instructor scores even when there has been no training in peer assessment (Saito, 2008). The value of peer assessment in the Japanese context has also been affirmed by a number of other studies (Asaba & Marlowe, 2011; Sato, 2013; Taferner, 2008; Wakabayashi, 2008).

The ways in which students engage with feedback is essential to the practice of formative assessment, as the whole endeavor revolves around using the results of assessment tasks to move students to the next stages of learning. Reugg (2015) found that students in a Japanese EFL class engaged more with teacher feedback than peer feedback, although peer feedback led to more successful revision attempts. Concerning methods of giving feedback, there seems to be little difference between form-focused and content-focused feedback in improving writing accuracy (Peloghitis, 2011) and little difference in learning effects between models and recasts in oral feedback (Sakai, 2004).

The aforementioned studies are a small sample of research that has looked into alternative methods of practicing assessment and using feedback from that assessment to further learning without relying on summative tests. However, summative testing still remains dominant in Japan, as it does in all Confucian-heritage cultures (Carless, 2011). The excessive influence of testing in Japanese schooling has led to its educational system being labeled as "examination oriented" by Sugimoto (2014) and as an "examocracy" by McVeigh (2006). This describes a situation in which most EFL lessons in formal schooling are not focused on promoting genuine communicative competence and other skills associated with the *guroubaru-jinzai*, but rather on preparing students for university entrance examinations (Kikuchi & Sakai, 2009).

This situation persists despite the fact that there is widespread agreement among scholars as to the benefits that derive from formative assessment as against summative assessment (Black & Wiliam, 1998). There are also a number of theoretical models of formative assessment that do exist, such as authentic assessment (O'Malley & Valdez Pierce, 1996), assessment for learning (Gardner, 2012), teacher-based assessment (Davison & Leung, 2009), and learning-oriented assessment (Carless, 2014). At present, however, these models of formative assessment are not in widespread use. Harlen and Winter (2004) argued that the main inhibiting factors are as follows:

- a reliance on assessment practices that give more attention to grading and assigning learners to "levels" rather than giving feedback about how work could be improved;
- teachers' lack of awareness of pupils' learning needs; and
- the high stakes attached to national test results, which encourage teachers to focus on the content of the tests and practicing test-taking. (p. 393)

In Japan, students are often ranked in respect to their peers, and the high-stakes nature of standardized tests exerts a powerful influence over all stages of education. Entrance exam pressure fuels EFL education in junior high school and high school, and tests such as TOEIC and EIKEN strongly influence English study in much post-secondary education.

The ways in which student learning is assessed are extremely influential on the teaching–learning process. In fact, within Japan itself, it has been observed that policies concerning assessment have had a transformative effect on EFL education throughout the last 150 years (Sasaki, 2008). The importance of assessment has been reflected in MEXT policy documents, as in the following example:

> 英語力の評価及び入学者選抜における英語力の測定については、４技能の総合的なコミュニケーション能力が適切に評価されるよう促す。(Eigoryoku no hyouka oyobi nyuugakusha senbatsu ni okeru eigoryoku no sokutei ni tsuite wa, 4 ginou no sougoutekina komyunikeishon nouryoku ga tekisetsu ni hyouka sareru you unagasu.) [*Regarding the measurement of English ability for the purposes of English assessment and admission to school, we assert that all four skills for comprehensive communication ability be assessed appropriately*.] (MEXT, 2014, p. 10)

But just what does it mean to assess English "appropriately"? Certainly, when students perceive an assessment method to be inappropriate, it encourages a surface approach to learning in which there is little deep thought and little cognitive gain (Struyven, Dochy, & Janssens, 2005). However, the qualities that make an assessment event appropriate in the eyes of Japanese students and their teachers need to be clarified. The present paper is aimed at contributing to a professional discourse concerning appropriate assessment as it is conceptualized and practiced in Japanese EFL. First, assessment will be defined and the relationship between summative and formative assessment will be explored. Second, the sociocultural factors that work to hinder formative assessment practice will be described. Finally, some suggestions will be made as a way forward to a contextually based formative assessment practice.

The Summative and Formative Functions of Assessment

Brown, Hudson, Norris, and Bonk (2002) defined assessment as "the process of gathering information in consistent ways via tests (and related sources), making interpretations based on that information, and then applying those interpretations to reach decisions or initiate particular actions within the language classroom or program" (p. 13). The implication of this definition is that assessment processes should ideally feed back into classroom instruction, so it is better for teachers to have a certain amount of freedom and authority in planning and implementing procedures for assessment. In this way assessment is best conceptualized as an approach (which allows for greater teacher autonomy in implementation), rather than a method (prescriptive techniques used irrespective of context). Abrams (2014) wrote,

> Best practices research distinguishes between methods and approaches and asserts that educators must have the flexibility

> to adapt their pedagogical approaches . . . as necessary in order to best meet their learners' needs, just as those approaches must reflect the philosophical beliefs and values of the educator. (p. 1)

When making policy for language assessment in the Japanese context, encouraging an approach will yield better results than imposing a method.

The definition of assessment can be further broken down into two broad categories: summative and formative. Although there is no agreed-upon definition of these terms, it may be broadly stated that summative assessment is concerned with summarizing the amount of learning that has been achieved up to a certain point. It seeks to rank or certify students in a process that could be called "assessment *of* learning." Formative assessment, on the other hand, could be described as "assessment *for* learning," in that it is primarily concerned with promoting effective learning by using the results to modify the teaching and learning activities that are taking place. For the purpose of this paper, formative assessment is conceptualized in line with Wiliam (2004), who stated, "formative assessment refers not to an assessment, nor even to the purpose of an assessment, but the function that it actually serves" (p. 3). The fact is that there is no inherent quality of a test that makes it either formative or summative. Rather, these two labels refer to the ways in which test results are used (i.e., to grade students on their level of achievement or to direct the focus of further study).

Although both functions of assessment have their place, an over-emphasis on the summative function of tests will be to the detriment of learner development. Any Japanese EFL program that focuses heavily on summative tests risks sabotaging its own efforts at improving students' educational attainment. Boud (2000) wrote,

> Ironically, summative assessment drives out learning at the same time it seeks to measure it. It does this by taking responsibility for judgments about learning away from the only person who can learn (the student) and placing it unilaterally in the hands of others. It gives the message that assessment is not an act of the learner, but an act performed on the learner. (p. 156)

Boud (2000) argued that although summative assessment is useful and desirable, there needs to be a significant shift of balance towards the formative function to empower students to be competent assessors of their own performance and consequently effective lifelong learners.

The heart of formative assessment, as stated by Black and Wiliam (2009), is grounded in five key strategies:

1. clarifying and sharing learning intentions and criteria for success;
2. engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding;
3. providing feedback that moves learners forward;
4. activating students as instructional resources for one another; and
5. activating students as the owners of their own learning. (p. 8)

Somewhat propitiously, research in the Japanese EFL context has suggested that each of these five integrated strategies can be used to good effect. The first strategy has been increasingly implemented in English learning via “can-do” statements. In particular, the Common European Framework of Reference (CEFR) descriptors (O’Dwyer, 2011) and its Japanese variation, the CEFR-J (Tono, 2013) have been used as a broad roadmap to direct learning paths and make the often hidden goals for success more visible. The second strategy, that of eliciting evidence of understanding through learning tasks, has been most widely recognized through the method of communicative language teaching and its most popular offshoot, task-based language teaching (Sybing, 2011). The meaning-focused nature of tasks arguably provides better evidence of students’ deeper understanding of language as it is actually used. The third strategy, that of feedback, has been the focus of much research activity, mainly centered on oral corrective feedback (Mori, 2002; Sakai, 2004) and written feedback (Peloghitis, 2011). The activation of students as instructional resources for one another has been done successfully in Japan through collaborative learning (Kato, Bolstad, & Watari, 2015), cooperative group work (Hirose & Kobayashi, 1991), and peer assessment (Asaba & Marlowe, 2011; Okuda & Otsu, 2010). The final strategy has likewise been advocated in Japanese EFL, mainly in the guise of promoting learner autonomy (Irie & Stewart, 2012) and self-regulated learning (Sullivan, 2014).

It is important to note that tests that have heretofore been used for summative purposes can be subsumed into a formative assessment framework. In other words, according to this conceptualization, summative tests can be used for formative purposes. There is no need for a radical departure from all high-stakes testing as it is currently conducted in Japan, but rather for a strategy that appropriates summative tests and uses them to promote productive student learning processes. This has been called "the formative use of summative tests," or FUST (Carless, 2011). In countries like Japan that have an entrenched system of high-stakes EFL testing, FUST provides a workable compromise between institutional interests that require certification and ranking and pedagogical interests that seek to promote greater learning achievement. Brown (2000) suggested one way in which this might be achieved: by using university entrance examinations in Japan to promote learning.

The positive aspect of summative testing has also been addressed by Taras (2005), who argued that for any assessment to be used for formative purposes, it must be grounded upon a summative judgment that has preceded it. If appropriate feedback is to be given to the learner to help aid further learning, that feedback must be based upon a summative assessment of that student's achievement. And so, "by recognizing that summative assessment is central and necessary to all assessment, it should stop the demonization of assessment for validation and certification, and instead see it as a stepping stone to learning" (Taras, 2005, p. 476).

Although the practice of assessment needs to be contextually grounded, debates about formative and summative assessment must not be limited to within the boundaries of the educational institution in which it takes place. Rather, as sociocultural theory attests,

> Activities within schools and practices associated with them are part of the broader cultural systems of relations, and social structure in which they have meaning. Summative, and increasingly formative, assessment practices are shaped by broader social and educational policies and structures. (Elwood & Murphy, 2015, p. 184)

As learning institutions seek to comply with MEXT objectives in developing *guroubaru-jinzai*, a well-articulated strategy of formative assessment is indispensible. However, to implement such a strategy a number of hindrances relating to student beliefs and teacher roles within the classroom will need to be negotiated.

Hindrances to Formative Assessment Practice: Student Beliefs and Teacher Role Conflict

Major potential obstacles to the development of formative assessment practice are the divergent beliefs that are sometimes held within the classroom. It seems that students can hold quite different beliefs about assessment from those held by their teachers. Brown (2009), in an exploratory study into teacher and student beliefs about effective language teaching, found that there were significant discrepancies between the beliefs of the two groups. Generally, the teachers believed more strongly in the principles of communicative language teaching (CLT) than did their students. Brown noted, "Unfortunately, in many cases, the participating teachers' perceptions of this ideal communicative classroom did not parallel their students' perceptions" (Brown, 2009, p. 54). As formative assessment finds natural expression through the practice of CLT, divergent teacher–student beliefs will need to be negotiated. For example, it has been noted in Japan that many learners appreciate form-focused instruction and explicit correction (Sato, Fukumoto, Ishitobu, & Morioka, 2012), which some teachers may be unwilling to give. Identifying and clarifying differences in assessment belief would help nurture a classroom ecology more conducive to language learning.

Of course, this ecology is best sustained when the learner is motivated to learn and the teacher is motivated to teach. However, this does not happen as often as we would like. Perrenoud (cited in Black and Wilian, 1998) wrote,

> A number of pupils do not aspire to learn as much as possible, but are content to "get by," to get through the period, the day or the year without any major disaster, having made time for other activities other than school work . . . Every teacher who wants to practice formative assessment must reconstruct the teaching contracts so as to counteract the habits acquired by his pupils. (p. 21)

This tendency towards apathy is arguably the case for many EFL students in higher education in Japan (McVeigh, 2001), where English is a compulsory subject required across most disciplines and a proverbial bitter pill that must be swallowed.

It is vital that student beliefs about assessment be addressed, as this is potentially one of the greatest hindrances to formative assessment in Japanese EFL. Even though teachers are invested with authority and responsibility to ensure that conditions are ripe for learning, it is really the learner who has ultimate sway over the learning process. Concerning assessment for learning (AFL), Davison and Leung (2009) noted that "the learner's role is crucial because it is the learner who does the learning. This point seems obvious, even trite, but it is central to the AFL philosophy and, if treated seriously, clearly highlights where formative assessment can go wrong" (p. 399). The old adage, *You can lead a horse to water, but you can't make it drink,* is especially applicable to those classrooms where EFL is a required subject and student motivation is low. The teacher may spend a great deal of time and effort in creating assessment tasks that are interesting and designed to promote effective learning habits, but if the student does not want to learn, then no learning will take place. Therefore one of the teacher's primary tasks is to foster within students those beliefs that are conducive to active participation in learning and assessment processes.

This is no easy feat, and it is made all the more difficult because of the need to balance teacher roles that are often in conflict during regular classroom interaction. Two main roles of EFL teachers have been identified as that of rater (or assessor), which requires teachers to identify levels of language achievement, and that of facilitator, in which the teacher continually appraises students through the ebb and flow of classroom discourse (Rea-Dickins, 2006). The facilitator role has a focus on academic improvement, and the assessor role is concerned with external accountability (Ewell, 1991). These roles sometimes conflict, as teachers are often "at the confluence of different assessment cultures and faced with significant dilemmas in their assessment practices: sometimes torn between their role as facilitator and monitor of language development and that of assessor and judge of language performance as achievement" (Rea-Dickins, 2004, p. 253). Their role as language-learning facilitator requires teachers to be responsive to individual needs and sensitive to personalities; their role as assessor requires them to be impartial and somewhat removed from personal circumstances. The result of such internal conflict is that there is often a substantial gap between teachers' beliefs and their actual practice (Saad, Sardarah, & Ambarwati, 2013). A dialogic negotiation and renegotiation of teacher and student roles within the EFL classroom will pave the way for an assessment practice that effectively assesses language achievement and also meets the needs of individual learners.

Towards a Contextually-Based Formative Assessment Practice

Encouraging teachers to engage more actively in formative assessment will not in itself be enough to make a significant contribution to the development of *guroubaru-jinzai*. The way in which Japanese students conceptualize assessment and their beliefs about assessment must also be shaped and molded in ways that are conducive to productive learning. Individual students will interpret and make meaning of assessment practices according to their own belief structures. Learners may lack the cognitive insight to understand the intentions of assessment even when these are explicitly stated (Rea-Dickins, 2006); or they may find it expedient to consciously work against the teacher's stated goals in order to pursue their own ends (Spence-Brown, 2001). In this way, students construct a personal version of the hidden curriculum based upon their previous experiences with assessment and the assumptions to which they give rise (Sambell & McDowell, 1998). Students who have been brought up in a heavily testing-oriented culture will have developed certain predispositions towards assessment that will influence the way they approach any kind of activity in which they will be evaluated. The aims of formative assessment need to be clearly explained to students unfamiliar with it if they are to become active participants in the teaching–learning process.

Traditional methods of assessment used throughout the past decades have left a deep imprint on educational culture. Any attempt to impose a system of assessment practices or techniques on top of existing beliefs and values, as one may apply a patch to an old garment, is bound to be an exercise in futility. Rather, the effort to move existing assessment practice towards a more learning-oriented approach will necessarily entail a shift in core understanding about teacher and learner roles. This is because, as Pedder and James (2012) noted,

> [The alteration of assessment practice] involves transformations in classroom processes, and this entails change not only in what teachers and students do but also in how they relate. It involves behavior imbued with deeper understanding and values. These understandings and values are informed by norms associated with particular conceptions of appropriate roles for teachers and for students. (p. 37)

Traditionally, and perhaps even now in Japan, teachers have been seen as the provider of knowledge, with students expected to soak up that knowledge in order to mechanically recite it on the final exam. However, approaches to assessment that are focused on promoting learning "accompany a necessary shift in learning relationships, with students becoming part of effective learning communities, involved in co-construction of knowledge, rather than passive recipients of teachers' knowledge" (Sambell, McDowell, & Montgomery, 2013, p. 8).

Thus, in an ideal situation, how would teachers and students view their respective roles? Black and Wiliam (2009) argued,

> The teacher is responsible for designing and implementing an effective learning environment, and the learner is responsible for the learning within that environment. Furthermore, since the responsibility for learning rests with both the teacher and the learner, it is incumbent on each to do all they can to mitigate the impact of any failures of the other. (p. 7)

According to this view, when learning fails to occur, blame cannot be laid squarely at the feet of either the teacher or the learner. Both have to work together to achieve the educational goals that have been set.

Once the foundation has been laid, strategies for encouraging productive assessment practice put in place, and teachers and students understand their roles in the teaching-learning process, a framework for formative assessment will be a practical help in EFL curriculum design. The following six questions can support educators in their efforts to implement assessment procedures that will encourage productive learning. These questions are based on Sambell, McDowell, and Montgomery (2013), who consolidated much of the research into assessment for learning and university education.

1. Does the assessment develop students' abilities to evaluate their own progress and direct their own learning?
2. Is the assessment rich in informal feedback (such as peer evaluation and group work)?
3. Is the assessment rich in formal feedback (such as teacher comments and corrections)?
4. Does the assessment offer extensive confidence-building opportunities and practice?
5. Is there an appropriate balance of summative and formative assessment?
6. Is there an emphasis on authentic and complex assessment tasks?

The extent to which each of these questions can be answered positively is the extent to which the assessment can be considered efficacious in promoting learning.

Conclusion

The Japanese government is to be lauded for its efforts to promote globally minded and internationally focused young men and women who can communicate across cultures. The development of English language skills and the assessment of those skills play a major part in this. If assessment of English is to be done "appropriately" (MEXT, 2014), then greater emphasis must be placed on formative procedures. At present, the summative function is the primary focus of much assessment practice in Japanese EFL. High-stakes tests that exert enormous influence over the future prospects of young people overshadow formative assessment procedures that aim to promote genuine learning. This imbalance needs to be redressed. The principles and strategies of formative assessment aim to create just the kind of person MEXT desires. These are people who are able to critically analyze their own work and that of peers, who are confident and competent in practical English use, and who are autonomous and self-regulating as they continue along a trajectory of life-long learning. It is hoped that all levels of the educational system will recognize the value of encouraging an approach to assessment that prioritizes its formative functions and consequently work to implement formative assessment practices in their places of learning.

Paul Wicking is currently doing doctoral research at Nagoya University Graduate School of International Development and is an associate professor at Meijo University. His research interests include learning-oriented assessment, teacher beliefs and practice, and task-based language teaching.

References

Abrams, D. J. (2014, December 3). What we talk about when we talk about best practices: Methods and approaches. *Multibriefs*. Retrieved from http://exclusive.multibriefs.com/content/what-we-talk-about-when-we-talk-about-best-practices-methods-and-approaches

Apple, M., & Shimo, E. (2004). Learners to teacher: Portfolios, please! Perceptions of portfolio assessment in EFL classrooms. In T. Newfields (Ed.), *Proceedings of the 2004 JALT Pan-SIG Conference* (pp. 53-59). Tokyo: JALT.

Asaba, M., & Marlowe, J. (2011). Using peer assessment in the language classroom. *The Language Teacher, 35*(1), 29-33.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5,* 7-74. http://dx.doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability,* 21, 5-31. http://dx.doi.org/10.1007/s11092-008-9068-5

Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education, 22,* 151-167. http://dx.doi.org/10.1080/713695728

Brown, A. (2009). Students' and teachers' perceptions of effective foreign language teaching: A comparison of ideals. *The Modern Language Journal, 93,* 46-60. http://dx.doi.org/10.1111/j.1540-4781.2009.00827.x

Brown, J. (2000). University entrance examinations: Strategies for creating positive washback on English language teaching in Japan. *Shiken: JALT Testing & Evaluation SIG Newsletter, 3*(2), 4-8.

Brown, J., Hudson, T., Norris, J., & Bonk, W. (2002). *An investigation of second language task-based performance assessments*. Technical Report #24. Honolulu: Second Language Teaching and Curriculum Center, University of Hawai'i.

Carless, D. (2011). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings*. New York, NY: Routledge.

Carless, D. (2014). Exploring learning-oriented assessment processes. *Higher Education, 69,* 963-976. http://dx.doi.org/10.1007/s10734-014-9816-z

Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods and procedures. *Language Testing, 21,* 360-389. http://dx.doi.org/10.1191/0265532204lt288oa

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly, 43,* 393-415. http://dx.doi.org/10.1002/j.1545-7249.2009.tb00242.x

Elwood, J., & Murphy, P. (2015). Assessment systems as cultural scripts: A sociocultural theoretical lens on assessment practice and products. *Assessment in Education: Principles, Policy and Practice, 22*(2), 182-192. http://dx.doi.org/10.1080/0969594X.2015.1021568

Ewell, P. (1991). To capture the ineffable: New forms of assessment in higher education. *Review of Research in Education, 17*, 75-125. http://dx.doi.org/10.3102/0091732X017001075

Gardner, J. (2012). Introduction. In J. Gardner (Ed.), *Assessment and learning* (2nd ed., pp. 1-8). London, UK: Sage.

Gonzales, R., & Aliponga, J. (2012). Classroom assessment preferences of Japanese language teachers in the Philippines and English language teachers in Japan. *MEXTESOL Journal, 36*(1), 1-19.

Harlen, W., & Winter, J. (2004). The development of assessment for learning: Learning from the case of science and mathematics. *Language Testing, 21,* 390-408. http://dx.doi.org/10.1191/0265532204lt289oa

Hirose, K., & Kobayashi, I. (1991). Cooperative small group discussion. *JALT Journal, 13,* 57-72.

Howrey, J. (2011). Promoting active learning through writing portfolios. *OnCue Journal, 5*(2), 3-18.

Howrey, J., & Tanner, P. (2009). Writing portfolios: Empowering students, teachers, and the curriculum. In A. M. Stoke (Ed.), *PAC7 at JALT2008 Conference Proceedings* (pp. 1157-1167). Tokyo: JALT.

Irie, K., & Stewart, A. (Eds.). (2012). *Realizing autonomy*. Basingstoke, UK: Palgrave Macmillan.

James, M. (2012). Assessment in harmony with our understanding of learning. In J. Gardner (Ed.), *Assessment and learning* (2nd ed., pp. 33-48). London, UK: Sage.

Kato, Y., Bolstad, F., & Watari, H. (2015). Cooperative and collaborative learning in the language classroom. *The Language Teacher, 39*(2), 22-26.

Kikuchi, K., & Sakai, H. (2009). Japanese learners' demotivation to study English: A survey study. *JALT Journal, 31,* 183-204.

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

McVeigh, B. (2001). Higher education, apathy and post-meritocracy. *The Language Teacher, 25*(10), 29-32.

McVeigh, B. (2006). *The state bearing gifts: Deception and disaffection in Japanese higher education*. Lanham, MD: Lexington Books.

MEXT. (2011). *Guroubaru jinzai no ikusei ni tsuite* [Concerning the development of global human resources]. Retrieved from http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo3/047/siryo/__icsFiles/afieldfile/2012/02/14/1316067_01.pdf

MEXT. (2014). *Kongo no eigo kyouiku no kaizen, juujitsu housaku ni tsuite houkoku: Guroubaruka ni taiou shita eigo kyouiku kaikaku no itsutsu no teigen* [About the future English educational improvement and report on fulfillment measures: Five proposals on English education reform corresponding to globalization]. Retrieved from http://www.mext.go.jp/b_menu/shingi/chousa/shotou/102/houkoku/attach/1352464.htm

Mori, R. (2002). Teachers' beliefs and corrective feedback. *JALT Journal, 24,* 48-69.

O'Dwyer, F. (2011). Using a European language portfolio for self-assessment, goal-setting and reflection. *OnCUE Journal, 5*(2), 49-61.

Okuda, R., & Otsu, R. (2010). Peer assessment for speeches as an aid to teacher grading. *The Language Teacher, 34*(4), 41-47.

O'Malley, J. M., & Valdez Pierce, L. (1996). *Authentic assessment for English language learners*. Boston, MA: Addison-Wesley.

Pedder, D., & James, M. (2012). Professional learning as a condition for assessment for learning. In J. Gardner (Ed.), *Assessment and learning* (2nd ed., pp. 33-48) London, UK: Sage.

Peloghitis, J. (2011). Form-focused feedback in writing: A study on quality and performance in accuracy. In A. Stewart (Ed.), *JALT2010 Conference Proceedings* (pp. 260-268). Tokyo: JALT.

Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing, 21,* 249-258. http://dx.doi.org/10.1191/0265532204lt283ed

Rea-Dickins, P. (2006). Currents and eddies in the discourse of assessment: A learning-focused interpretation. *International Journal of Applied Linguistics, 16,* 163-188. http://dx.doi.org/10.1111/j.1473-4192.2006.00112.x

Ruegg, R. (2015). Differences in the uptake of peer and teacher feedback. *RELC Journal, 46*(2), 131-145. http://dx.doi.org/10.1177/0033688214562799

Saad, M. R., Sardarah, S. A., & Ambarwati, E. K. (2013). Iranian secondary school EFL teachers' assessment beliefs and roles. *Life Science Journal, 10,* 1638-1647. Retrieved from http://www.lifesciencesite.com/lsj/life1003/246_B01918life1003_1638_1647.pdf

Saito, H. (2008). EFL classroom peer assessment: Training effects on rating and commenting. *Language Testing, 25,* 553-581. http://dx.doi.org/10.1177/0265532208094276

Sakai, H. (2004). Roles of output and feedback for L2 learners' noticing. *JALT Journal, 26,* 25-54.

Sambell, K., & McDowell, L. (1998). The construction of the hidden curriculum: Messages and meanings in the assessment of student learning. *Assessment & Evaluation in Higher Education, 23*, 391-402. http://dx.doi.org/10.1080/0260293980230406

Sambell, K., McDowell, L., & Montgomery, C. (2013). *Assessment for learning in higher education*. New York, NY: Routledge.

Sasaki, M. (2008). The 150-year history of English language assessment in Japanese education. *Language Testing, 25,* 63-83. http://dx.doi.org/10.1177/0265532207083745

Sato, M. (2013). Beliefs about peer interaction and peer corrective feedback: Efficacy of classroom intervention. *The Modern Language Journal, 97,* 611-633. http://dx.doi.org/10.1111/j.1540-4781.2013.12035.x

Sato, K., Fukumoto, Y., Ishitobi, N., & Morikoa, T. (2012). Focus-on-form instruction and student learning in Japanese junior high schools. In A. Stewart & N. Sonda (Eds.), *JALT2011 Conference Proceedings*, 283-303. Tokyo: JALT.

Spence-Brown, R. (2001). The eye of the beholder: Authenticity in an embedded assessment task. *Language Testing, 18,* 463-481. http://dx.doi.org/10.1177/026553220101800408

Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment and Evaluation in Higher Education, 30,* 325-341. http://dx.doi.org/10.1080/02602930500099102

Sugimoto, Y. (2014). *An introduction to Japanese society* (4th ed.). Port Melbourne, Australia: Cambridge University Press.

Sullivan, K. (2014). Reconsidering the assessment of self-regulated foreign language courses. *Studies in Self-Access Learning Journal, 5,* 443-459. Retrieved from http://sisaljournal.org/archives/dec14/sullivan/

Sybing, R. (2011). A response to criticism of TBLT in Japan's language classrooms. *JALT Journal, 33*, 67-76.

Taferner, R. (2008). Toward effective EFL writing revision: Peer review. *OnCUE Journal, 2*(2), 76-91.

Taras, M. (2005). Assessment—summative and formative—some theoretical reflections. *British Journal of Educational Studies, 53,* 466-478. http://dx.doi.org/10.1111/j.1467-8527.2005.00307.x

Tono, Y. (Ed.). (2013). *Eigo totatsudo shihyo CEFR-J gaido bukku: CAN-DO lisuto sakusei katsuyo* [The CEFR-J handbook: A resource book for using can-do descriptors for English language teaching]. Tokyo: Taishukanshoten.

Wakabayashi, R. (2008). The effect of peer feedback on EFL writing: Focusing on Japanese university students. *OnCUE Journal, 2*(2), 92-110.

Wiliam, D. (2004, April). *Assessment and the regulation of learning.* Paper presented the annual meeting of the National Council on Measurement in Education, San Diego, CA. Retrieved from http://www.dylanwiliam.org/Dylan_Wiliams_website/Papers_files/NCME%2004%20paper.pdf

International Perspectives on Motivation. Ema Ushioda (Ed.). London, UK: Palgrave MacMillan, 2013. xii + 243 pp.

Reviewed by
Matthew T. Apple
Ritsumeikan University

Language learning motivation has been, is, and will continue to be of concern to language teachers and researchers across the globe. Yet much language learning motivation research and theory may seem at times abtruse and inaccessible to language practitioners. The volume *International Perspectives on Motivation*, edited by Ema Ushioda, has two stated aims. First, the book underlines the increasing global importance of understanding language learning motivation in countries where languages are taught by L1 users as well as by advanced L2 users to novice L2 users in nonnative contexts. Second, although in the older, more traditional view L2 motivation is a static concept that can be generalized across cultures, the book highlights what could be termed a paradigm shift toward an understanding of L2 motivation as related to a sense of self and identity within a local, yet globally connected context. What this means in practical terms is a book whose content is informed by the experiences of individual language teachers in diverse countries across five continents that has applications for teachers in any educational context in which motivation is a concern.

Most chapters in the volume each focus on a single EFL educational context, ranging from a junior high school in Indonesia (Chapter 1) to a medical school in Bahrain (Chapter 6), with ample quotations from student and teacher interviews and detailed descriptions of the educational and political context. Two chapters (7 and 11) are based in the ESL contexts of Australia and North America, respectively, and two chapters (8 and 9) deal with the issue of "authentic" English outside the classroom in digital, online contexts. Finally, the penultimate chapter (12) examines reactions and motivations of native speaker British teachers with overseas teaching experiences. Two interesting features of the book are the discussion questions at the conclusion of each chapter (presumably for teacher-training courses) as well as brief one- or two-sentence descriptions of two or three suggested readings.

One strength of the volume is the reader-friendly presentation of the research. Most of the chapters are based on action research conducted by classroom teachers, primarily via interview, allowing the students' voices to clearly ring out. Although many other language motivation books (e.g., Apple, Da Silva, & Fellner, 2013; Dörnyei & Ushioda, 2009) contain complicated statistics such as structural equation modeling that may seem to have little connection to individual practitioners' classroom situations, the descriptions and quotations from students and teachers in this volume are easily understandable and applicable for those concerning themselves primarily with what happens at the chalkface. Despite claims that qualitative findings are "often not transferrable to other contexts" (p. 41), there are numerous findings common among the various chapters: a sense among students that they are only studying grammar and vocabulary for the purpose of passing high stakes exams, that the textbooks used in class are uninteresting and irrelevant to their language learning needs, and that they are not receiving enough opportunity in class for communicative self-expression. The importance of social support from family and friends, positive teacher-student rapport, and an overall supportive, nurturing classroom atmosphere are also emphasized throughout the studies in the book.

The role of the native speaker teacher is also examined (Chapter 2 and Chapter 12), relating intercultural awareness and teacher motivation to the pervasive "native speaker fallacy" (Phillipson, 1992) and "native speakerism" (see Houghton & Rivers, 2013). The role of the native speaker teacher is particularly relevant in Japan, where the native speaker of English in particular is exoticized as the "other" and Japanese learners come to have no sense of authority ownership over the language (Matsuda, 2003). This idealization of the native speaker undermines language learners' sense of themselves as a user of the language, affecting their motivation to learn and use the language. Thus, these two chapters serve to highlight the connections among teacher motivation, language ownership, and learner motivation.

Although Chapters 8 and 9 contain no study data and are essentially summaries of previous research and theory, I found them particularly compelling in their examination of authentic language tasks outside the classroom, as well as the difficulties concerning the potential "invasion" of students' online space by well-meaning language teachers seeking to connect classroom tasks to the real world. Although several studies in the volume unequivocally locate themselves within the ESL-EFL dichotomy decried by the editor in the introductory chapter, these two computer-assisted language learning (CALL)-related chapters exemplify the editor's opinion that the clear-cut distinction between ESL and EFL has begun to break down. These chapters discuss the ubiquitous nature of the Internet, which now is evident to some degree in daily life everywhere, in every cultural context, and describe the Internet and social networking services (SNSes) as borderless online communities that both shape and are shaped by the forces of globalization. Thus, even if there are no speakers of the foreign language physically present in daily life, language learners can easily interact in their L2 online. The rapid adoption of smartphone technology has only increased the degree of access to online language-use opportunities.

The volume does have a number of flaws frequently found in action or classroom-based research. Although presumably based on some form of grounded theory, most of the chapters do not describe any research methodology prior to data collection. For example, several chapters present the results of interviews without first specifying the qualitative interpretative framework (e.g., neo-Marxist, structuralist, deconstructionist) being used to analyze the data or even the criteria for selecting study participants. As a result, the writing often gives the impression that those interviewed were simply chosen on a whim or as a convenience sample, with no particular research questions or hypothesis in mind. In addition, several chapters seem to gloss over existing motivational studies or present brief summaries of motivational theories without the use of citations or data to support claims. Finally, a handful of chapters use Likert-scale questionnaires but misuse the statistics derived from them. For example, the author of a study in one chapter simply adds item means, claims that students have "high scores" on certain motivational orientations, then conducts multiple *t* tests without using Bonferroni adjustments to control for Type II error. Although this use of questionnaires may seem intuitive and is quite prevalent in investigations of motivation in the SLA field, it comes as somewhat of a surprise to see quantitative data used in such fashion in an otherwise well-written volume of qualitatively oriented research studies.

Overall, although the studies in this volume could have been conducted with a bit more academic rigor, the book seems an ideal introduction to the field of second language motivation research. The irony is that the teachers who will benefit the most from this book may be the most reluctant to read it in the first place. As the findings of the final study in the book (Chapter 12) illustrate, native speaker teachers typically disregard the opinions and writings of second language researchers and rely almost entirely on their own intuition and previous teaching experiences. Such teachers see little or no value in improving either their teaching qualifications or their understanding of the field in general due to a perception that academic research and academic papers have nothing to do with what happens in the language classroom. However, as the editor herself points out in the final chapter of the volume, it is teachers who ultimately have the greatest impact on learner motivation in the classroom, making a bridging of the gap between motivation researchers and teachers all the more important. Despite their limitations and flaws, the qualitative studies in multiple educational contexts in the volume represent a good first step for teachers to find out how language learning motivation theory and practice can inform each other.

References

Apple, M., Da Silva, D., & Fellner, T. (2013). *Language learning motivation in Japan*. Bristol, UK: Multilingual Matters.

Dörnyei, Z., & Ushioda, E. (2009). *Motivation, language identity, and the L2 self*. Bristol, UK: Multilingual Matters.

Houghton, S. A., & Rivers, D. (2013). *Native-speakerism in Japan: Intergroup dynamics in foreign language education.* Bristol, UK: Multilingual Matters.

Matsuda, A. (2003). The ownership of English in Japanese secondary schools. *World Englishes, 22,* 483-496. http://dx.doi.org/10.1111/j.1467-971X.2003.00314.x

Phillipson, R. (1992). *Linguistic imperialism.* Oxford, UK: Oxford University Press.

***Critical Thinking and Language: The Challenge of Generic Skills and Disciplinary Discourse*. Tim John Moore. London, UK/New York, NY: Bloomsbury Academic, 2013. x + pp. 247.**

Reviewed by
James Crocker
Kobe Women's University

For anyone who believes that we English teachers should contribute to the development of critical thinking (CT) skills in our students, this book will be an interesting read. There must be quite a significant number of us, because these days it seems that many new ESL and EFL textbooks make the "Now with added critical thinking!" promise in bursting stars on their covers.

Publishers are obviously responding to a growing sense, not only among English teachers but in many disciplines, that CT skills are not only valuable in general but can enhance learning.

But what exactly do we mean when we refer to CT? And are those who teach it all teaching the same skills? These are the questions that Tim John Moore tackles in this well-written and thoughtful study. He interviews a group of academics from three different disciplines at an Australian university searching for "networks of 'coherence' and 'logic' around the various usages of the term 'critical thinking'" (p. 38).

Specifically there are three questions he addresses (see p. 5):

1. What does the term mean to individuals teaching in different disciplines in an Australian university?
2. In what ways, and to what extent, are there variations in the meaning of the term in different disciplines?
3. What are the implications of the answers to the first two questions for teaching critical thinking in universities?

The first two chapters discuss the history of CT, the problems associated with trying to define it, and how general a skill we can consider CT to be. Moore is acutely aware that a study with this approach walks straight into a semantic minefield. It is not just the term CT that has a multitude of semantic nuances—its similes and terms for its subskills are often equally hard to nail down. What is rational thinking, for example? How does it differ from CT? Moore compares his task with Wittgenstein's (1958) example of trying to define the word *game*. This term could include board games, card games, Olympic Games and so on, which all have a complicated network of similarities.

Problems arise when defining terms in isolation. Therefore, Moore attempts to give the term CT some context by looking at three practical aspects of its usage. In Chapters 3 and 4 he introduces the academics he interviewed for the study and reports on their perceptions of the term, how they apply it, and what skills they try to develop in their students. Then in Chapter 5 he examines how the term is used in material introducing CT that is distributed to students by the different disciplines and in the tasks students are given as assignments.

The disciplines included are history, philosophy, and literary or cultural studies, purely because they shared the same building that Moore worked in. It's obviously a limited selection and unfortunately for us, does not include language teaching. However, it does give an indicative example of how conceptions can both vary and overlap, and as such, the book will be as relevant and thought-provoking to language teachers as it would be to those in any other discipline.

Unsurprisingly, Moore finds a great variety in the ways that the 17 academics describe the term CT. They differ not only between disciplines but within them, and there are even contradictions in the conceptions articulated by single informants. Moore is well aware of the main reason for these differences. It's all a matter of semantics. One wonders whether Moore would try to define a term like *university education* using the same methods as he does to try and define CT, and whether the results would mean that we have a problem because there is no simple, unified definition of a university education?

Moore identifies some patterns in the academics' descriptions of the CT skills required in their disciplines. For example, the historians mention the ability to judge the reliability of a source more often than academics in the fields of philosophy, literature, or cultural studies. Of course this is not to say that this skill is any less valuable to students in the other faculties. Thus, the reader may wonder how useful this information is in relation to the third part of Moore's task: the implications for teaching, as it may be argued that all students should develop this skill.

Those looking for a straight answer in Chapter 6 "Critical Thinking: So What Is It?" will be disappointed. Moore compares the task of describing CT to describing a family. Although a family has resemblances, there are also various differences—eye colour, hair, build, temperament, and so on. There are also similarities or resemblances in the various characteristics referred to when describing CT. Rather than identifying the commonalities in the concept and recommending that these be taught, Moore, rather unhelpfully, focuses on the differences. He creates a list of seven categories with two descriptive options in each. The idea is that academics consider the two options in each category, or "dimension of difference" (p. 212), to establish where their particular discipline sits on a continuum between the two terms. For example, is the thinking style in a given discipline more heuristic or hermeneutic? Is it more epistemic or deontic? The implication is that this will help achieve mutual understanding among academics of exactly how CT varies between disciplines and guide students who want to know exactly what is expected of them. Whether you can get the academics in a given discipline, philosophy for example, to agree on where they collectively sit on the continuum in each category and whether this might resemble the pattern determined by a philosophy faculty in another university is another matter entirely. And because Moore devotes 11 pages to explaining these abstract terms and their differences, the reader may wonder how helpful the resulting information would be to freshman students who may be wondering what kind of CT styles are expected in their chosen faculty.

Although it addresses one of the author's three key questions, the final chapter's "Conclusions and Implications for Teaching" are brief—only eight out of 234 pages. Moore suggests a transdisciplinary or unified approach to teaching CT skills, which my own experience has shown already happens in some universities with required freshman CT courses.

Moore's research data indicate more commonalities than differences among the various academics' conceptions of CT, even though they may use different words to describe what they value and teach. However, this book is focused more on the differences than the commonalities—something the reader may find frustrating and ultimately not useful in the application of a transdisciplinary approach. One gets the sense that, rather than quibbling about differences, semantic or otherwise, it would have been more productive to ask the academics in this study to consider a taxonomy of thinking skills (see Ennis, 1987, for example), and discard from it the skills they felt were of no use to students studying in their discipline. We could then compare the skills that all disciplines find useful and go ahead with the business of ensuring that all students acquire these common skills. Individual disciplines could then expand on this core knowledge according to their own preferences.

These limitations notwithstanding, university teachers and program administrators looking to foster critical thinking in Japan can find much food for thought in this contribution from Tim John Moore.

References

Ennis, R. (1987). A taxonomy of critical thinking abilities and dispositions. In J. Baron & R. Sternberg (Eds.), *Teaching thinking skills* (pp. 9-26). New York, NY: W. H. Freeman.

Wittgenstein, L. (1958). *Philosophical investigations* (2nd ed.). Oxford, UK: Basil Blackwell.

***Social Dimensions of Autonomy in Language Learning*. Garold Murray (Ed.). Basingstoke, UK: Palgrave Macmillan, 2014. xi + 277 pp.**

Reviewed by
Huw Davies
Komazawa University

Autonomous learning is often viewed as an individual act, an idea supported by Henri Holec's seminal definition of language learner autonomy: "the ability to take charge of one's learning" (1981, p. 3). However, this book takes another stance: Autonomy requires social interaction, with the focus moved to "the role of social and contextual processes" rather than individual agency (p. 135). This book is likely to appeal to a wide range of readers because it encompasses so many language learning contexts from all over the world and online and bridges both theory and practice.

The book is ordered into three sections that explore how emotions, space, and politics interact with autonomy in language learning. These sections are sandwiched between Garold Murray's informative Introduction and Conclusion, in which he draws on the literature to illustrate how proponents in the field are developing the definition of autonomy: from independence to interdependence to the inclusion of a wider social dimension.

In Part I, Christine O'Leary presents a study on strategy use and awareness raising to promote learner autonomy. She has found that learners use language learning strategies to cope with anxiety issues (p. 32). For O'Leary, autonomy is multidimensional, and there is no autonomy without social interaction. Tim Lewis also focuses on the complexity of autonomy, suggesting that unadulterated individualism is both unusual and inappropriate in society. Lewis draws on theory and research in the areas of psychology and anthropology to explore how social behavior impacts autonomous learning. Based on her study of high school students in Japan, Tomoko Yashima develops the powerful idea of autonomous dependency, in which the decision to accept advice and follow instructions may be seen as a demonstration of assertiveness and autonomy, rather than regarded as submission.

In Part II, Garold Murray, Naomi Fujishima, and Mariko Uzuka look at how social learning spaces can offer a different quality of interaction to classrooms and can encourage learners to develop their L2 identities. The writers suggest that although autonomy comes from the individual, it can only be fostered in a space that allows freedoms not often found in a classroom. Alice Chik and Stephan Breidbach, in a cross-cultural study, discover the difficulty of agreeing on an online space in which two groups from different cultures are comfortable to share and communicate. Linda Murphy, in a chapter on distance learning, explores how technology alters the manner in which learning takes place and focuses on shifting learner identities and the relationship between learner autonomy and self-determination. Diego Mideros and Beverly-Anne Carter describe and evaluate an interdependent autonomy project in which the students in a listening class identify and recommend Spanish Youtube videos as learning materials and comment on each other's choices. They identify the various outcomes, including learner reflection on their own learning through this approach, as well as the growth that resulted from coping with the unexpected linguistic and social demands of the project.

In Part III, chapters by Liliane Assis Sade and Andy Barfield focus on ways of developing and empowering new voices. For Sade, autonomy means using "one's own voice" (p. 155), and learners are empowered by actively expressing themselves in the community. She gives insights into the nonlinear, complex nature of autonomy once its social dimension is acknowledged and explored. Barfield, reviewing the evolution of JALT's Learner Development Special Interest Group (LD SIG), looks at the creation and shaping of local learning communities of teachers. This has been achieved through the development of opportunities for shared teacher reflection and collaborative research and writing, with a deliberate positioning of teacher identities to include the status of colearners both with other teachers and with students. (Incidentally, even this book review is part of the LD SIG's collaborative, community-building ethic.) David M. Palfreyman's concern is with learners' help-seeking behavior and the ways in which sociocultural prejudices and individual feelings affect their ability to maximize their learning potential. He describes how autonomous behavior involves drawing on resources found in the local community. E. Desirée Castillo Zaragoza explores the relationship between autonomy and social class, confirms the importance of learning with a purpose, and describes the visualization of a future self helps to encourage autonomous learning. Her focus is on the ways in which a self-access center functions as a point of contact with the L2 for people from different backgrounds.

One theme running through the book is the critical importance of developing an identity or voice in the target language. Castillo Zaragoza shows socioeconomic class to be a potentially decisive factor in creating an imagined self (p. 204). This aspect of autonomy, although individual in the sense that it is unique to one person, is undoubtedly related to interaction with other people. Murray, Fujishima, and Uzuka demonstrate the importance of having a safe and comfortable space where the learner has control (p. 91), but Yashima confirms the role of the educator as a guide because emerging L2 identities are unstable and not fully formed (p. 76). Also, it is important that learners' identities beyond that of language students are acknowledged and respected in the learning setting. As Murphy suggests, unexpected learning stemming from informal interactions is an essential part of developing an L2 identity (p. 131).

Social Dimensions of Autonomy in Language Learning presents persuasive arguments for viewing autonomy in language learning as innately social. This is a clear theme throughout the book, with the majority of authors using Holec (1981) as a starting point before emphasizing the importance of the social context of autonomy. Development of learner autonomy is an individual pursuit, but in the sense that it is personal, rather than in the sense that it is individualistic. Lewis sums this up well by explaining that working effectively with others does not necessitate having the same goals as they do (p. 58). In much the same way, by being anchored in such different learning settings, each individual chapter offers a fresh perspective to the message of the book as a whole. This message, expressed through the many voices and contexts, balancing theoretical discussion and reports of practice, is one that deserves the attention of all language educators who seek to increase their understanding of and support for language students' growth as autonomous learners.

Reference

Holec, H. (1981). *Autonomy and foreign language learning*. Oxford, UK: Pergamon Press.

***What English Language Teachers Need to Know Volume III: Designing Curriculum*. MaryAnn Christison and Denise E. Murray. New York, NY: Routledge, 2014. xxii + 255 pp.**

Reviewed by
Rick Derrah
Kwansei Gakuin University

Authors MaryAnn Christison and Denise E. Murray have certainly crafted a catchy primary title—*What English Language Teachers Need to Know*—for their books in the Routledge ESL & Applied Linguistics Professional Series. *Volume I: Understanding Learning* (2011a) and *Volume II: Facilitating Learning* (2011b) are followed by the authors' latest, *Volume III: Designing Curriculum*, written to help the reader "understand and work with the theory and practice of developing ELT curricula in a variety of contexts and for a variety of levels" (p. xx). These companion texts are designed for "pre-service teachers and teachers new to the field of ELT" (p. xviii), but *Volume III* can also provide a good refresher for those looking to revisit the basics of curriculum development or be used as classroom textbook for a course on curriculum design.

The latter purpose is reflected in the consistent structure of the chapters, which makes the material more approachable and easier to understand. Each chapter starts with a vignette, an introduction, and a definition of key terms. The vignette is a short story that places the theme of the chapter into a real world situation. These situations, in locations such as the United States, Australia, and Thailand, provide an interesting set of stories and place the role of curriculum in language teaching within a larger international context. In addition to the vignette, introduction, and definitions of key terms, the authors have added three types of tasks for readers: *reflect*, *expand*, and *explore* tasks. The reflect task, found in the early sections of each chapter, consists of questions designed to encourage the reader to focus on certain elements of the chapter. In the middle of each chapter, explore tasks lead the reader to apply the theme of the chapter to their own context by conducting interviews, examining textbooks and materials, or considering their practice. Near the end of the chapter are expand tasks. These tasks challenge the reader to go beyond their own context to look at the larger picture and to seek out additional resources linked to the unit theme. Finally, each chapter ends with a set of discussion questions and references.

The book is divided into six parts. Part I "Context for ELT Curriculum" contains three chapters. Chapter 1 "The Nature of Curriculum Design" lays out the chapters that follow and gives a short description of the book overall. Chapter 2 "Social, Political, and Historical Contexts" presents the realities of curriculum beyond simply the written form with a discussion of the differences between recommended, written, supported, taught, tested, learned, and hidden curricula. The final two chapters, "The World of English Teaching" and "The Technological Context," address two contexts that are surely of interest to individual teachers and those teams developing courses and programs.

Part II "Key Processes in Curriculum Design" includes Chapters 5 through 7. Chapter 5 introduces the "Cycle of Curriculum Design" in three parts: planning, implementing, and evaluating. This chapter is also where the authors explain the planning part of the cycle. Chapter 6 "Using Curriculum to Connect Lessons, Courses, and Programs" describes implementing curriculum across these three phases, and Chapter 7 "Quality Assurance and the Curriculum" discusses evaluation within the cycle of curriculum design. Evaluation here does not mean a discussion of learner assessment, but rather an overall method for quality assurance such as accreditation. The chapter also includes lists of methods for continuous improvement and quality standard development, but does not go into detail on how to develop or implement these ideas.

Parts III to VI then describe 14 different types of curricula. The authors organize their description into the following categories: Part III "Linguistic-based Curricula" (Chapters 8-13), Part IV "Content-based Curricula" (Chapters 14-15), Part V "Learner Centered Curricula" (Chapters 16-18), and Part VI "Learning Centered Curricula" (Chapters 19-21).

The chapters in Parts III to VI follow the same structure as in the earlier parts of the book. Two elements of this structure, the vignette and the definition, serve as useful introductions to each type of curriculum. They will be particularly helpful for new teachers who may be encountering some of these terms and distinctions for the first time. A wide range of curricula is introduced; however, each chapter is only 10 to 12 pages long, thus limiting the depth of coverage. In addition, there is little connection to the process of curriculum design as detailed in Part II. New teachers should gain an understanding of each of the curricula discussed, but may find it difficult to transfer this understanding to practical implementation within their own contexts. These later chapters could be strengthened with more explicit connection to the earlier parts of the book, specifically the stages of planning, implementing, and evaluating each of the curricula described. Readers looking to explore practical applications in context further may be intrigued by the case studies from Macalister and Nation (2011) or become interested in seeing what non-EFL, cross-disciplinary titles on instructional design might also offer. Finally, there is no concluding chapter, a lost opportunity to compare, contrast, and highlight the varied curricula introduced and serve as a guide to learners and decision makers.

Overall, Christison and Murray's *What English Language Teachers Need to Know Volume III: Designing Curriculum* has three strong points. First, it is an easy-to-read and approachable book with varied examples, definitions of key terms, and a clear structure to each chapter. Second, it covers a lot of material, from the context and the development of curriculum to an introduction of 14 different types of curricula. Third, the resources within each chapter allow the book to be easily utilized in a classroom setting or in the great number of distance education TESOL courses now available. Despite the lack of a stronger connection between the first two parts of the book and the introduction of various curricula, this volume is a good introduction to curriculum design for novice teachers and a useful refresher for experienced teachers looking to revisit the subject or taking on course coordinator or program administration roles.

References

Macalister, J., & Nation, I. S. P. (Eds.). (2011). *Case studies in language curriculum design: Concepts and approaches in action around the world*. New York, NY: Routledge.

Murray, D. E., & Christison, M. (2011a). *What English language teachers need to know volume I: Understanding learning*. New York, NY: Routledge.

Murray, D. E., & Christison, M. (2011b). *What English language teachers need to know volume II: Facilitating learning*. New York, NY: Routledge.

Japanese: A Linguistic Introduction. Yoko Hasegawa. Cambridge, UK: Cambridge University Press, 2015. xxii + 392 pp.

Reviewed by
Mark Rebuck
Meijo University

The word *introduction* may be in the title, but Yoko Hasegawa's book is a comprehensive, 28 chapter exploration of the Japanese language. Part 1 consists of five chapters, the first providing an overview of the typological characteristics of Japanese. There follows a further three chapters surveying the dialects, sound, and writing system of what is classified as an agglutinative language with "the [world's] ninth largest native speaker population" (p. 3). The Japanese lexicon is covered in Part 2, foundations of grammar in Part 3, the main clause types in Part 4, and clause linkage in Part 5. Covering well over 100 pages, Part 6 addresses topics related to pragmatics.

The book opens with an historical overview of the Japanese language placed in the context of social and political changes. Hasegawa describes, for example, how the topic particle *wa* is a product of a sound change that occurred during the Late Old Japanese period (794-1192), the start of which saw the relocation of the capital from Nara to Kyoto. At the end of Chapter 1 the author describes how the Meiji government's push for industrialization and militarization shaped the development of the modern Japanese language (1867 to present). An unprecedented movement of the population at this time was a factor leading to calls to establish *hyoujungo* (a standard language) "not only to foster communication but also to awaken nationalism" (p. 14).

Chapter 2 surveys dialects, an appreciation of which is important to understand the diversity of modern Japanese as well as to "probe its historical development" (p. 17). Readers without a solid foundation in phonetics may find this and the following chapter on the Japanese sound system somewhat challenging. However, the authentic video clips on the website that accompanies the book should help to clarify the written explanation. The website also complements other chapters, including that on backchanneling (Chapter 25), where a clip from a chat show brings to life a conversation analysis highlighting *aizuchi*, "short utterances [that] do not claim the floor" (p. 320).

Reading this book gave me fresh insights into fundamental features of the language. In Chapter 3, for example, I was reminded that "vowel length is not distinctive in spoken English . . . [but in Japanese] elongating a vowel can change the meaning." (p. 35). So, *slow* in English still means *slow* even when it is pronounced *sloooow*, which is different from Japanese where a long vowel can denote another word (as I am forever being reminded when I confuse *obasan* [aunt] with *obaasan* [grandmother]).

The above example illustrates a feature of the book that English teachers may find of value: There are numerous comparisons between the two languages. In Chapter 5, on the Japanese lexicon, for example, Hasegawa explains that, although English has many verbs expressing both the action and the way it is performed, in Japanese "manner is . . . typically expressed by an ideophone" (p. 72). Compare, for instance, the English *howl* (which expresses both action and manner) with the Japanese *wanwan* (manner) and *naku* (action).

In places, some of the more involved linguistic analyses were difficult for me to grasp. In the Preface, Hasegawa states that the book is intended as "a college-level reference book . . . that can also serve as the principal textbook in an introductory course in Japanese linguistics" (xix). Many future readers in that target group (students taking a course) would thus have access to an expert who could provide supplementary explanations. Those making the journey alone may be slightly disappointed by the absence of both a key to the phonetic symbols used in the early chapters and a glossary of the main linguistic terminology.

My struggle to follow certain sections was by no means a reflection of Hasegawa's writing, which is lucid and generally considerate of the reader. She provides numerous example sentences that serve to illustrate and clarify the explanations. It should be noted, however, that these examples are written in *romaji* (the Roman alphabet). This feature will increase the book's accessibility for many readers, but may initially be distracting to those accustomed to authentic Japanese texts. Such readers, however, will perhaps get a frisson of pleasure from the opening lines of Chapter 4, an enjoyable survey of the Japanese writing system, which state that "Japanese . . . employs what is arguably the most convoluted writing system ever devised in human history" (p. 43). Those who are still struggling with Chinese characters may wish that the *kana no kai* (Kana Club)—who we learn campaigned in the 1880s to abolish kanji—had succeeded (p. 50).

Readers wishing to quickly check specific language points, such as the difference between *kara* and *node*, may find more straightforward explanations elsewhere (see, for example, Seiichi & Tsutsui, 1995). Yet, Hasegawa's book is not short on practical language advice. In the section on pronouns, for instance, she explains that the use of the second person pronoun should be limited because it can create "the impression of pointing to . . . the addressee with a finger and saying 'YOU!'" (p. 68).

The book is studded with illuminating examples and studies. For example, in Chapter 11 on causatives, Hasegawa mentions a movie of the 1930s, whose title *Nani ga kanojo o so saseta ka* [What made her do it?] sounded "peculiar to many Japanese ears" (p. 146) because it juxtaposed an inanimate, abstract subject with a causative verb. The film's success, however, owed much "to its linguistically eccentric title" (p. 146). In the chapter on sentence-final particles, in the pragmatics' section of the book, Hasegawa describes a study into the acquisition of *yo*, *no*, and *ne* by children, including those children with cognitive disabilities (p. 305).

Hasegawa also occasionally inserts short anecdotes about the learners she has taught. In the discussion of speech acts (Chapter 19), for example, she illustrates the difficulty of rejecting compliments in Japanese by recalling a student of hers at the University of California, Berkeley (where Hasegawa is a professor), who responded to praise by saying "*Ie, watashi wa baka desu*" [No, I'm an idiot]. Hasegawa explains why this response was inappropriate (p. 250). Stories of the trials and tribulations of learners of Japanese make good reading, as attested to by the popularity of a particular manga series (Tomita, 2009), and perhaps such content could fill a less scholarly future publication by Hasegawa. The strength of the present book, however, is its eruditeness, particularly reflected in the various perspectives presented when dealing with linguistic controversies. Chapter 21, for example, covers in depth the objections from several scholars to Brown and Levinson's (1987/1978) politeness theory as applied to the Japanese honorific system.

The final chapter on gendered language encapsulates the good features of this book. There is historical background, an analysis that introduces several linguistic viewpoints (in this chapter on gendered language in soliloquy), and a touch of anecdote (Hasegawa tells us what she noticed about male and female speech after 25 years away from Japanese media).

Depending on one's previous experience studying linguistics and knowledge of Japanese, this book may be challenging in parts. Although I did not fully understand some pages, it was a read that, after my own two decades in Japan, provided a timely linguistic overview and left me with new insights into Japanese, a language will likely continue to challenge and fascinate me for my remaining years.

References

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge, UK: Cambridge University Press.

Seiichi, M., & Tsutsui, M. (1995). *A dictionary of intermediate Japanese grammar*. Tokyo: The Japan Times.

Tomita, T. (2009). *Nihonjin no shiranai nihongo* [The Japanese the Japanese don't know]. Tokyo: Media Factory.

English for Academic Purposes. Edward de Chazal. Oxford, UK: Oxford University Press, 2014. xix + 380 pp.

Reviewed by
Gregory Strong
Aoyama Gakuin University

With the Japanese Ministry of Education's ambitious goal of attracting 300,000 foreign students by 2020 (MEXT, 2008), this book is more relevant than ever for university educators and policymakers in Japan who are developing English for Academic Purposes (EAP) courses. In a refreshingly modest foreword, Edward de Chazal urges his readers to examine his ideas, to discuss them, and even to discard some. To start, he distinguishes between ELT, covering the language education of children to adults, and a more narrowly focused EAP. The latter "aims to develop students' skills and language to a target level of proficiency" at the same time as developing the "academic literacies" needed at a postsecondary institute (p. 11).

De Chazal's perspective is of an experienced educator in Britain, which has one of the largest foreign student enrolments in the world. With Britain's Higher Education Statistics Authority reporting 435,000 foreign students in 2013-14 (UK Council for International Student Affairs, 2015), much has been done to standardize university entrance requirements and academic support. Frequently, de Chazal refers to BALEAP, the British Association of Lecturers in EAP. For instance, when discussing teacher preparation, he references BALEAP's Competency Framework of teacher knowledge of academic contexts, disciplinary differences, academic discourses, and professional development (Chapter 2, pp. 49-54).

Impressively, the book summarizes most of the important research in EAP, beginning in Chapter 1 with a look at the influences on EAP. The volume is divided into chapters on a history of the field, texts, language, reading, writing, speaking, listening, materials assessment, and technologies. A bestselling volume in the Oxford *Handbooks for Language Teachers* series, it is valuable as a reference book for educators and for teacher training.

His thoughtful Chapter 3 "Texts" refers to text genre and cohesion as an aid to teaching. He offers a series of tables cataloguing types of academic texts such as a comparison essay, a review, and a case study with the essential elements of comparison, criticism, and a report.

Chapter 4 "Language" proves the most compelling and comprehensive in the book as he tackles such linguistic characteristics of academic language as text density, morphologically complex words, technical vocabulary, and complex noun phrases. Most textbooks teach grammar based on verb tenses and modals, but de Chazal emphasizes noun phrases. As proof, he cites a paper on fair trade with noun phrases like "changing global markets" and "participation in organic and fair trade markets" (p. 94). He argues that students should learn vocabulary in context, learning words with text rather than in word lists. Provocatively, he critiques academic word lists for often omitting useful nonacademic words, including those in journal articles for hedging or limiting claims or useful prepositional verb phrases like *applied to* and *referred to*. In the related debate in the field of English for Specific Purposes on whether teachers should be specialists in language or content, de Chazal argues for the former, stating that EAP teachers should be responsible for carefully selecting target words from their students' texts or field.

These strong sections make de Chazal's chapter on critical thinking (Chapter 5) all the more disappointing. He starts well by categorizing the thinking processes in academic writing. Summarizing consists of "identifying the author's stance in the text" (p. 130) and citing references as a means of providing support as well as "determining an appropriate way of presenting this evidence" (p. 128). Next, he addresses critical reading by urging that students assess whether an author has supplied adequate justification for his or her claims. Unfortunately, he offers precious few specifics for actually teaching critical thinking in EAP.

De Chazal's book gets much stronger when he discusses academic reading and writing (Chapters 6 and 7). His detailed tables outline different aspects of reading and he describes classroom activities for developing greater proficiency and speed and provides a student questionnaire on reading habits. On academic writing, he presents corpora for different academic disciplines. Helpfully, he describes essay introductions as more than just a thesis statement, but rather as a contextualization. Writers should situate the topic, time, place, and context by providing a rationale and introducing terms. In teaching writing, he contrasts the functional, process-writing, genre-based, critical literacy, and student-centered approaches. He includes reflective questions for student writers such as "Why am I writing?" and "Have I made connections between the material within and throughout the text clear and coherent?" (p. 286). Finally, he offers exemplars with prompts for student writing tasks and explanations for marking.

The other two skills are given briefer treatment. In "Listening" (Chapter 8), de Chazal argues that this skill is now a multimodal experience with images, video, audio, and online materials. His helpful table categorizes the types and characteristics of academic listening, from lectures, presentations, seminars, discussions, and tutorials to group projects. He also analyzes the text of an academic lecture. Chapter 9 then includes the author's description of the academic speaker as a persuader. He offers different types of practice, contrasts effective and ineffective content and delivery, covers the stages for students preparing presentations, and highlights how teachers should evaluate them.

In Chapter 10, de Chazal urges teachers to undertake needs analyses before developing materials for their classes. He includes some practical advice—the types of informants and the processes a teacher should follow, questionnaires, readings, and criteria for using published EAP materials. He offers ideas for "low-stakes tests and assessments" and he writes on the pros and cons of utilizing institutional tests such as IELTS or the Pearson Test of English. In this brief section, he also suggests further resources, particularly Carr (2011).

The final chapter is contributed by Aisha Walker, who attempts to cover the rapidly changing field of digital literacy. Unfortunately, it is the least satisfying chapter in an otherwise splendid book. It discusses students citing the *grey literature* of blogs, postings, and tweets and it provides some convenient websites including those for creating word clouds and word mapping. However, these are familiar to many and they can easily be googled.

Concluding with a user-friendly glossary and references, de Chazal contends that EAP teachers can help shape the field of ELT. In this excellent overview, he shows that EAP can indeed become a transformative discipline as inspired teachers help improve their students' lives.

References

Carr, N. (2011). *Designing and analyzing language tests.* Oxford, UK: Oxford University Press.

MEXT. (2008, July 29). *Outline of the student exchange system: Study in Japan and abroad*. Retrieved from http://www.mext.go.jp/a_menu/koutou/ryugaku/081210/001.pdf

UK Council for International Student Affairs. (2015, May 18). *International student statistics: UK higher education*. Retrieved from http://www.ukcisa.org.uk/Info-for-universities-colleges--schools/Policy-research--statistics/Research--statistics/International-students-in-UK-HE/

Translanguaging: Language, Bilingualism and Education. Ofelia García and Li Wei. Basingstoke, UK: Palgrave MacMillan, 2014. x + 165 pp.

Reviewed by
Blake Turnbull
Kyoto University

Ofelia García and Li Wei's *Translanguaging: Language, bilingualism and education* is a revealing insight into what has become an ever relevant field in today's multilingual world. Expanding on their own earlier work (see for example García, 2009; Wei, 2011), as well as on that established by others (e.g., Canagarajah, 2011; Creese & Blackledge, 2010; Hornberger & Link, 2012; Lewis, Jones, & Baker, 2012), García and Wei present a refined and developed concept of translanguaging in this two-part book: "Part I, Language and Translanguaging" and "Part II, Education and Translanguaging." The authors take the reader through a conscious-raising process that breaks down historic conceptions of language education, examining a range of pertinent topics including languaging, bi/multi/plurilingualism, and language education from both monolingual and bi/multilingual perspectives to address two key questions: *What is translanguaging?* and *What does a translanguaging approach mean for language and bilingualism on the one hand, and education and bilingualism on the other?*

Part I, "Language and Translanguaging" consists of two chapters and seeks to answer the question *What is translanguaging*? García and Wei begin Chapter 1 with a review of traditional concepts relating to language, bilingualism, multilingualism, and plurilingualism as they postulate the epistemological and theoretical grounding for their translanguaging approach to education. They examine the transformations in traditional understandings of language and bilingualism and consider the recent shift in research focus from language as a *product* to language as a *practice*, whereby the focus becomes "the speaker's *creative* and *critical use* of linguistic resources to mediate cognitively complex activities" (p. 10). The authors refer to the concept of *languaging*, which, they claim, views language not as a product or structure in the speaker's mind but as an ongoing process created through linguistic interaction with the world. García and Wei consider, and then dismiss, the traditional view of bilingualism as two autonomous linguistic systems and critique, in part, bilingualism from a linguistic interdependence perspective (based on Cummins's [1979] common underlying proficiency model). Instead, they advocate for *dynamic bilingualism*, entailing a single, expanded linguistic system from which speakers draw features in accordance with the rules of societally constructed languages. It is here, in the emergent paradigm of dynamic bilingualism, that the authors situate their concept of translanguaging.

In Chapter 2, García and Wei trace the term translanguaging from its original Welsh inception, defined as "a pedagogical practice where students are asked to alternate languages for the purposes of receptive or productive use" (p. 20) and examine the various extensions, interpretations, and related terminology that have risen since. The authors define and elaborate on their own interpretation of translanguaging as the fluid language practices of bi- or multilinguals who move between and beyond the systems in their linguistic repertoire, drawing upon multiple semiotic resources appropriate to given contexts to language and make meaning of both themselves and their surrounding environments.

Part II "Education and Translanguaging" consists of five chapters and examines the transition of translanguaging from its theoretical grounding to an applied pedagogical practice. In Chapters 3 and 4, García and Wei explore the transformational role of translanguaging on various educational programs. The authors revisit the traditional understandings of bilingual and foreign or second language education in which the learners' two languages are deliberately kept separate in accordance with education and institutional policies. They then dismiss this act of language separation, instead advocating for the integration, expansion, and extension of learners' new and existing language practices. Building upon the work of Wei (2011), García and Wei discuss the importance of creating a *translanguaging space*, in which linguistically diverse learners integrate social spaces and language codes previously practiced separately, giving learners the chance to contest the language separation ideologies of traditional monolingual and bilingual education.

In Chapters 5 and 6, García and Wei explain how, and for what purposes, translanguaging can be used by students to learn and by educators to teach. The authors claim that, for students, *pupil-directed translanguaging* is a way to develop new language practices in interrelationship with practices they are already doing, in order to become more knowledgeable. They provide several examples of translanguaging in the classroom, from kindergarten children to older students, in which learners draw on all of their linguistic resources to complete a variety of tasks. The authors also discuss *teacher-directed translanguaging*, which they define as a planned and structured transformative pedagogy that is holistic in nature and teaches to all students in a given class. They provide five case studies of US teachers utilising translanguaging to teach secondary school math, social studies, science, and English language arts, and primary school ESL.

Finally, in Chapter 7, García and Wei summarise the principles and strategies of translanguaging as a substantive pedagogy for teaching and learning. The authors provide recommendations on how translanguaging can be used by both students and teacher in L2 reading, writing, speaking, and listening, emphasizing its adaptability to all types of educational programs involving all types of students. The authors stress the importance of teachers developing a critical consciousness about the linguistic diversity of their learners and recommend that they learn to construct curricula and pedagogies that build on these through differences. They then present two major challenges that remain for translanguaging and education, *teaching to do translanguaging* and *using translanguaging in assessment*, and suggest that, to address these issues, an epistemological change is needed surrounding the negative ideologies of native language use in L2 education, which is beyond what most institutions and teachers currently accept.

Overall, this book is an incredibly accessible and well-rounded insight into the ever-growing field of translanguaging as an approach to bi- and multilingual education. García and Wei include abstracts at the beginning of all seven chapters to provide an overview and guide for the discussions presented within; 11 diagrams and images engage readers and help them visualise the many multimodal representations of translanguaging; and relatable real-life examples clarify the concept of translanguaging for anyone who has had difficulty understanding it in the past.

One of the book's greatest strengths is its ability to push the boundaries and question long-held beliefs of traditional language education. It must be said, however, that this book can come across as both provocative and controversial in that it challenges old conceptions of mono/bi/multilingual language education. Readers must therefore be open and prepared to consider and accept the concept of learners having but a single linguistic repertoire as is presented in the text.

Although the authors suggest that translanguaging has the potential to revolutionise education not only for bi- and multilingual learners, but also for emerging bilinguals, the majority of the transcripts and examples presented in this book are situated in either the UK or US. This limitation largely ignores language education in Asian and European countries in which EFL is the dominant form of instruction, and where I believe a translanguaging approach to language learning may have significant benefits to language education. Thus, whilst this book is particularly accessible to a readership of scholars and teachers involved in bi- and multilingual fields, I would also recommend this book to teachers and educators in EFL countries where first language use is often frowned upon, in the hope that a translanguaging approach may one day be integrated into foreign language classrooms as a standard practice of instruction and learning.

References

Canagarajah, S. (2011). Translanguaging in the classroom: Emerging issues for research and pedagogy. *Applied Linguistics Review, 2,* 1-28. http://dx.doi.org/10.1515/9783110239331.1

Creese, A., & Blackledge, A. (2010). Translanguaging in the bilingual classroom: A pedagogy for learning and teaching? *The Modern Language Journal, 94,* 103-115. http://dx.doi.org/10.1111/j.1540-4781.2009.00986.x

Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism Toronto, 19,* 197-202.

García, O. (2009). *Bilingual education in the 21st century: A global perspective*. Malden, MA: Wiley-Blackwell.

Hornberger, N. H., & Link, H. (2012). Translanguaging and transnational literacies in multilingual classrooms: A bilingual lens. *International Journal of Bilingual Education and Bilingualism, 15,* 261-278. http://dx.doi.org/10.1080/13670050.2012.658016

Lewis, G., Jones, B., & Baker, C. (2012). Translanguaging: Developing its conceptualisation and contextualization. *Educational Research and Evaluation*, *18,* 655-670. http://dx.doi.org/10.1080/13803611.2012.718488

Wei, L. (2011). Moment analysis and translanguaging space: Discursive construction of identities by multilingual Chinese youth in Britain. *Journal of Pragmatics, 43,* 1222–1235. http://dx.doi.org/10.1016/j. pragma.2010.07.035

Information for Contributors

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

Editorial Policy

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

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

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

Guidelines

Style

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

Format

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

Submission Procedure

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

1. Cover sheet with the title and author name(s).
2. One (1) copy of the manuscript, with no reference to the author. Do not use running heads.
3. Contact information sheet, including one author’s full address and, where available, a fax number.
4. Abstract (no more than 150 words).
5. Japanese translation of the title and abstract, if possible (no more than 400ji).
6. Biographical sketch(es) (no more than 50 words each).

Submissions will be acknowledged within 1 month of their receipt. All manuscripts are first reviewed by the Editor to ensure they comply with *JALT Journal* Guidelines. Those considered for publication are subject to blind review by at least two readers, with special attention given to (1) compliance with *JALT Journal* Editorial Policy, (2) the significance and originality of the submission, and (3) the use of appropriate research design and methodology. Evaluation is usually completed within 3 months. Each contributing author of published articles and Book Reviews will receive one complimentary copy of the *Journal* and a PDF of the article (Book Reviews are compiled together as one PDF). *JALT Journal* does not provide off-prints. Contributing authors have the option of ordering further copies of *JALT Journal* (contact JALT Central Office for price details).

Restrictions

Papers submitted to *JALT Journal* must not have been previously published, nor should they be under consideration for publication elsewhere. *JALT Journal* has First World Publication Rights, as defined by International Copyright Conventions, for all manuscripts published. If accepted, the editors reserve the right to edit all copy for length, style, and clarity without prior notification to authors.

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

Please send submissions in these categories or general inquiries to:

jj-editor@jalt-publications.org

Anne Howard, *JALT Journal* Editor

Japanese-Language Manuscripts

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

jj-editorj@jalt-publications.org

Ken Urano, *JALT Journal* Japanese-Language Editor

Reviews

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

jj-reviews@jalt-publications.org

Greg Rouault, *JALT Journal* Reviews Editor

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

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

日本語論文投稿要領

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

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

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

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

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

投稿原稿送り先またはお問い合わせ:

〒062-8605 北海学園大学　経営学部
JALT Journal 日本語編集者　浦野　研
電話:(011)841-1161(代) Fax:(011)824-7729
jj-editorj@jalt-publications.org

JALT Journal 第 38 巻　第 1 号

2016年 4月20日　印刷
2016年 5月 1日　発行
編集人　アン・マクレラン・ハワード
発行人　ロイド・キャロライン
発行所　全国語学教育学会事務局
〒110-0016　東京都台東区台東1-37-9 アーバンエッジビル5F
TEL(03) 3837-1630; FAX(03) 3837-1631
印刷所　コーシンシャ株式会社
〒530-0043　大阪市北区天満1-18-4天満ファーストビル301 TEL(06)6351-8795