

Japan Association for Language Teaching

Articles

- 5 Cognitive Processes in Second Language Word Association — *John P. Racine*
- 27 Japanese Secondary School Teachers' Beliefs and Practices Regarding Communicative Language Teaching: An Exploratory Survey — *Takako Nishino*
- 51 日本語話者大学生の英語学習動機の変化—国際イベントへのボランティア参加の効果 — [Changes in Japanese University Students' Motivation to Learn English: Effects of Volunteering in an International Event]
— 酒井英樹・小池浩子 (*Hideki Sakai, Hiroko Koike*)
- 69 An Investigation into the Effect of Raw Scores in Determining Grades in a Public Examination of Writing — *David Coniam*
- 85 A Comparative Analysis of the Japanese University Entrance Senta Shiken Based on a 25-Year Gap — *Michael (Mike) Guest*
- 105 Tracking and Targeting: Investigating Item Performance on the English section of a University Entrance Examination over a 4-Year Period
— *Christopher Weaver & Yoko Sato*

Reviews

- 129 *Language Policy, Culture, and Identity in Asian Contexts* (Amy B. M. Tsui & James W. Tollefson) — Reviewed by R. A. Brown
- 131 *Teaching English as a Foreign or Second Language: A Teacher Self-Development and Methodology Guide* (2nd ed.) (Jerry G. Gebhard) — Reviewed by Mieko Fukushima
- 134 *The Experience of Language Teaching* (Rose M. Senior) — Reviewed by Darren Elliott
- 136 *A Friendly Approach to English for Academic Purposes* (Sonia Sonoko Strain) — Reviewed by John Nevara
- 139 *Connecting Speaking & Writing in Second Language Writing Instruction* (Robert Weissberg) — Reviewed by Christian Perry
- 141 *Motivation, Language Attitudes and Globalisation: A Hungarian Perspective* (Zoltán Dörnyei, Kata Csizér, & Nóra Németh) — Reviewed by Akiko Tsuda
- 143 *Motivation and Experience in Language Learning* (Yoshiyuki Nakata) — Reviewed by Phillip Barkman

JALT Journal

Volume 30 • No. 1

May, 2008

Editor

Steve Cornwell

Osaka Jogakuin College

Associate Editor

Ian Isemonger

Kochi Women's University

Reviews Editor

Yuriko Kite

Kansai University

Japanese-Language Editor

Yoshinori J. Watanabe

Sophia University

Editorial Board

David Aline

Kanagawa University

David Beglar

Temple University—Japan Campus

William Bradley

Ryukoku University

James Dean Brown

University of Hawai'i, Manoa, USA

Charles Browne

Aoyama Gakuin University

Christine Pearson Casanave

Temple University—Japan Campus

Eton Churchill

Kanagawa University

Ann Ediger

Hunter College, City University of New York, USA

Rod Ellis

University of Auckland, New Zealand

Sandra Fotos

Senshu University

Peter Gobel

Kyoto Sangyo University

Eli Hinkel

Seattle University, USA

Yuri Hosoda

Kanagawa University

Nicholas O. Jungheim

Waseda University

Sayoko Yamashita

Tokyo Medical and Dental University

Laura MacGregor

Gakushin University

Tim Murphey

Kanda University of International Studies

Mary Goebel Noguchi

Ritsumeikan University

Aek Phakiti

University of Sydney, Australia

Cynthia Quinn

Konan University

Timothy J. Riney

International Christian University

Carol Rinnert

Hiroshima City University

Thomas Robb

Kyoto Sangyo University

Gordon Robson

Showa Women's University

David Shea

Keio University

Greg Sholdt

Kobe University

Donna Tatsuki

Kobe City University of Foreign Studies

Deryn Verity

Osaka Jogakuin University

Minoru Wada

Meikai University

Additional Readers: Michael Lessard Clouston, Neil Cowie, Dale Griffie, Neil Heffernan, Akihiro Ito, Darren Lingley, Hirohide Mori, Setsuko Mori, Naoyuki Naganuma, Bonny Norton, Robert O'Mochain, Hidetoshi Saito, Mark Sawyer, Andrea Simon-Maeda, Ken Urano, Deryn Verity, Martin Willis

JALT Journal Proofreading: Mayumi Fujioka, Seiji Fukazawa, Megumi Kawate-Mierzejewska, Aleda Krause, Steve McGuire, Cynthia Quinn, Jack Yohay

JALT Publications Board Chair: Kim Bradford-Watts

Layout: Pukeko Graphics **Cover Design:** The Word Works

JALT Journal on the Internet: www.jalt-publications.org/jj/

Website Editor: Salem Hicks

- 3 In this Issue
- 3 From the Editor

Articles

- 5 Cognitive Processes in Second Language Word Association
— *John P. Racine*
- 27 Japanese Secondary School Teachers' Beliefs and Practices Regarding Communicative Language Teaching: An Exploratory Survey
— *Takako Nishino*
- 51 日本語話者大学生の英語学習動機の変化—国際イベントへのボランティア参加の効果
— [Changes in Japanese University Students' Motivation to Learn English: Effects of Volunteering in an International Event]
— 酒井英樹・小池浩子 (*Hideki Sakai, Hiroko Koike*)
- 69 An Investigation into the Effect of Raw Scores in Determining Grades in a Public Examination of Writing — *David Coniam*
- 85 A Comparative Analysis of the Japanese University Entrance Senta Shiken Based on a 25-Year Gap — *Michael (Mike) Guest*
- 105 Tracking and Targeting: Investigating Item Performance on the English section of a University Entrance Examination over a 4-Year Period
— *Christopher Weaver & Yoko Sato*

Reviews

- 129 *Language Policy, Culture, and Identity in Asian Contexts* (Amy B. M. Tsui & James W. Tollefson) — Reviewed by R. A. Brown
- 131 *Teaching English as a Foreign or Second Language: A Teacher Self-Development and Methodology Guide* (2nd ed.) (Jerry G. Gebhard)
— Reviewed by Mieko Fukushima
- 134 *The Experience of Language Teaching* (Rose M. Senior)
— Reviewed by Darren Elliott
- 136 *A Friendly Approach to English for Academic Purposes* (Sonia Sonoko Strain)
— Reviewed by John Nevara
- 139 *Connecting Speaking & Writing in Second Language Writing Instruction* (Robert Weissberg) — Reviewed by Christian Perry
- 141 *Motivation, Language Attitudes and Globalisation: A Hungarian Perspective* (Zoltán Dörnyei, Kata Csizér, & Nóra Németh) — Reviewed by Akiko Tsuda
- 143 *Motivation and Experience in Language Learning* (Yoshiyuki Nakata)
— Reviewed by Phillip Barkman

JALT Journal Information

- 146 Information for Contributors (English and Japanese)

All materials in this publication are copyright ©2008 by JALT and their respective authors.

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 in Japan, along with 16 special interest groups (SIGs), and one forming chapter. JALT is one of the founders of PAC (Pan-Asian Consortium), which is an association of language teacher organizations in Pacific Asia. PAC holds 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 is a branch of IATEFL (International Association of Teachers of English as a Foreign Language).

JALT publishes *JALT Journal*, a research journal; *The Language Teacher*, a monthly periodical containing articles, teaching activities, reviews, and announcements about professional concerns; and *JALT International Conference Proceedings*.

The JALT International Conference on Language Teaching and Learning and Educational Materials Exposition attracts some 2,000 participants annually and offers over 400 papers, workshops, colloquia, and poster sessions. Each JALT chapter holds local meetings and JALT's SIGs provide information 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 automatic assignment to the nearest chapter or the chapter you prefer to join, copies of JALT publications, and reduced admission to JALT-sponsored events. JALT members can join as many SIGs as they wish for an annual fee of ¥1,500 per SIG. For information, contact the JALT Central Office or visit the JALT website at <<http://www.jalt.org>>.

JALT National Officers, 2008

President: Steve Brown
Vice President: Cynthia Keith
Auditor: Tadashi Ishida
Director of Treasury: Kevin Ryan
Director of Records: Donna Tatsuki
Director of Programs: Philip McCasland
Director of Membership: Ann Mayeda
Director of Public Relations: Sayoko Yamashita

Chapters

Akita, Chiba, East Shikoku, Fukui, Fukuoka, Gifu, Gunma, Hamamatsu, Himeji, Hiroshima, Hokkaido, Ibaraki, Iwate, Kagoshima, Kitakyushu, Kobe, Kyoto, Matsuyama, Miyazaki, Nagasaki, Nagoya, Nara, Okayama, Okinawa, Omiya, Osaka, Sendai, Shinshu, Shizuoka, Tokyo, Toyohashi, Wakayama (forming), West Tokyo, Yamagata, Yokohama.

Special Interest Groups

Bilingualism; College and University Educators; Computer-Assisted Language Learning; Extensive Reading (forming); Gender Awareness in Language Education; Global Issues in Language Education; Japanese as a Second Language; Junior and Senior High School Teaching; Learner Development; Lifelong Language Learning; Materials Writers; Other-Language Educators; Pragmatics; Professionalism, Administration, and Leadership in Education; Study Abroad (forming); Teacher Education; Teaching Children; Testing and Evaluation.

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; E-mail: jco@jalt.org;

Website: <www.jalt.org>

In this Issue

Articles

This issue contains a broad assortment of topics that range from cognitive processes and lexicon to teachers' beliefs to student motivation to university entrance exams. The issue starts off with an article that examines second language word association. **John P. Racine** studies similarities and differences in the cognitive processes native speakers and Japanese speakers of English use to access mental lexicons. In the next article, **Takako Nishino** presents her results from an exploratory survey of Japanese high school teachers' beliefs and practices regarding communicative language teaching. Next, **Hideki Sakai** and **Hiroko Koike** report on how volunteering at an international event changed university students' motivation when learning English. Then next three articles all examine university examinations. **David Coniam** compares the different results obtained on a Hong Kong public writing exam when using raw scores and when using Rasch analysis scores to determine pass levels. Next, **Mike (Michael) Guest** analyzes and reports on changes to the *Senta Shiken*. Finally, **Christopher Weaver** and **Yoko Sato** look at item performance (using Rasch analysis) over 4 years on a university entrance exam.

Reviews

In this issue we have seven book reviews. In the first one **R. A. Brown** reports on a volume that examines language policy, culture, and identity in an Asian context. Next, **Mieko Fukushima** reports on a book designed to be a self-study guide that looks at teaching English as a foreign or second language. **Darren Elliott** looks at a volume that examines how people develop into teachers after choosing to enter intensive language teaching courses. **John Nevara** examines a book that takes readers through the design and implementation of a course on English for Academic Purposes. Our next review, by **Christian Perry**, is of a book that examines how speaking and writing can be purposefully connected to provide better instruction in a second language writing class. This issue ends with two reviews on books that deal with motivation, a perennial interest among language teachers and students. **Akiko Tsuda** reports on a book that examines motivation, attitudes, and their relationship to globalization, while **Phillip Barkman** reviews a book that shows how language learning motivation and experiences are intertwined.

From the Editor

Mark Twain once quipped, "I am not an editor of a newspaper, and shall always try to do right and be good so that God will not make me one."

Hmmm...well, I must have done something wrong or have been bad because I have been an editor for JALT since 1994 but now it is time for me to pass the baton on.

Seriously, I count myself fortunate to have gotten involved with JALT Publications back in 1994 editing the handbook for the Nagoya JALT Conference in November 1995. Since then I have volunteered for JALT in different ways on both the local and national level, but usually with a publication focus, be it as a reader or reviewer, or as a newsletter, handbook, proceedings, or most recently, journal editor. With this issue I am stepping down as Editor of the *JALT Journal* and just want to say thank you to JALT for letting me serve in this position. It has been an honor and a joy! We have page restrictions so I cannot thank everyone I would like to but I trust you will bear with me as I thank a few people or groups: I want to thank the JALT Board of Directors who since 2003 have been very supportive of me, the *Journal*, and publications in general, **Junko Fujio** and the JALT Central Office, who are always helpful (and patient), the Publication Board and the Publication Board Chairs I have served under (**Brad Visgatis**, **Amanda O'Brien**, **Kim Bradford-Watts**), the Associate Editors I have worked with (first, **Deryn Verity**, and now, **Ian Isemonger**), the other *JALT Journal* editors (**Yuriko Kite**, book reviews, and **Yoshinori J. Watanabe**, Japanese language), the proofreaders, who, when not making me look good, have helped keep me out of trouble (currently, **Mayumi Fujioka**, **Seiji Fukazawa**, **Megumi Kawate-Mierzejewska**, **Aleda Krause**, **Steve McGuire**, **Cynthia Quinn**, **Jack Yohay**), the Editorial Advisory Board, and additional readers who put in countless volunteer hours reading prospective manuscripts. [Like the speakers at the Academy Awards, I am sure I am forgetting to thank someone. If so, I am very sorry!] Finally, a special thanks needs to go to **Malcolm Swanson** and Pukeko Graphics who have been a joy to work with on the *Journal* and other publication projects (and like the proofreaders have saved me from myself more than once).

I introduced **Ian Isemonger** as Associate Editor in the November 2007 issue. Ian will be the new editor and his first issue will be published in November 2008; Ian and I have been working together for the past year and I know the *JALT Journal* will be in good hands! Joining Ian as the new Associate Editor is **Darren Lingely**. Welcome Darren! I am not the only editor leaving the *Journal* at this time; **Yuriko Kite**, the Book Review Editor, is also stepping down—she has done a wonderful job and I have enjoyed working with her and wish her well as she moves on. And **Cynthia Quinn** is stepping down from her proofreading work, and I want to thank her for her work too. Finally, I want to welcome **Laura MacGregor** and **Greg Sholdt** to the Editorial Advisory Board. Both Laura and Greg have been serving as additional readers for the past year and I am glad they have agreed to join the other EAB members, who continually provide an invaluable service to the *Journal*.

Steve Cornwell

Articles

Cognitive Processes in Second Language Word Association

John P. Racine

Ibaraki Christian University

This study utilizes a word association (WA) paradigm to infer similarities and differences between processes used to access the mental lexicons of native speakers (NS) and Japanese nonnative speakers of English (NNS). Three hypotheses were examined: a) grammatical word stimuli will elicit proportionately fewer paradigmatic responses than will content words; b) the proportion of phonologically-related responses will increase when stimuli are presented aurally rather than in written format; and c) NNS responses to infrequent words will not differ from responses to common words if a loan word equivalent exists in their first language (L1). Generally speaking, results concurred with established findings. Where results failed to validate the hypotheses, cognitive models are outlined to account for the data. In particular, a process model involving access to explicit knowledge of grammar rules is presented to account for the fact that NNS were less likely to respond to grammatical word stimuli with syntagmatic responses than were NS ($\chi^2 = 15.22$, $p < .001$, $df = 1$). Also, during aural presentation, only NNS responses, not NS responses, displayed more phonological similarities to their stimuli, suggesting the NNS rely on phonological cues in the absence of semantic knowledge. Similarly, NNS produced fewer semantic associates to low-frequency nouns with loan word equivalents than they did to commonly occurring nouns ($\chi^2 = 3.89$, $p < .05$, $df = 1$). In fact, NNS produced marginally *more* semantic responses to low-frequency nouns without loan word equivalents at all. A model postulating competition between cognitive processes that precipitate semantic responses and those instigated by the salience of phonological similarities between the stimuli and their loan word equivalents is proposed.

語連想による第二言語の認知処理の検証

本研究は言語連想法を使い、英語の母語話者と非母語話者(日本語を母語とする英語学習者)のメンタル・レキシコンへのアクセスの過程の類似点と相違点を検証したものである。単語を被験者に提示し、それに対して返された連想語を文法、意味、音声の観点から分析した。総じて従来行われてきた調査結果とほぼ同様の結果が得られたが、中には異なる結果もあった。これまでとの相違点については、認知モデルに基づいて再解釈した。

A large body of knowledge has accumulated concerning the nature of the mental lexicon, the storehouse of vocabulary in the human mind (for overviews, see Aitchison, 2003; McCarthy, 1990, chap. 3; Singleton, 1999). Research findings in areas as diverse as neuropsychology and linguistics as well as the development of electronic databases have enabled theorists to infer a great deal about the ways in which we store language. Research to date has primarily involved participants' first language (L1) mental lexicons. Recently, however, extensive research and theory have attempted to reveal the nature of learners' second language (L2) lexicon as well. Among the many methodologies available to researchers in this field, and one of the simplest to implement, is word association (WA). The WA paradigm consists simply of the presentation of lexical stimuli to which participants respond with either written or verbal responses. Examination of these stimulus-response pairs allows researchers to make inferences concerning the ways in which lexical items are stored in human memory. This study, too, attempts to utilize word association as a means of inferring similarities and differences between the mental lexicons of native speakers (NS) and Japanese nonnative speakers (NNS) of English, as well as the relationship between the L1 and L2 lexicons within learners themselves.

L1 Lexical Organization

The complexity of the relationship between L1 and L2 lexicons is seen in the diversity of researchers' opinions. Some have argued for the similarities between the two (e.g., Wolter, 2001), while others have highlighted the differences (e.g., Meara, 1983). Still others have focused on the connections between the two (e.g., Channell, 1988; de Groot, 2002; Hall, 1992). To elucidate the nature of this complex relationship, it is necessary to first examine the manner in which L1 vocabulary items are connected

within the mind. The most common links between words in our mental store are semantic and phonological connections.

Although word meaning itself tends to be a “slippery customer” with “fuzzy,” fluid boundaries (Aitchison, 2003, chap. 4), many semantic connections have been revealed through L1 word association studies (Aitchison, 2003, chap. 8; Carter, 1998, chap. 2; McCarthy, 1990, chap. 2). Of all the semantic links to be discovered between words in the L1 mental lexicon, Aitchison (2003) considers coordination, collocation, superordination, and synonymy to be the “most important” (p. 86). McCarthy (1990) adds encyclopaedic knowledge to this list.

For native speakers, coordination, or cohyponymy (Carter, 1998), is the most common link between words, involving “words which cluster at the same level of detail” (Aitchison, 2003, p. 86). This type of connection includes such associates as *salt-pepper*, *butterfly-moth*, and *black-green*, as well as antonymous pairs (e.g., *left-right*, *on-off*). Collocation refers to words which appear together regularly in normal text or speech. These include associates that appear in direct sequential relation such as *butterfly-net*, *salt-water*, or *bright-red*. Superordinates, also called hypernyms (Carter, 1998), are cover words or categorical descriptors often used in defining the more specific associate. Thus, *flower* is the superordinate of *rose* or *tulip*. Synonymy refers to links between words that have the same or similar meanings (e.g., *hungry-famished*, *discover-find*, *begin-start*). Finally, encyclopaedic knowledge refers to the “web-like set of associations” that all human beings develop in their L1 mental lexicons through personal experience, “origins, causes, effects, histories, and contexts” (McCarthy, 1990, p. 41). An example of such an encyclopaedic link from my own mental lexicon is the connection between *sunny* and *slide*. These two words are linked to a childhood memory of when I stood on top of a slide and looked up at the sun. As these kinds of links are clearly idiosyncratic to the individual respondent, classifying this kind of connection can be difficult for the researcher.

Besides the semantic connections outlined above, there is a great deal of evidence revealing phonological links between words in the L1 mental lexicon as well. WA studies involving NS children show that rhyming responses, alliterative responses, or responses with similarly prominent consonant clusters are common for children up to 7 years of age (Meara, 1983). Phonological connections have also been inferred from a number of studies of slips of the tongue or pen. In particular, studies of malapropisms (i.e., errors in speech or in writing in which the intended word and the mistaken word have no semantic similarities) provide clear evidence of phonological links in the mind. For example, Fay and Cutler (1977)

found that the majority of these slips have the same number of syllables (87%) and the same stress pattern (98%) as the words participants had intended to say or write. Examples of word pairs displaying both of these properties include *determination* - *denomination*, *tambourines* - *trampolines*, and *operations* - *occupations* (from Fay and Cutler's 1977 corpus). From the fact that such semantically unrelated slips can occur, researchers have posited that a single, phonologically arranged mental lexicon exists and it is accessed by two different networks, one phonological and one semantic (Channell, 1988; Fay & Cutler, 1977). Indeed, it seems unlikely that both semantically related and unrelated errors could occur without such an arrangement. Further evidence for phonological links between items in the L1 mental lexicon comes from studies in which phonological features of words are remembered despite the apparent absence of links to their meanings (see Aitchison, 2003, for an overview).

Word Association Research

Typical analyses of word association data categorize responses according to combinations of the types of links discussed above. Two types of semantically related response are distinguished: paradigmatic and syntagmatic (e.g., Meara, 1983; Soderman, 1993; Wolter, 2001). Paradigmatic responses belong to the same word class (grammatical paradigm) as the stimulus (Greidanus & Nienhuis, 2001; Meara, 1983). In the case of nouns, then, this would include cohyponyms, superordinates, subordinates, synonyms, antonyms, and so forth. A syntagmatic response, on the other hand, is similar to a collocation in that it forms "an obvious sequential link with the stimulus" (Meara, 1983; also Read, 1993, 2004). Responses that share phonological features (rhymes, assonance, etc.) with the stimulus, but have no apparent semantic connection, are referred to as *clang responses* (Meara, 1983; Soderman, 1993). A fourth category of response type is simply referred to as *errors*. These associates are elicited in response to mistaken words sharing similar phonological or orthographical features to the actual stimuli. For example, *disclose* may elicit responses such as *door* or *far*. In these cases, participants are responding to *close* (as in *to shut*) and *close* (as in *near*) respectively, rather than the actual stimulus word. Finally, a null response category is used when respondents fail to produce a response at all. With these association types clearly defined it is possible to examine word association research results in detail.

Of particular relevance to the current study are the differences between responses of native (NS) and nonnative speakers (NNS) of English.

While NS tend to produce a preponderance of paradigmatic responses, a number of researchers (e.g., Meara, 1983; Piper & Leicester, 1983; Schmitt & Meara, 1997; Soderman, 1993) have found that NNS produce larger proportions of syntagmatic and clang responses than paradigmatic responses. Clang responses are common for young L2 learners, but are gradually replaced by meaning-based responses as L2 proficiency improves (Schmitt & Meara, 1997). A similar "shift" is seen in the fact that syntagmatic responses are gradually replaced by paradigmatic responses as learner proficiency increases (Meara, 1983; Soderman, 1993). Similarly, NNS and NS children also produce more errors than do NS adults (Meara, 1983; Schmitt & Meara, 1997). These findings indicate that mature NS responses are predominantly paradigmatic in nature. Differences in learners' response patterns have been attributed to age, language proficiency, and the relative unfamiliarity of stimuli to learners (Soderman, 1993).

By varying the types of stimuli presented to respondents, researchers have found that WA responses are to some extent dependent upon the word class to which their stimuli belong. Strong intraclass links have been revealed in error analyses of native speakers (Aitchison, 2003) and lexical database research has confirmed that semantic relations and lexical organization differ according to word class (Miller & Fellbaum, 1991). Similarly, L2 WA studies have shown that bilinguals "respond to nouns with nouns and adjectives with adjectives, even across languages, more frequently than they make syntagmatic associations" (Channell, 1988, p. 92). Aitchison (2003) confirms these findings, citing WA studies where nouns elicited nouns 80% of the time, while verbs and adjectives elicited their respective word classes in at least 50% of cases. Piper and Leicester (1983) found that significantly more paradigmatic responses were elicited by verbs ($F = 3.68, p < .05$) and adjectives ($F = 6.259, p < .01$) in NS than in beginning L2 learners. However, this difference was not found in the case of nouns. On the contrary, Soderman (1993), while finding a connection between paradigmatic responses and L2 proficiency overall, failed to find differences between NS and NNS in the number of paradigmatic responses to adjectives. To account for this anomaly, Soderman postulated that each word must pass through different stages of development regardless of the proficiency of the respondent. Thus the timing of the shift toward paradigmatic responses would differ for each word. Soderman's conclusions account for the minor discrepancies in the results cited above, and allow for the conclusion that content words produce a relatively high proportion of paradigmatic responses.

Frequency of stimulus words is also a determining factor in what types of response will be elicited. Postman (1970) found that NS respond to infrequently occurring words with responses typical of NNS (i.e., larger proportions of clang responses). Presumably, a lack of familiarity with the meanings of these words led participants to make connections to phonological characteristics of the stimuli, rather than semantic ones.

Purpose and Hypotheses

The purpose of this study was to examine areas not fully covered in the WA research literature to date. As cited above, studies examining word class have found that content words typically elicit paradigmatic responses (i.e., content words). The use of grammar-function words as stimuli, however, has not been actively pursued in WA research thus far. One of the purposes of this study is to address this gap. There are two reasons to assume that such function words will produce proportionately fewer paradigmatic responses than content words. First, given the relatively small number and fixed nature of membership in this class, there are simply fewer intra-class words from which to choose a response. Second, functional words hold no inherent meaning of their own. They only acquire meaning in relation to the words around them as they appear in discourse. Thus, to make sense of these terms, they must be placed in the context of other words (i.e., in syntagmatic relations). While *the* seems incomplete on its own, it begins to make sense in constructs like *the gang*, *the cake*, or *the city*. Thus, one of the purposes of this study (Hypothesis 1 below) was to determine if, in fact, function words elicit proportionately more syntagmatic responses than do content words.

Another area of research yet to be deeply explored concerns the modes by which WA stimuli are presented to participants and by which responses are expressed. The overwhelming majority of word association research to date has been conducted in written-written (i.e., written presentation/written response) mode. Generally speaking, L1 responses elicited in this manner tend to be semantically related to their stimuli. That is, there is a preponderance of paradigmatic and syntagmatic responses. Relatively few clang responses are elicited. Although Kudo and Thagard (1999) conducted their research in aural-oral (aural presentation/oral response) mode, they did not utilize a written-written comparison group, thus failing to draw conclusions concerning differences in response types as a function of mode of presentation. This study, on the contrary, attempts to compare responses elicited by written-written vs. aural-oral modes

of presentation. It is expected (Hypothesis 2) that as phonological characteristics of the stimuli are made more salient (via aural presentation), responses will become more phonologically rather than semantically related to their stimuli. That is, the relative proportion of clang responses will increase in relation to paradigmatic and syntagmatic responses.

A third area of interest in this study concerns the unique place that loan words hold in the L2 mental lexicons of Japanese learners of English. As cited above, relatively infrequent words tended to produce proportionately more clang responses than do common words as far as native-speaking respondents are concerned (Postman, 1970). This NNS-like response pattern was attributed to a lack of familiarity with the meanings of the cues. In the case of NNS whose L1 is Japanese, however, research into the effects of word frequency on associations may be confounded by the presence of a great many English loan words in the Japanese language. Words such as *helicopter* and *asbestos*, which are quite unfamiliar to many L2 learners, already exist as *herikoputa* and *asubesuto* in Japanese. Thus, strong semantic connections to these terms should already exist in the L2 lexicon of native Japanese. These terms should elicit approximately the same number of semantic responses as those elicited by such common everyday words as, for example, *tree* and *car* (Hypothesis 3).

In accordance with these purposes, and in light of the research findings cited above, this study was designed to test the following hypotheses:

1. Function word stimuli will elicit proportionately fewer paradigmatic responses than will content words.
2. The proportion of phonologically related responses will increase when stimuli are presented aurally rather than in written format.
3. NNS responses to infrequent words will not differ from responses to common words if a loan word equivalent exists in the L1.

Method

Participants

Forty-four participants took part in this study: 11 native English speakers (NS; mean age, 31.1), 11 adult Japanese learners of English as a second language (NNS-Adult; mean age, 49.3), and 22 Japanese university students (NNS-University; mean age, 19.4). Although no objective

test of vocabulary or language ability was administered, a subjective evaluation of the NNS groups' language abilities was made. Participants were also asked to report the period of time during which they had studied English. The NNS-Adult group were evaluated as intermediate to high-intermediate; mean length of English study was 13.5 years. NNS-University respondents were judged to be high beginners; mean length of English study was 7.5 years.

Lexical Items and Procedure

The basic design of this study is based upon a task suggested by McCarthy (1990, p. 152), in which a number of words from differing word classes are used as cues in a word association study. McCarthy suggested that responses be evaluated in light of previous findings from WA research, and to make inferences concerning the development of the L2 mental lexicon. In fact, words were presented via two modes of presentation: an interview to elicit verbal responses, and a printed form on which written responses were recorded. (An example form appears in the Appendix.) There were 16 lexical items in each presentation. Lexical items were selected from categories suggested by McCarthy (1990): grammatical/function words, everyday items from the physical environment, low frequency items, and other word classes. The actual number of items employed was increased to improve reliability and generalizability of results. Grammatical words included articles, prepositions, conjunctions, and pronouns. To test Hypothesis 3, low frequency items included two types of nouns: four items for which the Japanese equivalent term was phonetically unrelated: *hospital* (*byouin*), *morning* (*asa*), *rabbit* (*usagi*), and *November* (*juichigatsu*) and four items whose equivalent terms were loan words borrowed from English: *helicopter* (*herikoputa*), *asbestos* (*asubesuto*), *orchestra* (*okesutora*), and *escalator* (*esukareta*). The stimuli appear in Table 1.

Written and verbal cues were presented to each participant in one of eight randomly selected orders respectively, making 64 possible presentation orders for each respondent, substantially eliminating the influences of priming and order effects. For the same reasons, respondents were randomly placed in either "aural-first" or "written-first" order of presentation. Written instructions informed them to answer with the first English word that came to mind. They were told that they need not respond to any items they did not understand or for which no response readily came to mind. They were also informed not to be concerned about correct spelling to ensure that the first word they thought of (rather than a word that was easier to spell) was entered. Instructions appeared in

Table 1. Word Association Stimuli

	Written items	Verbal items
Grammatical / Function words	the of and he or she	a at but this or that
Items from the everyday physical environment	table car tree pen	store desk house spoon
Relatively low- frequency words	helicopter asbestos hospital morning	orchestra escalator rabbit November
Verbs	eat walk	sleep jump
Adjectives	happy soft	bright heavy

both Japanese and English for ease of understanding, and all participants received identical instructions to increase reliability across respondents. The aural presentation interview consisted of the researcher’s reading the list of cues and waiting for a verbal response to each. If the respondent failed to respond after approximately 5 seconds, the item was repeated. If the participant still failed to respond, a null response was recorded and the next item was presented.

Results and Discussion

Two separate analyses of the data were conducted. The first, to examine Hypotheses 1 and 3, involves the categorization of syntagmatic, paradigmatic, and clang responses as defined above. These categories incorporate the response types suggested by Aitchison (2003) and McCarthy (1990): Coordination, superordination, and synonymy are types

of paradigmatic responses. Collocational responses are syntagmatic. Responses based on sound similarities in the absence of clear semantic links were coded as clang responses. A fourth category, null responses, includes errors, cases where the participant did not respond at all, and the kinds of responses McCarthy (1990) refers to as encyclopaedic. Encyclopaedic links are presumed to be semantic links based on the subjective experience of the respondents, but without a measure to assess this connection they cannot be attributed to any other category. Responses to the stimulus word *November* were not included in these analyses as it became clear that Japanese learners usually learn the months of the year as an ordered list. Thus it was impossible to classify common responses such as *December* as either paradigmatic or syntagmatic in nature.

To test Hypothesis 2, a second analysis was conducted in which responses were coded according to phonological characteristics shared with their respective stimuli. Examined phonological features included number of syllables and whether either the first or last phoneme matched those of the stimulus. These were selected as they are typical phonological features to be examined in lexical research (see Aitchison, 2003). Another common measure of phonological similarity, the examination of stress patterns, was not conducted as the majority of stimulus words were one syllable in length.

Analysis 1: Response Type

Figure 1 shows the percentages of response types per participant group for both written and verbal responses. It is clear that paradigmatic responses are predominant. Generally speaking, response patterns appear to be similar for all groups (paradigmatic > syntagmatic > clang). It should be noted, however, that NS produced syntagmatic responses (e.g., *tree-green*, *walk-home*) significantly more often ($\chi^2 = 21.83$, $p < .001$, $df = 1$) than the L2 learners, while making slightly fewer paradigmatic responses as well. Very few clang responses were produced by any of the groups while null responses were much more prominent for L2 learners than for NS. These results are in line with previous findings, suggesting that semantic links in the L2 lexicon are somewhat tenuous, leading NNS respondents to mistakenly respond to phonologically similar but misperceived stimuli. Similarly, a combination of paradigmatic and syntagmatic scores shows that NS are more likely ($\chi^2 = 19.74$, $p < .001$, $df = 1$) to respond with semantic associates than are the NNS groups. A comparison of Figures 2 and 3 allows for a more detailed examination of this data.

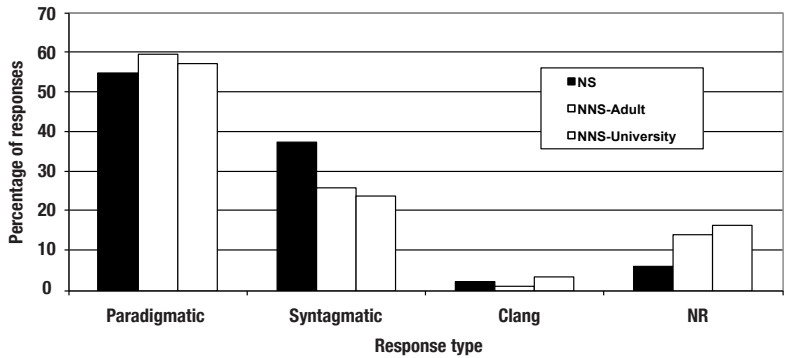


Figure 1. Percentage of response types per group

Figure 2 shows that responses to content words do not deviate substantially from the overall response pattern illustrated in Figure 1. Participants are more likely to respond to content words paradigmatically (e.g., *table-chair*) whether they are native speakers or not. Conversely, Figure 3 illustrates how native speakers are more likely to respond to grammatical-function words with syntagmatic responses (e.g., *and-you*) rather than

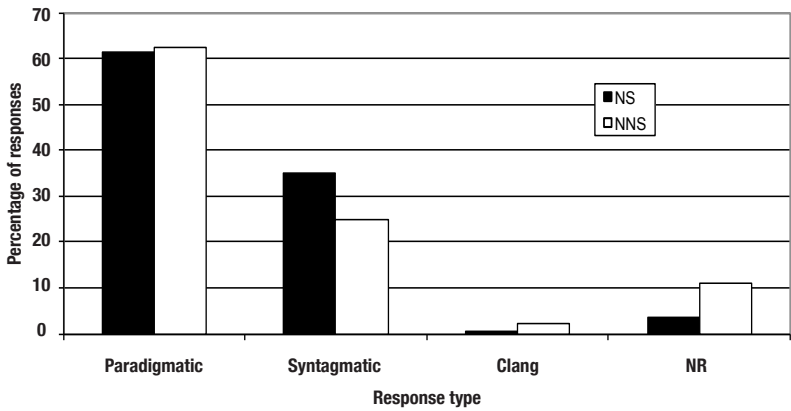


Figure 2. Responses to content word stimuli

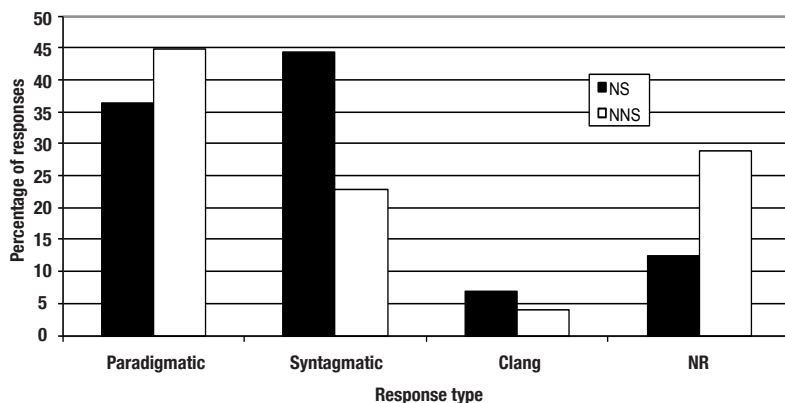


Figure 3. Responses to grammatical word stimuli

paradigmatic ones (e.g., *and-but*). The proportionately fewer paradigmatic responses seen in Figure 3 lend support to Hypothesis 1 which predicted higher percentages of paradigmatic responses in the case of function-word stimuli. Chi-square tests demonstrated statistically significant differences in frequency of syntagmatic responses as a function of group membership ($\chi^2 = 15.22, p < .001, df = 1$) as well as marginal differences in frequencies of paradigmatic responses ($\chi^2 = 1.88, p < .25, df = 1$).

From a linguistic perspective, the fact that NS respond to grammatical-function items with proportionately more syntagmatic than paradigmatic responses can be interpreted as a function of the relatively small and fixed membership of the grammatical word class. In other words, there are simply fewer function words from which to select a (paradigmatic) response. Thus responses are retrieved from other word classes. This interpretation, however, does not account for the high levels of paradigmatic responses in the case of NNS. A cognitive interpretation may better account for these findings: As grammatical-function stimuli are relatively meaningless in isolation, it would appear that native speakers impose meaning on them by generating contexts in which these words occur. That is, they produce the necessary collocations within which function words acquire their meaning and thus respond with collocational associates, (i.e., syntagmatic responses). This process is illustrated in the upper half of Figure 4 where the function word *and* elicits the syntagmatic response *pepper* as a result of the respondent's having generated the context phrase *salt and pepper*. Thus, responding syntagmatically to grammatical

word stimuli is contingent upon the respondent’s working knowledge of the stimuli’s occurrence in text.

Certainly NNS have less experience with authentic English text than native speakers. With the exception of very advanced learners, then, they would have less knowledge of the contexts in which function words occur. Thus, with limited knowledge of collocations to draw upon, NNS must rely instead on some other mechanism by which to impose meaning on these stimuli. In this case, learners may consult explicit knowledge of grammatical rules to make sense of function words. In so doing they may access lists of other words adhering to these rules (cohyponyms) and respond accordingly, with paradigmatic responses (e.g., *and-but*). This process is illustrated

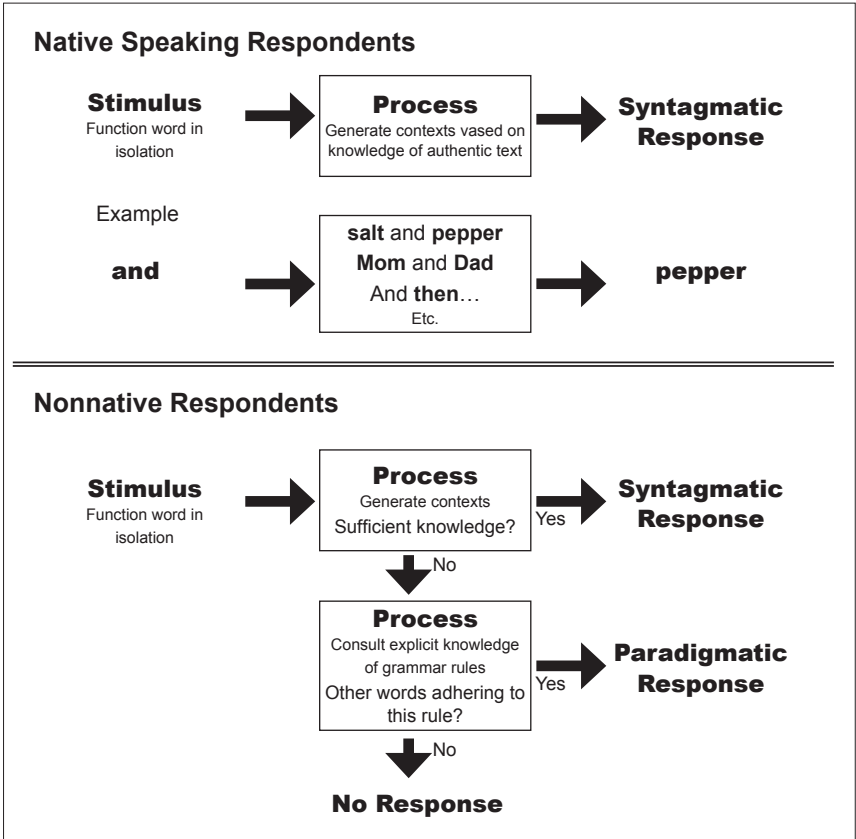


Figure 4. A cognitive model for word associations involving functional-word stimuli

in the lower half of Figure 4 where the NNS respondent with sufficient knowledge of the contexts in which the stimuli might appear can respond with a syntagmatic response. One can imagine the grammatical word *the* eliciting a train of thought like "I always see *the* in '*The end*.'" Hence, the stimulus would elicit the collocation *end*, a syntagmatic response. Where collocational knowledge is insufficient or where syntagmatic connections in the L2 mental lexicon are weak, a secondary process is initiated. Here, the NNS respondent consults explicit knowledge of grammar rules and thus encounters other words adhering to these rules. Again, in the case of the stimulus *the*, for example, the respondent's thinking may resemble "*The* goes in front of nouns like *a* does." Thus *the* would elicit the paradigmatic response *a*. The difficulty in successfully completing both of these processes is reflected in the inordinately large number of null responses for NNS (28.8%) as illustrated in Figure 3. In these cases, nonnative respondents may simply be giving up en route to discovering possible cohyponymous responses to grammatical stimuli as their cognitive resources become taxed. Undoubtedly, numerous other cognitive and motivational factors affect this process as well, though it is beyond the scope of this paper to further elaborate on them here.

Word frequency. In the case of native speakers, proportions of response types did not significantly differ as a function of word frequency. This finding was expected, as the low frequency nouns (*helicopter*, *asbestos*, *hospital*, *morning*, *orchestra*, *escalator*, and *rabbit*) had been selected for their relative unfamiliarity to the NNS groups. It was assumed that all of these terms are very familiar to native speakers of English. NNS groups, on the contrary, showed marginal, but statistically insignificant differences ($\chi^2 = 1.62$, $p < .25$, $df = 1$) in the proportion of semantic responses as a function of stimulus frequency. That is, infrequent stimuli elicited slightly fewer semantic responses from NNS. Presumably, the lack of familiarity with the meanings of these words led respondents to respond with phonologically related associates or no response at all. Utilizing even less frequently occurring stimuli could potentially produce statistically significant effects.

To further examine the effects of word frequency, low-frequency stimuli were split into two categories: those for which an equivalent loan word exists in Japanese (*helicopter*, *asbestos*, *orchestra*, and *escalator*) and those items perceived as unique to English (*hospital*, *morning*, and *rabbit*). A comparison of NNS responses to these two types of stimuli as well as to the high-frequency nouns allows a test of Hypothesis 3 which stated that NNS response patterns to infrequent stimuli for which a loan word

equivalent existed in Japanese would not differ from response patterns elicited by commonly occurring stimuli. In fact, a chi-square test revealed statistically significant differences ($\chi^2 = 3.89$, $p < .05$, $df = 1$) between NNS responses to high-frequency stimuli and to low-frequency stimuli despite the presence of loan word equivalents in Japanese to the low-frequency nouns. Specifically, NNS produced significantly *fewer* semantic associates to low-frequency nouns with loan word equivalents than they did to commonly occurring nouns. Further, a comparison of responses to the two types of low-frequency nouns shows that NNS produce *more* paradigmatic and syntagmatic responses to nouns *without* a loan word equivalent (e.g., *hospital-sick*). Albeit only a statistically insignificant difference ($\chi^2 = 2.56$, $p < .25$, $df = 1$), this trend opposes that predicted by Hypothesis 3: NNS do not in fact respond to low-frequency nouns with loan word equivalents as they do to high-frequency nouns. In fact, NNS respond more often to low-frequency nouns with semantic responses than they do to high-frequency stimuli.

These results can perhaps best be accounted for by inferring cognitive interference between the dominant processes that usually result in semantic responses and an alternative process instigated by the salience of phonological similarities between the English stimulus and its loan word equivalent. This model is illustrated in Figure 5. When encountering a noun with a loan word equivalent (e.g., *asbestos*), a respondent aware of the stimulus' similarity to the loan word (*asubesuto*) initiates a phoneme-by-phoneme phonological check to confirm this similarity ("Is asbestos *really* asubesuto?"). This is shown as Process 1 in Figure 5. Only after enough similarity has been recognized ("This word *must* be the same thing as asubesuto.") will the next process be initiated. If comprehension of the loan word is confirmed in the second process ("I know what this means"), then strong semantic ties to the word prompt either a syntag-

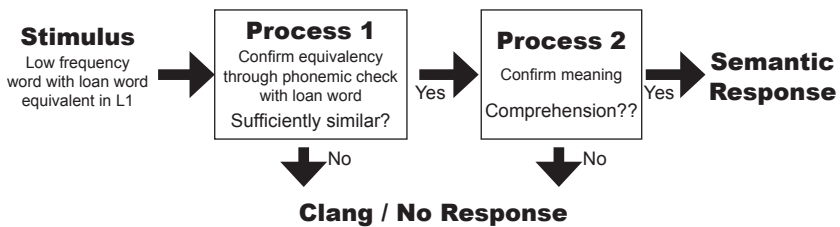


Figure 5. A cognitive model for NNS word association responses to low frequency stimuli with L1 loan word equivalents

matic or paradigmatic response (e.g., *dangerous* or *insulation*). If loan word similarity is not recognized in Process 1, or if the respondent gives up en route to this decision (because, say, motivation is insufficient or cognitive resources become taxed), then clang or null responses result. In the case of stimuli that do not alert respondents to the possibility of a loan word equivalent, only a process similar to Process 2 would be initiated. This second process, that of confirming the meaning of a stimulus and responding with a semantic associate, is the fundamental mechanism in all word association.

Analysis 2: Phonological features

The analysis of phonological features was based in part on memory research findings that first and last sounds of words are remembered better than those in the middle positions (see Aitchison, 2003). Comparisons were made between stimulus-response pairs in regards to numbers of syllables, and whether the first or last phonemes were identical. All valid (i.e., non-null) responses were examined. Results showed that shared phonological characteristics between the stimuli and the responses of NS were uniformly less frequent than those of NNS. These results are illustrated in Figure 6. For each measure, NS displayed noticeably lower percentages of phonologically similar responses than did the combined NNS groups. Chi-square tests were used to test the significance of these differences. Results of these tests appear in Table 2 where each cell represents the test score for differences between that particular NNS group’s scores and those of the NS group. It is clear that more experienced learn-

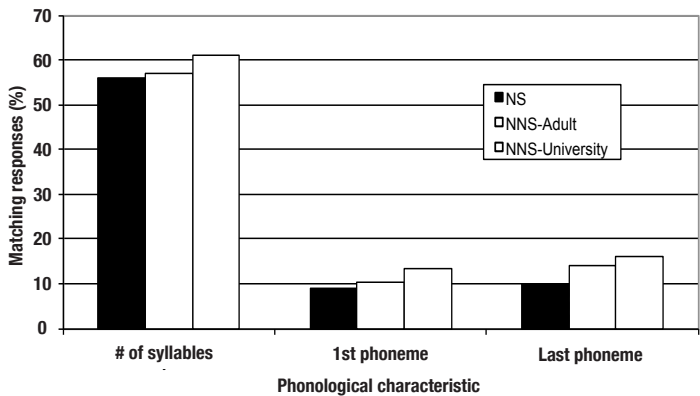


Figure 6. Phonological similarities between associates

ers (NNS-A) were less likely to respond to phonological cues than were less experienced learners (NNS-U). The first and last phonemes of NNS-U responses matched the stimuli's first and last phonemes significantly more often than did those of the NS responses. It would appear then that NNS are at least sometimes utilizing phonological characteristics of the stimuli as cues in generating responses. NS and experienced learners with stronger semantic connections to the words find it less necessary to rely on these kinds of cues. These three measures appear to support researchers' intuitions about learners' reliance upon phonological characteristics in the absence of strong semantic ties to the stimuli.

Table 2. Chi-square Test Results Comparing Differences in Phonologically Similar Response Patterns to NS Responses

	Total NNS	NNS-Adult	NNS-University
Number of syllables	1.38	0.10	2.23
First phoneme	2.78	0.45	3.90*
Final phoneme	5.67**	2.29	6.37**

Note. *Df* = 1 in all cells.

p* < .05. *p* < .025.

The same phonological criteria described above were used to test Hypothesis 2 which stated that the frequency of phonologically related associates would increase when stimuli are presented aurally. In order to determine this effect of mode of presentation on shared phonological characteristics, only responses from first presentations were examined. That is, only responses from the aural-first condition were used in calculating the effects of aural presentation. Aural responses from participants in the written-first condition were not included as they were considered likely to have been influenced by response processes involved during the written presentation. Likewise, the effects of written presentation were measured in the same way. The results of this analysis only partially validated the mode-of-presentation hypothesis: NNS produced responses with the same number of syllables as their stimuli significantly more often when the stimuli were presented aurally ($\chi^2 = 6.46, p < .025, df = 1$). Likewise, NNS responded with the same first phoneme significantly more often during aural presentations ($\chi^2 = 18.87, p < .001, df = 1$).

This effect was not observed with final phonemes. Interestingly, mode of presentation had no effect on NS responses as measured by any of these three phonological measures. These results point to the primacy of semantic associations in the responses of native speakers. Regardless of the salience of phonological cues in the stimuli, NS rely primarily on semantic connections in generating word associations. Where semantic ties are more tenuous, as in the case of NNS, salient phonological cues prompt phonologically related associations. This finding concerning L2 learners parallels the results of studies cited above involving L1 learners, that is, NS children (Meara, 1983; Schmitt & Meara, 1997). As learners, NS make more clang responses than they do as adults in the absence of strong semantic associations. The findings described here may indicate that the L2 lexicon also develops from being comprised of predominantly phonological connections to semantic ones.

Summary and Conclusion

This study found at least partial support for its three hypotheses. It was predicted that function word stimuli would elicit fewer paradigmatic responses than would content word stimuli. In fact, only native speakers responded in this manner. NNS produced significantly higher frequencies of paradigmatic responses to grammatical word stimuli than did NS. A cognitive process model was presented to account for these findings, suggesting that NS rely on collocational knowledge to generate syntagmatic responses while NNS rely on knowledge of explicit grammar rules to generate paradigmatic responses. It was also predicted that aural presentation of stimuli would precipitate an increase in phonologically related responses. Here, only NNS responses fit the predicted pattern. Results here were discussed in terms of the strength of semantic connections in the mental lexicons of native speakers and NNS' reliance on phonological cues in the absence of such strong semantic links. Finally, it was predicted that NNS responses to infrequently occurring stimuli for which an L1 loan word equivalent exists would not differ from responses to common stimuli. In fact, NNS responded with fewer semantic responses despite the existence of loan word equivalents. Here too, cognitive processes were inferred to account for the data. In this case, it was suggested that NNS initiate a cognitively taxing phoneme-by-phoneme check when a stimulus is recognized as a potential loan word. Only after this is complete can the usual process of semantic recognition occur.

This study was conducted in an attempt to address certain gaps in word association research to date. It is hoped that the focus on grammatical word stimuli, mode of stimuli presentation, and loan words contributes some interesting findings to the body of WA research knowledge and points to some clear differences between the manner in which L1 and L2 items are stored and accessed in the mental lexicon. Likewise, one hopes that the application of process models to WA data will impart a fresh focus on theorization concerning the mental lexicon and how it is accessed. Without further research, however, the models presented here remain somewhat speculative. In particular, further studies should be designed to uncover which specific word classes account for the effects attributed here to differences between function versus content word responses. Follow-up studies must also replicate these findings with larger respondent samples to ensure reliability of these results.

John Racine is a lecturer in the Department of Contemporary English at Ibaraki Christian University in Hitachi, and in the General English Department at Ibaraki University in Mito. His research interests include psycholinguistic approaches to the study of the second language mental lexicon.

Author Note

An alternate version of this paper was submitted to the University of Birmingham in partial fulfillment of the requirements for the degree of MA TEFL/ TESL. Portions of the results were presented at the general meeting of the Ibaraki Chapter of JALT, December 10, 2006, and at the Shinshu ELT Research Colloquium, March 4, 2007. Correspondence concerning this article should be addressed to John Racine, 2611-62 Higashi Ishikawa, Hitachinaka City, Ibaraki, 312-0052. E-mail: gaijira@gmail.com

References

- Aitchison, J. (2003). *Words in the mind* (3rd ed.). Oxford: Blackwell.
- Carter, R. (1998). *Vocabulary: Applied linguistic perspectives* (2nd ed.). London: Routledge.
- Channell, J. (1988). Psycholinguistic considerations in the study of L2 vocabulary acquisition. In R. Carter & M. McCarthy (Eds.), *Vocabulary and language teaching* (pp. 83-96). London: Longman.

- de Groot, A. M. B. (2002). Lexical representation and lexical processing in the L2 user. In V. J. Cook (Ed.), *Second language acquisition I: Portraits of the L2 user* (pp. 32-63). Clevedon: Multilingual Matters.
- Fay, D., & Cutler, A. (1977). Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry*, 8, 505-520.
- Greidanus, T., & Nienhuis, L. (2001). Testing the quality of word knowledge in a second language by means of word associations: Types of distractors and types of associations. *The Modern Language Journal*, 85, 567-577.
- Hall, C. J. (1992). *Making the right connections: Vocabulary learning and the mental lexicon*. Puebla, Mexico: Universidad de las Americas. (ERIC Document Reproduction Service No. ED363128)
- Kudo, Y., & Thagard, D. (1999). Word associations in L2 vocabulary. *University of Hawai'i Working Papers in ESL*, 17 (2), 75-105.
- McCarthy, M. (1990). *Vocabulary*. Oxford: Oxford University Press.
- Meara, P. (1983). Word associations in a foreign language [Electronic version]. *Nottingham Linguistics Circular*, 11 (2), 29-38.
- Miller, G. A., & Fellbaum, C. (1991). Semantic networks of English. *Cognition*, 41, 197-229.
- Piper, T. H., & Leicester, P. F. (1983). *Word association behaviour as an indicator of English language proficiency*. (ERIC Document Reproduction Service No. ED227651)
- Postman, L. (1970). The California norms: Association as a function of word frequency. In L. Postman & G. Keppel (Eds.), *Norms of word association* (pp. 241-320). New York: Academic Press.
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10, 355-371.
- Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 209-227). Philadelphia: John Benjamins.
- Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework. *Studies in Second Language Acquisition*, 20, 17-36.
- Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.
- Soderman, T. (1993). Word associations of foreign language learners and native speakers – different response types and their relevance to lexical development. In B. Hammarberg (Ed.), *Problems, process, and product in language learning* (pp. 157-169). Stockholm: AFInLA.
- Wolter, B. (2001). Comparing the L1 and L2 mental lexicon: A depth of individual word knowledge model. *Studies in Second Language Acquisition*, 23, 41-69.

Appendix

<p>Thank you for agreeing to participate in this study.</p> <p>The Japanese instructions on the right are the same as these written in English. So you may ignore them if you feel more comfortable completing this survey in English.</p> <p>Instructions: Read the following list of words and write down the first English word that comes to mind. There is no right or wrong answer, so you don't have to think about it too much. Don't worry about spelling either; just try to write down the first word that comes to mind. If you don't understand a word, you can leave it blank and continue to the next word.</p>	<p>この研究への参加に同意いただきまして、ありがとうございます。</p> <p>この日本語文は、左側の英語文と同じ文章です。日本語のほうが理解しやすい方は、左側の英語は無視してください。</p> <p>説明：左側にある単語を見て、一番に思い浮かんだ英単語を書いてください。正しい答えも誤った答えもあります。あまり考える必要はありませんので、最初に思い浮かんだ単語を書いてください。スペルも心配しなくて結構です。思い浮かんだままに書いてください。単語が分からなかった場合は、何も書かなくて結構ですので、次の単語に移ってください。</p>
---	--

soft	
asbestos	
car	
helicopter	
tree	
hospital	
she	
eat	

and	
walk	
pen	
the	
morning	
table	
of	
happy	

Japanese Secondary School Teachers' Beliefs and Practices Regarding Communicative Language Teaching: An Exploratory Survey

Takako Nishino
Temple University Japan

Since 1989, the Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT) has attempted to promote higher achievement in English communicative skills among secondary school students by urging teachers to use communicative activities. MEXT has also undertaken to achieve this goal by executing a 5-year Action Plan. This exploratory study investigates Japanese teachers' beliefs and practices regarding communicative language teaching (CLT) in their classrooms through a survey of 21 secondary school teachers. The results show that in order to employ CLT in their classrooms, teachers feel that a change in classroom conditions is a prerequisite. The results also show that CLT is beginning to be employed at the local level. In order to delineate ways to help this small local change lead to real English education reform in Japan, a comprehensive investigation of the beliefs of a larger number of language teachers is necessary.

1989年より、文部科学省(当時文部省)は、中学生・高校生の英語によるコミュニケーション能力を高めようと、外国語科授業でのコミュニケーション活動の採用を促してきた。さらに同省は、2003年から5年間の「英語が使える日本人」の育成のための行動計画の実施により、その方針の具体的な実現をねらっている。本探索的研究では、中学及び高等学校の英語教員21人にアンケートを実施し、コミュニケーション・アプローチ(CLT)を授業に採用することについてどのような信条を持ち、どのように実践しているかについて調査した。その結果、被験者の多くはCLTを採用するために教室の教育環境を変えてほしいと願っていることが判明した。また、CLTは一部の学校で利用され始めていることがわかった。この傾向をさらに確かなものにするためにも、より多くの英語教員を対象にした包括的調査が必要である。

Kagan (1992) claimed that most of a teacher's professional knowledge can be regarded as beliefs. But what shapes teacher beliefs and practices? Researchers have been focusing on this question since they started regarding teachers as active decision-makers in the 1980s (Freeman, 2002). Borg (2003) reviewed 64 studies and reported that teacher cognition, defined (p. 81) as "the unobservable cognitive dimension of teaching—what teachers know, believe, and think," plays a central role in teachers' lives and that contextual factors influence both teacher cognition and practice. Other research suggests that teachers' beliefs might have the strongest influence on classroom practices (Nespor, 1987; Pajares, 1992; Richardson, 1994). In this paper teacher knowledge will be used interchangeably with teacher beliefs.

In the field of ESL/EFL, this question led me to a narrower inquiry. What shapes teachers' beliefs concerning novel teaching methods: government policy, high-stakes examinations, previous teaching and learning experience, or contextual factors? In recent years, this question has been investigated in relationship to Japanese English teachers' perceptions and practices in communicative language teaching (CLT) (e.g., Gorsuch, 2000, 2001; Sakui, 2004).

Since 1989, the Ministry of Education, Culture, Sports, Science, and Technology (MEXT).¹ has attempted to promote higher achievement in English communication skills among secondary school students by urging teachers to incorporate CLT into their lessons. The 1989 version of *The Course of Study* (implemented in 1993) introduced new oral communication courses,² and the 1999 version (implemented in 2003) called for the development of "practical communication abilities" as a main goal of foreign language education (MEXT, 1989, 1999). In 2003, MEXT produced an Action Plan with the goal of cultivating English communication abilities in Japanese people (MEXT, 2003).

This study, conducted after the start of the Action Plan, aims to continue the line of research begun by Gorsuch (2000). A survey of Japanese secondary school teachers was conducted to ascertain their beliefs regarding CLT. Results from this pilot questionnaire appear to support Gorsuch's argument that school and classroom conditions have an impact on teacher perceptions concerning CLT.

In the following sections, I will briefly discuss CLT in EFL contexts and review the history of Japanese English education and MEXT policy. Then, I will outline the research on Japanese teacher beliefs and practices regarding CLT.

CLT in EFL Situations

Communicative language teaching is defined as an approach to foreign or second language teaching which aims to develop communicative competence (Richards, Platt, & Platt, 1992). CLT was designed to be applied to ESL situations, especially in Britain, North America, and Australasia, where English teachers support a skill-based, discovery-oriented, collaborative approach to education (Holliday, 1994) and where classroom language learning usually takes place in small classes through group and pair work. In contrast, in Japan, classroom conditions differ from those in ESL contexts. Features common to the Japanese context include large classes, a tradition of nonnative teacher-centered lessons, limited communicative needs among students, and minimal foreign language input outside the classroom. Under such conditions, it may be ineffective to try to use CLT in the same way as it is employed in ESL situations.

Given the potential incompatibility of CLT to the Japanese context, it may be helpful to review the historical background of English education and recent policy innovations in Japan.

Historical Background of English Education in Japan

MEXT Policy

After the Second World War, a new education system started in Japan; *The Course of Study* was first published in 1947. Since then, it has been revised seven times. In 1955, the third version of *The Course of Study* started to have legally binding force in prescribing the content of textbooks. Secondary school textbooks have been strictly checked and authorized by MEXT since then. As teachers must use authorized textbooks, syllabus design is constrained by *The Course of Study* (Imura, 2003).

In the 1980s, aiming at internationalization, the Ministry initiated English education reform. One of the policies adopted in 1987 was the Japan Exchange and Teaching (JET) program which invited young college graduates from overseas to participate in foreign language education throughout Japan as assistant language teachers (ALTs). They were called AETs—Assistant English teachers—when the JET program started. MEXT then revised *The Course of Study* again in 1989 and 1999, and implemented the 5-year Action Plan in 2003. This latest Action Plan advocated training 60,000 Japanese English teachers intensively and hiring 11,500 additional ALTs.

In spite of these initiatives taken by MEXT, various factors such as the use of the traditional grammar-translation method, *yakudoku*, and high-stakes entrance examinations have hindered the reform of Japanese English education.

Yakudoku and Examination English

English education in Japan has been dominated by the *yakudoku* method (Gorsuch, 1998; Suzuki, 1999). The main classroom activity in this method is word-by-word translation of written English into Japanese. The teacher gives grammatical explanations in Japanese; students have few chances to vocalize English except when they practice reading by repeating after the teacher.

Yakudoku has been a favored teaching method used to help students pass university entrance examinations which have mainly evaluated reading skills and grammatical knowledge. Many high school teachers believe they cannot ignore university entrance examinations and thus teach using this method (Gorsuch, 1998). Meanwhile, critics of this approach claim that *Juken Eigo* (examination English) requires high school students to learn decontextualized language and peripheral grammar (Law, 1994).

Teacher Beliefs about the Use of CLT

Gorsuch's Model

Research on classroom practices prior to 2003, the year when the most recent revision of *The Course of Study* became mandatory, appears to suggest that CLT was not being widely used. Brown (1995) claimed that very little oral English was used during English lessons. Gorsuch (1998) reported that 70 to 80% of the Japanese high school teachers she surveyed used *yakudoku* in their English classes. More recently, Taguchi (2005) found that high school teachers' concern about entrance examinations tended to lead them toward traditional methods such as choral repetition in oral communication courses.

In order to explore factors that influenced teachers' approval of CLT, Gorsuch (2000) employed a structural equation model based on empirical data from Japanese high school teachers, and examined the influence of school and classroom conditions on their approval of CLT activities in English I and English II. She identified four latent variables inferred from items on a questionnaire. These were a) a *school latent variable* de-

rived from questions on the local syllabus, teacher preservice license programs, colleagues, and principals; b) a *classroom latent variable* relating to class size, student expectations, student English abilities, and teacher's English-speaking ability; c) an *exam latent variable* reflected in questions on *The Course of Study*, university entrance exams, and parental expectations; and d) a *CLT approval latent variable* based on attitudes towards communicative activities.

Gorsuch's results showed that there were strong to moderate relationships between the exam latent variable and the school and classroom latent variables. The school and classroom variables had positive, though moderate to weak, effects on teachers' approval of CLT. However, there was a weak negative relationship between the exam and CLT approval variables. Thus, these findings support the long-standing view of the strong effects of university entrance examinations on secondary school education in Japan. However, they do not show that teachers' attitudes toward the examinations directly influence their approval of CLT activities.

In her analysis, Gorsuch suggested that if university entrance examinations were to include questions that tested students' communicative ability, teachers would think that individual school conditions (e.g., school curriculum) could change (for a discussion of the effect entrance examinations have on school curriculum see also Browne & Wada, 1998). Changes in school conditions might in turn moderately influence the approval of CLT among teachers who work at those institutions. Gorsuch suggested that Japanese teachers might resist CLT in the classroom because they believe in strong teacher control and memorization/translation in foreign language learning. She argued that without reforming school and classroom conditions, CLT may not become widely employed in Japan even if university entrance examinations change.

Gorsuch also wrote that "the influence of students' expectations concerning teachers' instruction is potentially powerful" (2000, p. 685). She argued that high school students expect to prepare for entrance exams in English courses and that they might not see the value of communicative instruction. It might be difficult for teachers to effectively use CLT without student cooperation.

Thus, contextual factors such as the *yakudoku* method, university entrance examinations, and learners' beliefs have all had negative effects on the use of CLT activities by English teachers in Japanese classrooms. However, a question remains as to whether teacher classroom practices have been changing given MEXT's active support of CLT methods.

Teacher Beliefs and Practices

Recent research findings suggest growing approval of CLT use. Gorsuch (2001) reported that Japanese high school teachers mildly approved of CLT activities, although there were still obstacles to implementing CLT activities in their classrooms. Taguchi (2002) found that even though high school teachers were still using exam-oriented teaching methodologies, they want to teach communicative skills. Both studies suggest that while there is inconsistency between the teachers' beliefs and practices, teachers' perceptions regarding the use of CLT may be gradually changing.

There is a need to delve into what might be influencing teachers' beliefs and practices. Fang (1996) claimed that contextual factors including classroom conditions can have a powerful impact on teachers and affect their classroom practice. Gorsuch's model (2000) also showed that school and classroom conditions influence teachers' perceptions regarding the use of CLT. Listening to teachers' voices can help us better understand the relationships between these contextual factors and teachers' perceptions. This study investigates Japanese secondary school teachers' beliefs regarding the use of CLT in their classrooms and what conditions, if any, they want to change in order to better facilitate CLT activities. The following research questions were posited:

1. What are Japanese secondary school teachers' beliefs and practices regarding CLT?
2. What contextual factors, if any, do Japanese secondary school teachers think should be changed in order to utilize CLT in Japanese secondary schools?

Method

Materials and Procedures

The questionnaire (see Appendix) used in this study was based on information gathered through interviews with three middle school teachers, and notes on teaching and learning gained from the researcher's recent experience teaching in high school. Questions, written in Japanese, were categorized as related to either the first or the second research question, then assigned as 15 closed-response and 3 open-response questions following the questionnaire format recommended by Brown (2001). A 6-point Likert scale was used following Lyberg, Biemer, Collins, Ieeuw, Dippo, Schwarz, and Dennis (1997) who suggest an optimal scale length

of between 5 and 7 points. The questionnaire was revised twice after receiving advice from active teaching professionals.

The questionnaire had four main sections. The first section (Questions 1 to 5) was designed to ascertain Japanese secondary school teachers' beliefs and knowledge about CLT. The second section (Questions 6 to 10) elicited information about how teachers use CLT activities. The third section (Questions 11 to 14) was designed to explore what difficulties teachers faced in using CLT activities. The last section (Questions 15 to 17) tracked the influence of entrance examinations on the teachers' perception of skills necessary for English. Questions 1 to 10 and Questions 11 to 17 correspond to the first and second research questions respectively.

The internal consistency estimate of reliability for the Likert-scale questions (Questions 15 to 17) was calculated, and Cronbach's Alpha was estimated at .78.

Participants

The sample of teachers used in this study was a sample of convenience. The researcher sent a Japanese version of the questionnaire to 30 teachers in October 2003; 21 were returned by December, a response rate of 70 percent.

Of the 21 participants, 5 were the researcher's former colleagues; 3 were the researcher's classmates in a TESOL doctoral program; and 6 were members of a teachers' association. These teachers introduced 7 additional participants.

Among the 8 male and 13 female participants, 6 were teaching in public middle schools, 11 in high schools (8 private and 3 public), and 4 in both. Five had taught for 1 to 5 years, 4 for 6 to 10 years, and 12 for more than 10 years. Twelve had experience abroad (6 for 1 to 6 months and 6 for more than 6 months). All 21 teachers worked in Tokyo.

Limitations of the Method

This study is a pilot study, so the number of participants was not large. In addition, the perceptions of the participants, many of whom were actively pursuing professional development, might not reflect those of the general population. Moreover, the participants work in Tokyo and may be more aware of CLT and MEXT guidelines than those living away from the center of political power. Thus, results of this study are not generalizable. In future studies, teachers' perceptions about CLT should be

more comprehensively investigated using a stratified random sample of teachers from every prefecture.

Results

Teachers’ Beliefs About and Knowledge of CLT

Questions 1 to 5 concern Japanese teachers’ beliefs and knowledge about CLT. In response to Question 1, 19 of the 21 participants indicated that they had heard of or studied CLT. However, responses to Question 2 show that they mainly learned by themselves, not from workshops held by local boards of education. Table 1 shows that the teachers learned about CLT mainly from books and journals, or seminars and lectures. Only two teachers reported learning about CLT from *The Course of Study*.

**Table 1. Where (Or from What) the Teachers
Heard/Learned about CLT (N = 21)**

Sources and places	No. of mentions
Books or journals	12
TESOL seminars/lectures	10
University	7
Teacher’s manual	5
Workshop held by a teachers’ association	4
The Course of Study	2
Workshop held by the Board of Education	0

The teachers’ responses to Questions 3 and 4 showed that they had a relatively clear understanding of CLT. As shown in Table 2, they thought it was most important for students in CLT classrooms “to communicate effectively” and “to enjoy communicating” in L2. Moreover, the teachers selected being a “material provider,” “co-communicator,” “communication model,” and “facilitator” as their main roles in CLT classrooms (Table 3). Only a few chose native-like pronunciation or native-like accuracy as a crucial factor.

Table 2. What the Teachers Think is Important for Students in CLT Classrooms (N = 21)

Important factors for students	No. of mentions
To communicate effectively in L2	13
To enjoy communicating in L2	11
To collaborate with each other	8
To talk to a native speaker	4
Never to use L1	4
To acquire native-like fluency	1
To acquire native-like accuracy	0
To acquire native-like pronunciation	0

Table 3. What the Teachers Think is Required of Teachers in CLT Classrooms (N = 21)

Roles of teachers	No. of mentions
To provide material	15
To be a co-communicator	14
To be a communication model	13
To be a facilitator	10
To have native-like fluency	4
To have native-like pronunciation	2
To be a native speaker	2
To have native-like accuracy	1

Although the teachers did not refer to Canale and Swain's (1980) four areas of communicative competence (grammatical, sociolinguistic, discourse, and strategic competence), one can infer from their answers to Question 5 that they have a solid understanding of communicative competence. Eleven defined communicative competence as the ability to understand others' messages and to convey one's message to others. Seven referred to the ability to express one's thoughts/ideas. Two held a different perspective from the others with one connecting communica-

tive competence to identity and the other to the establishment of human relationships.

Use of Communicative Activities in the Classroom

Because one of the aims of the JET program is to promote interaction in English between Japanese Teachers of English (JTEs) and ALTs (Wada & Cominos, 1994), communicative activities may be more likely to be used in team-taught classes. Questions 6 to 10 address issues related to ALTs and team teaching.

In all schools at which the participants in this study worked, native English speakers were employed as ALTs. In the teachers’ responses to Question 6, the frequency with which native speaker teachers visited the schools varied considerably with five participants reporting visits of once a month or less, one reporting visits by ALTs every other week, and 14 schools once a week or more.³

Table 4. Frequency of Group/Pair Activities (N = 17)

Frequency	No. of mentions	
	With ALTs	Without ALTs
Never	2	2
Hardly ever	0	2
Sometimes	2	5
Often	4	2
Usually	5	4
Always	4	2

Seventeen teachers who team taught with an ALT responded to Question 8. Table 4 shows that 13 of these 17 Japanese teachers often, usually, or always used group or pair activities with ALTs, while 8 teachers frequently used such activities without ALTs. As for Question 9, a variety of activities were provided either with or without ALTs. By examining these results carefully, it was found that the frequency of the use of communicative activities *with* ALTs was a little higher. Role plays and discussion were more likely to be employed by Japanese teachers when they taught *with* ALTs (Table 5).

Table 5. Range of Communicative Activities (N = 21)

Activity	No. of mentions	
	With ALTs	Without ALTs
Game (Bingo, board game, etc.)	14	13
Information gap, ranking, listing, etc.	13	11
Role play	10	4
Discussion (group/ whole class)	8	5
Debate	1	1
Speech	1	0
Quiz	0	1
Play	0	1

In response to Question 10, 14 of 21 teachers reported that their students' favorite activity was a game.

Problems and Difficulties in Implementing CLT

Questions 11 to 14 concern problems and difficulties the Japanese teachers thought they had when they employed CLT. In response to Question 11, 10 of 21 indicated that CLT was used effectively in their schools. Surprisingly, among the 11 who did not think that they were using CLT effectively, only 2 selected entrance examinations as the main reason from the list of answers for Question 12 (Table 6).

Question 13 revealed that 18 of 21 respondents wanted to provide their students with more communicative activities. Two wrote that communicative activities would be useful after students had learned grammar and vocabulary. One teacher did not respond to the question. As to why they want to use more communicative activities, seven teachers wrote that communication in the L2 was the main objective of learning a foreign language. Three teachers, feeling that they had given too few chances to date to use English, hoped to give their students more opportunities.

Table 7 presents the responses to Question 14: What do you think should be changed first in order for you to apply CLT more effectively in your lessons? Both Table 6 and Table 7 show that the teachers considered the "number of class hours" and "class size" serious problems.

**Table 6. Reasons Why CLT Cannot Be Used
in the Classrooms (N = 11)**

Problems	No. of mentions
Number of class hours	7
Class size	6
Textbook	5
Curriculum	5
Lack of teachers' English proficiency	4
Evaluation system	4
Lack of materials for communicative activities	3
Entrance examinations	2
Teachers' views	1

**Table 7. Conditions to Be Changed in Order to Use CLT
in the Classrooms (N = 21)**

Problems	No. of mentions
Number of class hours	8
Class size	7
Teacher training	3
Curriculum	3
Textbook	2
Education system	2
Lack of materials for communicative activities	2
Entrance examinations	2
Teachers' views	1
Teachers' English proficiency	1
Cooperation with ALTs	1

Importance of Domain-Specific Skills/Knowledge

Questions 15 to 17 asked about teachers' perceptions of the importance of skills and knowledge. A 6-point Likert scale (0 = not important;

1 = little importance; 2 = slight importance; 3 = somewhat important; 4 = important, and 5 = very important) was used.

Seven one-way analyses of variance (ANOVAs) were conducted. The independent variable was the purpose of English learning (learning in general, learning for passing high school entrance examinations, and learning to pass university entrance examinations). Dependent variables included the teachers' perceptions of the importance of English skills (reading, writing, listening, speaking, grammar, vocabulary, and yakudoku). A Bonferroni adjustment was made in order to avoid committing a Type I error; thus, $p < .0071$ ($.05/7$) was used to determine statistical significance. To determine whether the data met the assumptions of ANOVA, the data in each of the 21 cells (3 times 7) were checked for normality. Three dependent variables (reading skills, grammar, and vocabulary) were positively skewed, and thus, logit transformation of the reflected variables was performed.

Three means were found to be significantly different for five dependent variables: perceived importance of listening, speaking, grammar, vocabulary, and yakudoku. The strength of relationship between the purposes and the change in perceived importance, assessed by η^2 , was relatively strong: listening 25%, speaking 53%, grammar 31%, vocabulary 25%, and yakudoku 17% (Table 8).

Table 8. Means, Standard Deviations, and One-Way Analyses of Variance (ANOVAs) for Effects of Learning Purposes on Seven Dependent Variables (N = 21)

	In general		HS exam		Univ. exam		ANOVA		
	M	SD	M	SD	M	SD	F (2, 20)	p	η^2
Reading	4.23	1.38	4.67	0.66	4.90	0.30	1.89	.159	0.06
Writing	4.05	0.80	3.65	1.27	3.76	1.30	0.67	.513	0.02
Listening	4.55	0.94	3.52	1.33	3.24	1.30	9.75	.000	0.25
Speaking	4.57	0.60	2.05	1.43	2.15	1.46	33.68	.000	0.53
Grammar	4.10	0.83	4.86	0.48	4.90	0.30	13.45	.000	0.31
Vocabulary	4.48	0.51	4.86	0.36	4.95	0.22	10.00	.000	0.25
Yakudoku	3.05	1.22	3.94	0.90	4.00	0.88	5.92	.004	0.17

Because equal variances among the three groups were not assumed, post hoc comparisons were conducted with Dunnett's C tests. Significant mean differences were found between the perceived importance of learning English in general and for passing high school entrance examinations. Significant mean differences were also found between the perceived importance of learning English in general and for passing university entrance examinations for all the five variables (perceived importance of listening, speaking, grammar, vocabulary, and yakudoku).

The results show that listening and speaking skills were perceived to be more important for learning English in general than for passing high school or university entrance examinations. They also show that grammar, vocabulary knowledge, and *yakudoku* skills were considered more important for passing entrance examinations than for learning English in general.

Discussion

What Are the Japanese Secondary School Teachers' Beliefs and Practices Regarding CLT?

The participants in this study seemed to have relatively solid knowledge of CLT and a good understanding of learners' and teachers' roles in CLT classrooms (Tables 2 and 3). However, at the same time, their responses imply that there were some problems that negatively affect the implementation of CLT.

First, only two teachers answered that they had learned about CLT from *The Course of Study* (Table 1). According to my observations, most secondary school teachers appear indifferent to the guidelines. One of the reasons for this might be that *The Course of Study* is not meant to address specific methods of instruction, but rather to describe the overall purpose of English education at secondary schools. Although it lists language activities and elements that should be taught, it does not show teaching techniques or practices useful in teaching them. *The Course of Study* specifies only what teachers are to teach, not *how* they are to teach (Gorsuch, 2000). It appears that MEXT needs to make the guidelines more practical by addressing methods of instruction, so that information related to reforms can be disseminated more efficiently.

Second, none of the participants reported that they had learned about CLT from workshops held by local boards of education (Table 1). According to Browne and Wada (1998), in-service teachers may need to receive more training that exposes them to CLT theories and practices. The in-service training planned by MEXT may be ineffective in helping teachers

learn new subject-specific teaching methodologies. Prefectural boards of education in designated cities do conduct teacher training workshops for novice teachers, teachers with 5 years of experience, and teachers with 10 years of experience. However, the length of those workshops ranges from just one day to a maximum of one week. Moreover, novice teachers and teachers with 10 years of experience also study topics other than teaching methodologies (e.g., class and school management) (Kanatani, 2004).

Furthermore, financial support for in-service training seems inadequate. Since 2003, local boards of education have held intensive in-service teacher training workshops supported by funds (US\$ 6 million per year) from the Action Plan budget (MEXT, 2006). But the amount is hardly enough to train 60,000 secondary school teachers. The budget is by far lower than that of the JET Program for hiring ALTs every year (US\$ 480 million). In addition, it is not clear whether financial support for in-service training will continue after completion of the Action Plan in 2008.

If the government's intent is to help teachers learn CLT theories and practices, then teacher training workshops should be available for all teachers, be made longer, and include training in methodologies that help promote the development of communicative abilities. It should be proposed to MEXT and local boards of education that numerous different workshops be organized.

Third, ALTs are not required to have any previous teaching experience or training (see Wada & Cominos, 1994). According to teachers' responses, role plays and discussions were more likely to be used by Japanese teachers when they taught with ALTs than alone (Table 5). As Sakui (2004) pointed out, having an ALT makes CLT more salient both for teachers and students. Gorsuch (2002) suggested that we should view the presence of ALTs as "a dynamic, if unevenly available, form of in-service teacher education" (p. 24). As suggested by Gillis-Furutaka (1994), improved pre- and in-service training for ALTs should be given, so ALTs can perform their role in introducing new teaching methods to JTEs.

In short, the participants have relatively solid CLT knowledge, but in order to enable more teachers to learn about CLT, the content of *The Course of Study* and pre- and in-service training needs to be reviewed.

What Contextual Factors Do the Japanese Teachers Think Should Be Changed in Order to Utilize CLT in Japanese Secondary Schools?

The results of this study are compatible with Gorsuch's (2000) model which suggests that teachers' perceptions about the use of CLT activities

may be influenced by changes in school and classroom conditions.

In response to Question 13, teachers in this study reported that they wanted to use more communicative activities in class. However, they believe that listening and speaking skills are less important for passing entrance examinations. They also believe that grammar, vocabulary, and *yakudoku* were more important for passing entrance examinations than for general learning. Teachers' concern over entrance examinations had a strong influence on their perceived importance of English skills and knowledge.

It must also be noted that despite their worries about entrance examinations, only 2 of 21 teachers reported that they felt entrance examinations needed to change before employing CLT in the classroom. In contrast, approximately one third of the teachers reported that educational reforms in classroom conditions (the number of class hours and class size) are a prerequisite for the effective use of CLT methods (Table 7). A teacher emphasis on reforms at the classroom level to promote the use of CLT has also been reported elsewhere. Comparing school and classroom latent variables, Gorsuch (2000) reported that while teachers are sensitive to attitude shifts toward examinations at the institutional level, they may not be inclined to implement related changes in the classroom (p. 701). She noted that teachers might be more resistant to CLT activities at the classroom level than at the school level because of their concern over control in the classroom and over students' learning. As a counter example to Gorsuch, Browne and Wada (1998) reported that the rate of academic high schools choosing listening classes (67%) was higher than that of vocational high schools (25%) due to the recent trend among universities of adding listening comprehension to their entrance examinations. These results were for oral communication courses while Gorsuch's study focused on integrated courses. There seems to be a difference between teachers' perception of oral communication courses and integrated courses. Similarly, some Japanese secondary school teachers in this study reported that they might be able to use CLT if class size and hours changed. They could maintain better control in a small class while students are doing pair/ or group work, and could guarantee adequate learning provided they are given more class hours to spend time on communicative activities.

With regards to class size, two participants reported that in their school a native speaker taught one oral communication lesson per week to approximately 520 students (11 classes of 47-50 students). They pointed out that it was hard to check and assist 24 to 25 pairs in one lesson, that classrooms sometimes became very noisy, and that administering oral

tests to 520 students was almost impossible (on the effects of large classes, see also Holliday, 1994).

In order to solve the problems of large classes and limited instruction time, the Conference of English Education Reform, attended by nine teachers' associations, has called for reform in English education since 1974. Their main proposals included having smaller class sizes (fewer than 20 students in one class) and additional lessons (Conference of English Education Reform, 1992). Consistent with these proposals, many teachers in this study also called for smaller classes and more class time.

Although MEXT has reacted to suggestions from teachers' groups and has tried to improve the situation, MEXT and local educational institutions appear to have difficulty shaping policy suited to the actual state of affairs in Japanese secondary schools. For instance, since 2001, the Tokyo Metropolitan Government Board of Education has attempted to decrease the number of students in English class by reducing the class size from 40 to 25 or 26. However, this was done only once a week and it also fell short of attaining the class size (less than 20) recommended by the Conference of English Education Reform. In addition, because the new class members were not from the same homeroom, group or pair work could not be conducted as smoothly as when students know each other well (see Kashimura, 2005). As this example suggests, MEXT and the boards of education need to listen more closely to teachers' voices. Otherwise the reforms they are aiming at may not be effective in creating a classroom environment conducive to CLT.

Li (1998) maintains that teaching methodologies developed in the West, such as CLT, are often difficult to introduce into EFL situations. Holliday (1992) argues that innovation can be effective only if appropriate to the actual conditions of host educational institutions. As such, the Ministry's educational policy, which has promoted the use of CLT in secondary schools, needs to take into consideration the educational context for teaching and learning of English in Japanese schools.

Conclusion

Gorsuch (2000) called for more research to understand teachers' concerns about school and classroom conditions and to find concrete ways to help teachers deal with these concerns when they use CLT activities. The present study contributes to this research agenda by investigating Japanese secondary school teachers' perceptions about using CLT methods following the introduction of the MEXT 1999 *Course of Study* and the

2003 Action Plan. Although this exploratory study cannot be generalized, the results suggest that for teachers to more effectively use CLT in the classroom, changes in educational conditions are necessary. The teachers in this study reported that they needed to have more class hours and smaller classes to employ CLT more effectively.

As Borg (2003) noted, contextual factors influence both teacher cognition and practice. It was found that to a certain extent beliefs and practices regarding CLT might be affected by contextual factors (class hours and class size). If educational conditions are improved, teachers' beliefs and practices may change. About half of the participants wrote that they had already begun to use CLT at the local level. In order for this small step to be the first toward real reform in English education in Japan, we need to listen more carefully to teachers' voices and learn what conditions teachers really want to change.

Notes

1. The Japanese Ministry of Education was combined with the Ministry of Science and Technology in 2002. Since then, it has been called the Ministry of Education, Culture, Sports, Science, and Technology. "The Ministry of Education," "the Ministry," and "MEXT" are all used in this paper.
2. The 1989 version of *The Course of Study* did not explicitly state that one of the goals of the course was to develop students' communication abilities. It simply stated that the goal was to "foster the positive attitudes toward communicating in a foreign language" (MEXT, 1989).
3. In 2005, two years after the administration of the questionnaire, the average frequency of ALT visits was 0.7 times a week in junior high schools and 0.6 times a week in high schools (MEXT, 2006).

Acknowledgements

I would like to thank the reviewers for their very helpful comments on previous drafts of this paper. I would also like to express my special thanks to David Beglar, Eton Churchill, Lyn Churchill, and Ethel Ogane for their valuable feedback and suggestions.

Takako Nishino, M. Ed., TESOL, Temple University Japan, is currently working on her Ed. D. degree. She teaches at Hosei University and Kanda University of International Studies.

References

- Abelson, R. P. (1979). Differences between belief and knowledge systems. *Cognitive Science*, 3(4), 355-366.
- Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, 36, 81-109.
- Brown, J. D. (1995). English language entrance examinations in Japan: Problems and solutions. In K. Bradford-Watts, C. Ikeguchi, & M. Swanson (Eds.), *Proceedings of the JALT 1995 Conference* (pp. 272-283). Tokyo:JALT.
- Brown, J. D. (2001). *Using surveys in language programs*. Cambridge: Cambridge University Press.
- Browne, C. M., & Wada, M. (1998). Current issues in high school English teaching in Japan: An exploratory survey. *Language, Culture, and Curriculum*, 11 (1), 97-112.
- Canale, M. and Swain, M., (1980), Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1, 1-47.
- Conference of English Education Reform. (1992). *Nihon eigo kyouiku kaizen kondankai: Apiiru youbousho, 1974-1992 [Conference of English Education Reform: Proposals, 1974-1992]*. Tokyo: Author.
- Fang, Z. (1996). A review of research on teacher beliefs and practices. *Education Research*, 38 (1), 47-64.
- Freeman, D. (2002). The hidden side of the work: Teacher knowledge and learning to teach: A perspective from North American educational research on teacher education in English language teaching. *Language Teaching*, 35, 1-13.
- Gillis-Furutaka, A. (1994). Pedagogical preparation JET program teachers. In M. Wada and & A. Cominos (Eds.), *Studies in team teaching* (pp. 29-41). Tokyo: Kenkyusha.
- Gorsuch, G. (1998). *Yakudoku* EFL instruction in two Japanese high school classrooms: An exploratory study. *JALT Journal*, 20 (1), 6-32.
- Gorsuch, G. (2000). EFL educational policies and education cultures: Influences on teachers' approval of communicative activities. *TESOL Quarterly*, 34 (4), 675-710.
- Gorsuch, G. (2001). Japanese EFL Teachers' Perceptions of Communicative, Audiolingual and *Yakudoku* activities: The plan versus the reality. *Education Policy Analysis Archives*, 9 (10). Retrieved January 3, 2005, from <http://epaa.asu/epaa/v9n10.html>
- Gorsuch, G. (2002). Assistant foreign language teachers in Japanese high schools: Focus on the hosting of Japanese teachers. *JALT Journal*, 24 (1), 5-32.
- Holliday, A. (1992). Tissue rejection and informal orders in ELT projects: Collecting the right information. *Applied Linguistics*, 13 (4), 403-424.

- Holliday, A. (1994). *Appropriate methodology and social context*. Cambridge: Cambridge University Press.
- Imura, M. (2003). *Nipponno eigokyouiku nihiyakunen [History of English education in Japan]*. Tokyo: Taishyukan.
- Kagan, D. M. (1992). Implications of research on teacher belief. *Educational Psychologist*, 27 (1), 65-90.
- Kanatani, K. (2004). *Eigojogyo kaizenno tameno shohousen: makuronni kangae mikuronni taishosuru [Prescription for improving English lessons: Think globally and act locally]*. Tokyo: Taishyukan.
- Kashimura, M. (2005). *Shuujukudobetsu-jugyo ga ataeru kyoushito kodomoeno eikyo [Influence of "a small class according to level" on students and teachers]*. *The New English Classroom*, 433, 7-10.
- Law, G. (1994). College entrance exams and team teaching in high school English classrooms. In M. Wada & A. Cominos (Eds.), *Studies in team teaching* (pp. 90-102). Tokyo: Kenkyusha.
- Li, D. (1998). "It's always more difficult than you plan and imagine": Teachers' perceived difficulties in introducing the communicative approach in South Korea. *TESOL Quarterly*, 32 (4), 677-703.
- Lyberg, L., Biemer, P., Collins, M., Ieeuw, E. d., Diplo, C., Schwarz, N., & Dennis, T. (1997). *Survey measurement and process quality*. New York: Wiley-Interscience.
- MEXT. (1989). *The Course of Study*. Tokyo: Author.
- MEXT. (1999). *The Course of Study*. Tokyo: Author.
- MEXT. (2003). *Regarding the establishment of an action plan to cultivate "Japanese with English abilities"*. Retrieved August 12, 2003, from <http://www.mext.go.jp/english/topics/03072801.htm>
- MEXT. (2006). *Eigokyouiku kaizen jisshi-joukyou chousakekka gaiyou [Statistical abstract on the reform of the English education]*. Retrieved December 16, 2006, from http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo3/siryo/015/05071201/004
- Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies*, 19 (4), 317-328.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62 (3), 307-332.
- Richards, J., Platt, J., & Platt, H. (1992). *Dictionary of language teaching & applied linguistics*. London: Longman.
- Richardson, V. (1994). Conducting research on practice. *Educational Researcher*, 23 (5), 5-10.
- Sakui, K. (2004). Wearing two pairs of shoes: Language teaching in Japan. *ELT Journal*, 58 (2), 155-163.
- Suzuki, T. (1999). *Nihonjinwa naze eigoga dekinai. [Why can't Japanese people function in English?]*. Tokyo: Iwanami Shoten.
- Taguchi, N. (2002). Implementing oral communication classes in upper secondary schools: A case study. *The Language Teacher Online*. Retrieved October 31, 2004,

- from <http://www.jalt-publications.org/tlt/articles/2002/12/taguchi>
 Taguchi, N. (2005). The communicative approach in Japanese secondary schools: Teachers' perceptions and practice. *The Language Teacher*, 29 (3), 3-12.
 Wada, M., & Cominos, A. (1994). Language policy and the JET program. In M. Wada & A. Cominos (Eds.), *Studies in team teaching* (pp. 1-6). Tokyo: Kenkyusha.

Appendix

CLT Questionnaire

I would like to know how Japanese teachers feel about communicative language teaching (CLT). Please answer the following questions.

Background Information: Please circle the item that best describes your background and current teaching situation. (Check all items that apply.)

Sex: Male Female

Present Teaching Position:

Junior high school	High school
University	Others ()

Type of school: Public Private National

Areas you teach: Reading Writing Oral communication
 Grammar Others ()

Numbers of years teaching English:

1-5 years 6-10 years Over 10 years

Experience of living abroad:

None 1-6 months 6-12 months
 1-3 years Over 3 years

Questions

Please check the items that apply to you. (Check all items that apply in Questions 2, 3, 4, 9, and 10.)

- Have you ever heard/learned about communicative language teaching (CLT)?
 _____ Yes _____ No. (If no, please skip the questions 2 ~ 4.)
- Where did you learn about communicative language teaching (CLT)?
 _____ books or journals _____ TESOL seminars/lectures
 _____ teachers' manual _____ *the Course of Study*
 _____ workshop held by the board of education

- ☐ University
☐ workshop held by a teachers' association
☐ others ()
3. What do you think is important for students in CLT classrooms?
- ☐ to talk to a native speaker
☐ to acquire native-like pronunciation
☐ to acquire native-like fluency
☐ to acquire native-like accuracy
☐ to communicate effectively in L2
☐ never to use L1 (Japanese)
☐ to collaborate with each other
☐ to enjoy communicating in L2
☐ others ()
4. What do you think is required for English teachers in CLT classrooms?
- ☐ to be a native speaker
☐ to have native-like pronunciation
☐ to have native-like fluency
☐ to have native-like accuracy
☐ to provide material
☐ to be a facilitator
☐ to be a communication model
☐ to be a co-communicator
☐ others ()
5. What is your understanding of "communicative competence?"
- _____
- _____
6. Does a native English speaker teach in your school?
- ☐ Yes ☐ No
 If yes, how often do they teach?
☐ not regularly ☐ once a month
☐ once in a few weeks ☐ once a week
☐ twice a week ☐ more than three times a week
7. Do you have a team-taught class with an ALT? If yes, in what class?
- ☐ Yes. [] ☐ No.
 (If no, please skip the questions regarding an ALT.)

8. How often do you use group / pair activities in your lesson?
 With an ALT: ☐ never ☐ hardly ever ☐ sometimes
☐ often ☐ usually ☐ always
 Without an ALT: ☐ never ☐ hardly ever ☐ sometimes
☐ often ☐ usually ☐ always
9. Which of the following activities have you used in your lessons?
 With an ALT: ☐ information gap ☐ problem solving
☐ discussion ☐ listing / ranking
☐ role-play ☐ games ☐ others ()
 Without an ALT: ☐ information gap ☐ problem solving
☐ discussion ☐ listing / ranking
☐ role-play ☐ games ☐ others ()
10. Which of the following activities do you think your students prefer?
☐ information gap ☐ problem solving ☐ discussion
☐ listing / ranking ☐ role-play ☐ games
☐ others ()
11. Do you think CLT is employed effectively in your school?
☐ Yes ☐ No
12. If no, which of the following factors do you think is the biggest problem?
☐ lack of materials for communicative activities
☐ entrance examinations
☐ lack of teachers' English proficiency
☐ curriculum (*The Course of Study*)
☐ textbook ☐ class size
☐ number of class hours ☐ evaluation system
☐ others ()
13. Do you want to provide your students with more communicative activities? Why?
☐ Yes ☐ No
 Reasons: _____

14. What do you think should be changed first in order for you to apply CLT more effectively in your lessons?

15. In your opinion, how important are the following areas for your students to learn English? (Circle the number that best describes the degree of importance that you attach to the item on the left.)

Importance	None	Little	Slight	Somewhat	Important	Very Important
Reading	0	1	2	3	4	5
Writing	0	1	2	3	4	5
Listening	0	1	2	3	4	5
Speaking	0	1	2	3	4	5
Grammar	0	1	2	3	4	5
Vocabulary	0	1	2	3	4	5
Yakudoku	0	1	2	3	4	5

16. In your opinion, how important are the following areas for your students to pass high school entrance examinations?

Importance	None	Little	Slight	Somewhat	Important	Very Important
Reading	0	1	2	3	4	5
Writing	0	1	2	3	4	5
Listening	0	1	2	3	4	5
Speaking	0	1	2	3	4	5
Grammar	0	1	2	3	4	5
Vocabulary	0	1	2	3	4	5
Yakudoku	0	1	2	3	4	5

17. In your opinion, how important are the following areas for your students to pass university entrance examinations?

Importance	None	Little	Slight	Somewhat	Important	Very Important
Reading	0	1	2	3	4	5
Writing	0	1	2	3	4	5
Listening	0	1	2	3	4	5
Speaking	0	1	2	3	4	5
Grammar	0	1	2	3	4	5
Vocabulary	0	1	2	3	4	5
Yakudoku	0	1	2	3	4	5

日本語話者大学生の英語学習動機の変化 —国際イベントへのボランティア参加の効果—

Changes in Japanese University Students' Motivation to Learn English: Effects of Volunteering in an International Event

酒井英樹(さかい ひでき)
小池浩子(こいけ ひろこ)
信州大学

The purpose of this study was to examine the effects of volunteer assistance at the 2005 Special Olympics World Winter Games (SO) in Nagano on Japanese university students' motivation to learn English. The construct of motivation was investigated within the framework of self-determination theory, which assumes three basic psychological needs: for competence, for relatedness, and for autonomy. According to the degree to which these psychological needs are satisfied, social-contextual factors are considered to facilitate or impede motivation. The theory posits that in terms of the degree of self-determination or autonomy, motivation is categorized as (a) amotivation, (b) extrinsic motivation, and (c) intrinsic motivation. Extrinsic motivation is further divided into (a) external regulation, (b) introjected regulation, (c) identified regulation, and (d) integrated regulation; in this order, the degree of self-determination increases. Previous studies on motivation to study English as a second language (L2) within the framework of self-determination theory (Hiromori, 2003a, 2006; Hiromori & Tanaka, 2006) have shown that L2 learners' motivation to study may change and that social-contextual factors facilitative for basic psychological needs may influence changes in motivation.

The research questions posed for this study were: (a) Did participation in the SO as volunteers cause changes in Japanese university students' motivation to study English? and (b) if so, how? We supposed that such participation would provide

the students with opportunities through which their basic psychological needs might be met. Thus, it was hypothesized that (a) participation in the SO would lead to changes in motivation to study English and that (b) the motivational types with higher self-determination (e.g., intrinsic motivation) would be enhanced and the motivational types with lower self-determination (e.g., amotivation and external regulation) would be diminished.

A 16-item questionnaire was administered twice: once before the event (survey 1) and then two months after the event (survey 2). Responses from 44 students were analyzed. Because of deviation from the normal distribution for the scores for amotivation, nonparametric tests (Wilcoxon signed-rank tests and Friedman's tests) were performed.

Descriptive statistics showed that the median scores of amotivation were the lowest for both survey 1 and survey 2, whereas the median scores of identified regulation were the highest and that the median scores of external regulation and introjected regulation decreased within the two-month period, whereas the median scores of amotivation, identified regulation, and intrinsic motivation increased. Friedman's tests and the post-hoc tests revealed that, before participating in the SO, amotivation was significantly lower than any other type of motivation, while identified regulation was significantly higher than any other type of motivation. There were no significant differences among the other combinations of motivational types on survey 1. After two months, changes were observed: since the median scores of amotivation and intrinsic motivation increased and the median score of introjected regulation decreased, the significant difference found between amotivation and introjected regulation on survey 1 disappeared on survey 2, and the difference between introjected regulation and intrinsic motivation became significant on survey 2. Wilcoxon signed-rank tests revealed that the change in the median scores of intrinsic motivation between survey 1 and survey 2 was significant. In summary, the results confirmed the hypotheses.

The discussion of the results is structured around the following three points. First, the findings support the arguments made by Hayami (1995) and Horino and Ichikawa (1997) that motivation should be treated as a dynamic changeable cognitive state. At the same time, considering the two-month interval after the SO, the changes of motivation observed in this study may not be temporal, but durable. Second, the results suggest that participating in international events like the SO may be facilitative for the improvement of motivation even if the events are short. Most of the participants in this study reported that they had assisted at the SO for three to five days ($n = 40$, 90.9%). Third, it is pointed out that satisfaction of the need for competence alone may not influence motivational changes because most ($n = 28$, 63.6%) reported that they had spent a total of two hours or less speaking English during the volunteer work. Thus, the participants may not have had sufficient experience in communicating with others in English.

Finally, several limitations are discussed: the sample characteristics, the design without any control groups, the questionnaire with a small number of items for

each type of motivation, and the unresolved question as to how participation in the SO influenced motivation. Future research is needed to overcome these limitations and thereby to obtain insight into how participation in international events should be incorporated into language programs.

本研究の目的は、2005年スペシャル・オリンピックス (SO) 冬季世界大会・長野にボランティアとして参加したことが、日本語話者大学生の英語学習の動機づけにどのような影響を及ぼしたのかを調べることである。動機づけの理論の1つである自己決定理論を援用して作成した質問票をSOの直前と2ヶ月後に実施した。SOにボランティアとして参加した大学生44名の回答を分析対象とした。結果は次の通りである。外的調整と取り入れの調整の中央値は、事後調査において下がり、一方、無動機、同一視的調整、内発的動機の中央値は事後調査において上がった。これらの変化のうち、統計的に有意であったのは内発的動機であった。すなわち、SOのような国際的イベントへのボランティアとしての参加が、日本人大学生の英語学習動機に影響を及ぼしていることが確認された。

本 研究の目的は、2005年スペシャル・オリンピックス (SO) 冬季世界大会・長野 (2005 Special Olympics World Winter Games, Nagano) にボランティアとして参加したことが、日本人大学生の英語学習の動機づけにどのような効果を及ぼしたのかを調べることである。SOなどの国際的なイベントに参加することは、世界のいろいろな地域の人と触れる、自分の英語力を実際に使う体験をする、他の人が英語を使っている場面を見て英語が国際語として使われている現実を知る、といった経験の場を与えることになると考えられる。これらの経験は、英語の学習動機に影響を与えるだろうと考えた。この影響を調べるために、本研究では、動機の変容という側面を扱っている動機づけ理論の1つである自己決定理論を援用した。

自己決定理論 (self-determination theory)

英語の学習動機とは、簡単に言ってしまうと、英語を学ぶ理由である。Dörnyei (2001) によれば、動機は、「人の行動の方向性と強さ」¹⁾に関するもので、人の意識を英語の学習に向かわせ、英語の学習を始めさせたり、その学習を維持させたりする原動力であるとされる (p. 8)。この動機が、どのようにして強まったり弱まったりするのかという問題は、教育的に重要な課題である (Dörnyei, 1994; 池野, 2003; 中田, 2006)。

動機づけに関して内発的・外発的動機という区分がある。鹿毛 (1994) は、動機づけの理論を概観し、内発的・外発的動機がどのように概念化されているか整理し、(a) 認知的動機づけによる概念化 (情報収集とその体制化が目標 vs 一次的欲求の充足が目標)、(b) 手段性・目的性による概念化 (自己目的性 vs 手段性、道具性)、(c) 自己決定による概念化 (行為の原因の所在が内的であるという認知 vs 行為の原因の所在が外的であるという認知)、(d) 感情による概念化 (行為に没頭してときに感じるフローという包括的感覚や興味・興奮が内発的動機づけの本質)、(e) 測定尺度の開発を通じた包括的な概念化、という 5 つに分類した (pp. 350-352)。鹿毛はその上で、「内発的動機づけとは、『自己目的的な学習の生起・維持過程』である」と述べ、「認知的動機づけ、好奇心、挑戦、成就といった概念を統合」する「熟達指向性」と、「自ら進んで学習に取り組む」という「自律性」の両性質を持っているもので

であると指摘した (p. 353)。この鹿毛の分類に従えば、3つ目の自己決定による概念化を行っているのが自己決定理論 (self-determination theory) である。

自己決定理論について、Ryan & Deci (2002) に基づいて簡略に説明する。自己決定理論は、まず、「あらゆる個人は、より精緻な、また統合された自己感覚を発達させたいという、自然で生得的で、建設的な傾向を持つ」(p. 5) と想定する。この生得的な傾向は、社会・文脈的な要因 (social-contextual factors) の影響を受けると考えられている。この生得的な傾向と外的な要因の間に仮定されているのが、基本的な心理欲求 (basic psychological needs) である。基本的な心理欲求として、有能性の欲求 (the need for competence)、関係性の欲求 (the need for relatedness)、そして自律性の欲求 (the need for autonomy) が提案されている (pp. 7-8)。有能性とは、社会環境との関わりにおいて自分が有能であると感じることであり、自信や効力感のことである。関係性とは、他の人とつながっていると感じることであり、他の人やコミュニティへの所属感を指す。自律性は、自分自身の意思によって当該の行為を開始していると認識していることである。社会・文脈的な要因がこれらの欲求をどの程度充足するのかによって、動機は高められたり、阻害されたりする (p. 9)。

また、自己決定理論は、自己決定性 (self-determination) や自律性 (autonomy) の度合いによって、動機づけとそれに関係している調整 (regulation) の種類を分類している。自己決定性の最も低い動機づけが、調整が存在していない無動機 (amotivation) である。自己決定性の最も高い動機づけは、内発的に調整されている内発的動機 (intrinsic motivation) である。その間に、外発的動機 (extrinsic motivation) が位置づけられている。外発的動機は、さらに、外的調整 (external regulation)、取り入れの調整 (introjected regulation)、同一視的調整 (identified regulation)、統合的調整 (integrated regulation) に分類されている。すなわち、無動機、外発的動機 (外的調整、取り入れの調整、同一視的調整、統合的調整)、内発的動機という順番に、自己決定性が高くなっていく。

無動機 (amotivation) は、「行動の意思が欠如した状態」(p. 17) である。一方、内発的動機は、「興味や本質的な満足から活動を行っている状態」(p. 17) である。無動機と内発的動機の間位置する外発的動機は、次のように説明されている。まず、外的調整は、「最も自律性のない外発的な動機」(p. 17) で、報酬の獲得や罰の回避などによってもたらされる。取り入れの調整は、「内在化されているが、深い意味で自分自身のものとして本当に受け入れていない外的な調整」(p. 17) である。自尊心の維持のための、罪や恥の意識の回避によってもたらされる。同一視的調整は、「ある行動的目標や調整の意識的な価値づけや、その行動を個人的に重要なものとして受け入れること」(p. 17) を含み、かなり高い自己決定性を示す。しかし、必ずしも個人の信念や価値と一致するとは限らない。最後に、統合的調整は、「同一視が評価され、すでに自己の一部となっている個人的に認識されている価値観、目標、欲求などと一致するようになっている」(p. 18) ときに生じる。内発的動機と異なるのは、興味や楽しさから行動するのではなく、「個人的に重要な結果」(p. 18) を得るために行動するという点である。

自己決定理論の特徴は、これらの区分が一元的に並べられるところであろう。これらの区分は、連続体 (a continuum) として捉えられており、自己決定性が高い方向に調整されていく過程は内在化 (internalization) と呼ばれている。鹿毛 (1995) は、自己決定理論の考え方について、「状況必然性を出発点とした学習意欲が、社会的・文

化的価値を内面化するにしたがって、自己必然性と内容必然性の内容を含むようになって、学習意欲が構造的に発達していくことを示唆している。」(p. 162) と述べている。つまり、自己決定理論は、動機の変容という側面を扱っているのである。

第2言語の学習に関して自己決定理論に基づく実証研究が数多く報告されている(e.g., 廣森, 2003a, 2003b, 2005; Hiromori, 2006; 廣森・田中, 2006; Honda & Sakyu, 2004; Noels, Pelletier, Clément, & Vallerand, 2000)。これらの先行研究の中でも、本節では学習者の動機の変化に扱ったものについて紹介する。

廣森 (2003b) は、英語学習の動機づけの変化を横断的に調査した。対象者は、高校1年生 (72名)、高校2年生 (68名)、高校3年生 (63名) であった。自己決定理論に基づく質問票は、5つの動機の種類 (内発的動機、同一視的調整、取り入りの調整、外的調整、無動機) に対して3項目ずつ計15項目を含み、5件法であった。クラスター分析の結果、各学年は3つのクラスターに分類された。学年が進むにつれて、外的調整が減少し、無動機が強いクラスターか、どの動機についても低く認知するクラスターか、受験などを控え英語の重要を認識しているクラスターかに分かれたことが指摘された。

廣森・田中 (2006) は、自己決定理論が内発的動機を高める要因として考えている3種類の心理的欲求 (自律性の欲求、有能性の欲求、関係性の欲求) が教室内において果たす役割を検討するために、113人の大学生を対象に、3種類の心理的欲求を充足すると考えられる授業実践を行った効果を調べた。実践した授業は、5週間にわたる英語によるグループ・プレゼンテーション活動であった。また、授業実践の前後に、心理的欲求と内発的動機を測定する質問票調査を実施した。その結果、心理的欲求も内発的動機も、授業実践の後に統計的に有意な上昇が見られた。また、潜在曲線モデルによる検討の結果、内発的動機の上昇に、自律性の欲求と有能性の欲求が影響を及ぼしていることが示された。

Hiromori (2006) は、3つの心理的欲求を刺激する可能性のある教育的介入を行うことによって、第2言語学習者の動機づけを促進することが可能かどうかを調べた。参加者は、100人の大学生であった。教育的介入の前後において、心理的欲求と動機づけを測定するための質問票調査が実施された。動機づけに関する質問票 (18項目) では、内発的動機だけでなく、外発的動機 (同一視的調整、取り込み的調整、外的調整) や無動機に関してそれぞれ3~4項目用意された。回答方法は5件法であった。実施された教育的介入は、12週間にわたる英語ライティング活動 (creative writing activities with a student self-monitoring technique) であった。動機づけに関する事前質問票調査の結果に基づき、クラスター分析を行い、参加者は4つの群に分類された。外的に動機づけられた群 (the externally-motivated group, 17人)、内発的に動機づけられた群 (the intrinsically-motivated group, 32人)、動機づけられていない群 (the unmotivated group, 13人)、そして内的プレッシャー群 (the internal-pressure group, 38人) である。教育的介入の結果、動機づけに肯定的な効果が見られた。例えば、動機づけられていない群では、無動機が統計的に有意に減少し、内発的動機と、取り込み的調整が統計的に有意に上昇した。一方、内発的に動機づけられた群では、外的調整が統計的に有意に減少し、取り込み的調整と同一視的調整が統計的に有意に上昇した。3つの心理的欲求の変化と内発的動機づけの変化の相関係数の分析の結果、群ごとに3つの心理的欲求の重要性が異なることが分かった。例えば、動機づけられていない群では、有能性の欲求の変化が内発的動機づけの変化と強

い相関を示したのに対して、内発的に動機づけられた群では、自律性の欲求が正の相関を、関係性の欲求が負の相関を示した。

これらの先行研究の結果は、次の2点にまとめられよう。まず、これら3つの研究(廣森, 2003b; Hiromori, 2006; 廣森・田中, 2006)は英語学習の動機は変化することであることを示している。別の言い方をすると、自己決定理論に基づいて動機の変容を捉えようとするのが妥当であることを示唆している。第2に、社会的・環境的な要因が英語学習の動機の変容に影響を及ぼしていることを示している。特に、英語グループ・プレゼンテーション活動(廣森・田中, 2006)と英語ライティング活動(Hiromori, 2006)という異なる教育的介入であっても、動機の変容をもたらしたことは興味深い。社会的・環境的要因(学習環境や教育的介入)が自律性、有能性、関係性の欲求をどの程度充足するかによって、学習者の動機は変化するという Ryan & Deci (2002)の主張を支持するものであるからである。これらの先行研究に基づいて、本研究では、英語学習の動機の理論として、自己決定理論を採用した。さらに、基本的心理欲求を満たす環境としてSOへのボランティア参加に焦点をあてた。

SO について

特定非営利活動法人2005年スペシャル・オリンピックス冬季世界大会・長野(SONA)のパフレットに基づいて、SOについて概要を説明する。SOとは、知的障害のある人たちの「健康や体力増進、競技力の向上を促進するだけでなく」、たくさんの人たちと関わる機会を増やし、「社会性を育てる」ために実施されている(p. 3)。そのため、「最後まで競技をやり終えた一人ひとりの健闘を称え、全員が表象され」という特徴がある(p. 3)。

2005年2月26日から3月5日までの8日間実施されたSO 冬季世界大会・長野は、約80ヶ国・地域から約2,500人の選手が参加する大会であった。競技は、アルペンスキー、クロスカントリースキー、スノーボード、スノーシューイング、スピードスケート、フィギュアスケート、フロアホッケーの7競技79種目であった。スポーツ以外のプログラムとしては、トーチラン、ホストタウンプログラム(ホームステイプログラム)、ヘルシー・アスリート・プログラム(HAP、参加者のヘルスケアの実施)、グローバル・ユース・サミット(競技に参加しない知的障害のある学生と知的障害のない学生との討論会)、スペシャルオリンピックスタウン(交流とレクリエーションの場)、及び文化芸術プログラムが実施された。

研究課題

本研究の研究課題は、次の通りである。

1. ボランティアとしてSOに参加した経験は、教育学部生の英語学習の動機に影響を及ぼすか。
2. もし及ぼすとすれば、どのような影響か。

1番目の研究課題に関して、SOの参加経験が英語学習の動機に影響を及ぼすだろうという仮説が立てられた。SOでは英語が共通語の1つであるという認識がなされており、英語によってコミュニケーションが達成されているという社会的状況が存在している。このような社会的環境に関わることによって、学習者は英語に対す

る価値観を内在化させる可能性がある。自己決定理論の基本的心理欲求についていえば、英語を使用する場面が与えられ、有能性への欲求を満たす可能性がある。さらに、英語を通じて、SOの参加者と交流したり、ボランティアの輪を広げるなど、関係性の欲求も満たされよう。また、自ら選択しボランティアに参加していることを考えれば、自律性の欲求を満たすことも考えられる。つまり、SOは、自律性・関係性・有能性の欲求を満たしうる場面であると考えられよう。その結果、学習者の英語学習の動機は変容すると予測される。²

2番目の研究課題については、より自己決定性の高い種類の動機が強まり、より自己決定性の低い種類の動機が弱まるという仮説が立てられた。Hiromori (2006) は、基本的心理欲求を満たす教育的介入によって、より自己決定性の高い種類の動機（内発的動機、同一視的調整、取り込み的調整）が上昇したのに対して、より自己決定性の低い種類の動機（無動機や外的調整）が減少したことを報告している。先に述べたように、SOへの参加は基本的心理欲求を満たす経験であると考えられるため、Hiromori (2006) と同様の変容が見られると予測される。

方法

参加者

2005年2月、SO冬季世界大会・長野にボランティアとして参加した日本語話者大学生に対して質問票調査を実施した。参加者は全員、日本の中部地区に位置するA大学の教育学部生であった。SO参加直前に行われた質問票調査（事前調査）に回答した人数は 72名だった。SOの終了後、約2ヶ月経ってから、もう一度同じ質問票調査（事後調査）を実施した。SO直後ではなく、2ヶ月後に質問票調査を実施した主な理由は、一時的なボランティア経験の効果ではなく、安定した効果を検証できると考えたからである。事後調査の回答者数は、56名であった。記入漏れのある回答者（4名）と研究目的のデータ使用の不許可者（1名）を分析から除外した。さらに、海外経験について、1ヶ月以上の回答者（5名）と、無回答者（1名）を除いた。また、外れ値を検討するために、標準化した z 値の絶対値が3.29より大きいかどうかを調べた（Field, 2005, p. 79）。外れ値を示した1名を除いた結果、最終的に分析対象となった人数は44名であった。回答者に関する情報は、表1を参照されたい。

表2は、ボランティアの参加状況についてまとめた結果である。最も多いボランティアの参加日数は、4日間であった。教育学部生が参加したボランティアの場所は、競技会場ではなく、ヘルシー・アスリート・プログラム（HAP）とスペシャルオリンピックスタウンであった。そのため、仕事内容は、道案内、通訳、受付、ガイド、事務作業、HAP などであった。HAP の仕事は、主に障害児教育専攻の学生が担当した。英語を使う機会については、「合計すると、何時間ぐらい英語を使う機会がありましたか」という質問を行った。1時間未満が最も多く、その次に1～2時間であった。

表1. 回答者の情報

特徴	回答者全員 (N = 51 ^a)		分析対象者 (N = 44)	
	人数	割合	人数	割合
性別				
男性	8	15.7%	6	13.6%
女性	43	84.3%	38	86.4%
年齢				
19歳	1	2.0%	0	0.0%
20歳	18	35.3%	16	36.4%
21歳	24	47.1%	23	52.3%
22歳	8	15.7%	5	11.4%
学年				
大学2年生	22	43.1%	18	40.9%
大学3年生	28	54.9%	25	56.8%
大学4年生	1	2.0%	1	2.3%
分野				
臨床学校教育学	5	9.8%	3	6.8%
総合・生活科教育	2	3.9%	2	4.5%
国語教育	1	2.0%	1	2.3%
社会科教育	1	2.0%	0	0.0%
数学教育	3	5.9%	3	6.8%
理科教育	1	2.0%	1	2.3%
美術教育	1	2.0%	1	2.3%
保健体育科教育	1	2.0%	1	2.3%
ライフプランニング教育	1	2.0%	1	2.3%
英語教育	7	13.7%	7	15.9%
国際理解教育	7	13.7%	5	11.4%
障害児教育	16	31.4%	16	36.4%
心理臨床	5	9.8%	3	6.8%
海外経験				
ない	20	39.2%	20	45.5%
1ヶ月未満	25	49.0%	24	54.5%
1ヶ月以上1年未満	4	7.8%	0	0.0%
1年以上2年未満	1	2.0%	0	0.0%
無回答	1	2.0%	0	0.0%

^a回答者56名のうち、記入漏れのあるもの4名と、不許可者1名を除いた51名の情報である。

表2. ボランティア参加状況 (N = 44)

特徴	人数	割合
参加日数		
2日間	1	2.3%
3日間	5	11.4%
4日間	29	65.9%
5日間	6	13.6%
6日間	2	4.5%
その他	1	2.3%
英語使用時間		
機会は少しもなかった	5	11.4%
1時間未満	13	29.5%
1～2時間	10	22.7%
2～3時間	3	6.8%
3～4時間	6	13.6%
4時間以上	6	13.6%
無回答	1	2.3%

調査を実施したとき、参加者が所属するA大学のカリキュラムでは、1年生と2年生は英語演習を受講することが必修となっていた。1学年は週2コマ（4技能を統合した総合的な英語演習を半期2コマずつ）であり、2学年は週1コマ（ライティングを中心とする授業とプレゼンテーションを中心とする授業を各半期ずつ）であった。したがって、本研究の参加者が質問票に回答したときには、すべての参加者は必修の英語演習の授業を受講し終わっていたことになる。ただし、英語教育分野の学生及び英語の教員免許を取得しようとする学生は、3年次以降も、英語科教育法、英語学、英米文学などの英語関連の専門科目をとることになる。

質問票

質問票は、24項目から成る（そのうち6項目は、錯乱的質問項目である）。先行研究（廣森, 2003a; Kimura, Nakata, & Okumura, 2001; Noels, Pelletier, Clément, & Vallerand, 2000など）を参考にして質問項目を作成した。質問項目の作成にあたっては、大学生3名に試行し、文言の修正を行ったり、曖昧な質問項目の削除を行った。また、回答時に英語の授業を履修していない参加者が答えられない質問項目を分析から除いた。³その結果、内発的・外発的動機づけに関する質問項目の詳細は、無動機（2項目）、外的調整（5項目）、取り入りの調整（3項目）、同一視的調整（2項目）、内発的動機（4項目）となった（表3参照）。なお、統合的調整は、同一視的調整と区別しづらいとする先行研究（Noels, Pelletier, Clément, & Vallerand, 2000）にならい、質問票

に含めなかった。質問項目の順番は、無作為に並べ替えられた。回答方法には5件法を用いた (1.「まったくそう思わない」、2.「あまりそう思わない」、3.「どちらともいえない」、4.「ややそう思う」、5.「非常にそう思う」)。事前調査と事後調査において、同じ質問票が用いられた。

表3. 動機の種類と質問項目

動機の種類	(項目数)	質問項目
無動機 (2項目)		
	1.	自分にとっての英語を学ぶ意義がわからない。
	2.	英語の学習は時間の無駄であるという感覚がある。
外的調整 (5項目)		
	1.	私が英語を学ぶ理由は、両親や先生を喜ばせたいからである。
	2.	私が英語を学ぶ理由は、試験 (入学試験、資格試験、就職試験など) に合格するためである。
	3.	私が英語を学ぶ理由は、最近では日本人全員が英語を理解できることが求められているからである。
	4.	もし英語を学ぶ必要性がなければ、英語を学ばないだろう。
	5.	英語ができれば、よい仕事が得られると思う。
取り入れの調整 (3項目)		
	1.	私が英語を学ぶ理由は、英語が話せれば有能であると自分自身が思えるからである。
	2.	私が英語を学ぶ理由は、英語を話す人たちと話せないと気まずいからである。
	3.	私が英語を学ぶ理由は、友だちに自分が有能な人間であると思わせたいからである。
同一視的調整 (2項目)		
	1.	私が英語を学ぶ理由は、英語が自分の成長にとって役立つと考えるからである。
	2.	私が英語を学ぶ理由は、二つ以上の言語を話せるような人間になりたいと思うからである。
内発的動機 (4項目)		
	1.	私が英語を学ぶ理由は、英語が好きだからである。
	2.	私が英語を学ぶ理由は、英語の学習が楽しいからである。
	3.	私が英語を学ぶ理由は、英語が話されているのを聞くのが心地よいからである。
	4.	私が英語を学ぶ理由は、英語を話していると気持ちがよいからである。

分析方法

5種類の動機に対する質問項目の数が異なるため、動機の種類ごとの平均点を求めた。正規分布からの逸脱を、歪度と尖度の点から検討した。歪度と尖度を標準誤差で割って得られる標準化された値の絶対値が 1.96 よりも大きいと、5%水準で正規分布から逸脱していると考えられる (Field, 2005, p. 72)。無動機の歪度 (事前調査と事後調査) と尖度 (事後調査) が有意に大きかった (表4参照)。そのため、正規分布の仮定を必要としないノンパラメトリック検定を実施した (p. 521)。

本研究の実験計画は、時間 (事前調査と事後調査) と動機の種類 (無動機、外的調整、取り入的調整、同一視的調整、内発的動機) の被験者内要因を2つ含んでいる。事前調査と事後調査における動機の各種類の比較に対しては、フリードマンの検定をそれぞれ実施した。また、動機の各種類における事前調査と事後調査の比較に対しては、5つのウィルコクソンの符号付順位検定を行った。タイプIの過誤を避けるため、ボンフェローニ法による調整を行い、5% を組み合わせ数 (検定数7) で割り、有意水準を 0.7% と設定した。

また、効果量として、連関の強さを示す相関係数 r を計算した (Field, 2005, p. 541, p. 566)。Field は、 r の解釈のために、.10 (小さい効果量)、.30 (中程度の効果量)、.50 (大きい効果量) という基準値を示している (p. 32)。統計処理は、SPSS 12.0J を用いて行った。

結果

表4に、事前調査と事後調査の記述統計量がまとめられている。クロンバック α による信頼性係数が .60 よりも低かったのは、取り入的調整 (事前調査と事後調査) と同一視的調整 (事後調査) であった。外的調整 (事前調査と事後調査) は .63 と .70 でやや低かった。項目数が少なかったにもかかわらず、無動機 (事前調査と事後調査)、同一視的調整 (事前調査)、内発的動機 (事前調査と事後調査) においては、高い値 (.79 ~ .88) が得られ、内的一貫性が確認された。

図1は、事前調査と事後調査における動機の種類ごとの中央値を示している。事前調査においては、無動機の中央値が最も低く ($Mdn = 1.75$)、取り入的調整と外的調整が続いた。同一視的調整の中央値が最も高く ($Mdn = 3.75$)、内発的動機づけの中央値 ($Mdn = 3.13$) よりも高かった。事後調査においても、同じような傾向が見られた。すなわち、無動機の中央値 ($Mdn = 2.00$) が最も低く、同一視的調整の中央値 ($Mdn = 4.00$) が最も高かった。内発的動機づけの中央値 ($Mdn = 3.50$) は、同一視的調整よりも低かった。また、事前調査よりも事後調査の値が大きかった動機の種類は、無動機、同一視的調整、内発的動機であった。一方、事前調査よりも事後調査の値が小さかった動機の種類は、外的調整と取り入的調整であった。

フリードマンの検定の結果、事前調査と事後調査における動機の各種類の中央値には有意な差が見られた ($\chi^2 = 48.77, N = 44, p = .000$; $\chi^2 = 58.07, N = 44, p = .000$)。ウィルコクソンの符号付順位検定による事後比較を行った。ボンフェローニの調整により、事前調査と事後調査の有意水準をそれぞれ 0.07% に設定した (0.7% \div 10 組み合わせ)。表5は、事後比較の結果と、それぞれの組み合わせにおける効果量 (r) を示している。事前調査においては、無動機が他の4種類の動機と比べて有意に低く、また、同一視的調整は他の4種類の動機と比べて有意に高かった。これらの組み

表4. 事前調査・事後調査における各動機の基礎統計量 (N = 44)

	無動機	外的	取り入れ	同一視	内発的動機
<u>事前調査</u>					
平均	1.97	2.77	2.83	3.60	3.15
95% 信頼性区間					
下限	1.65	2.53	2.60	3.30	2.86
上限	2.28	3.01	3.06	3.91	3.44
標準偏差	1.03	0.79	0.76	1.01	0.96
中央値	1.75	3.00	2.83	3.75	3.13
四分位範囲	1.50	1.10	1.25	1.00	1.44
歪度	1.00	-0.29	-0.11	-0.42	-0.17
尖度	-0.01	-0.63	-0.54	-0.33	-0.46
信頼性係数 (α)	.87	.63	.50	.79	.88
<u>事後調査</u>					
平均	2.05	2.80	2.67	3.75	3.39
95% 信頼性区間					
下限	1.75	2.56	2.43	3.45	3.11
上限	2.35	3.04	2.92	4.05	3.67
標準偏差	0.99	0.80	0.81	0.98	0.92
中央値	2.00	2.80	2.67	4.00	3.50
四分位範囲	1.38	1.20	1.33	1.50	1.25
歪度	1.20	-0.37	-0.26	-0.57	-0.32
尖度	1.51	-0.65	-0.21	0.01	-0.06
信頼性係数 (α)	.79	.70	.53	.58	.82

注. 外的 = 外的調整; 取り入れ = 取り入乐的調整; 同一視 = 同一視的調整; 歪度の標準誤差 = 0.36; 尖度の標準誤差 = 0.70.

合わせの効果量は、大きかった (.53 ~ .66)。事後調査においては、無動機は、3種類の動機（外的調整、同一視的調整、内発的調整）と比べて有意に低く、また、同一視的調整は他の4種類の動機と比べて有意に高かった。さらに、内発的動機は、取り入乐的調整よりも有意に高かった。これらの組み合わせの効果量は、大きかった (.50 ~ .75)。まとめると、事前調査においても、事後調査においても、最も低かった動機の種類は無動機であり、最も高かった動機の種類は同一視的調整であったことが統計的にも支持された。また、事後調査では、取り入乐的調整が下がった一方で無動

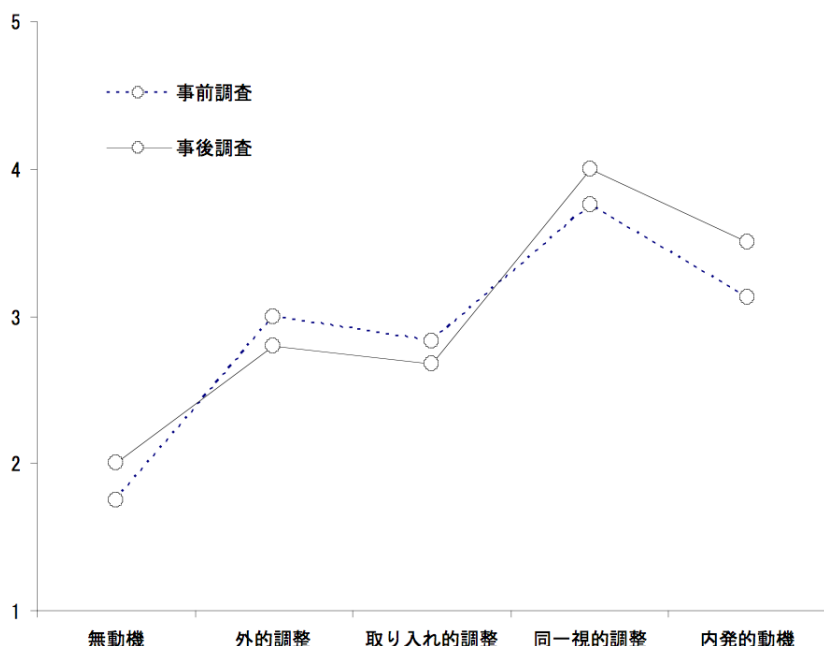


図1. 事前調査と事後調査の変化

機と内発的動機が上がったため、取り入れ的調整と無動機の間で有意な差が見られなくなり、また、取り入れ的調整と内発的動機の間で有意な差が見られるようになった。

次に、動機の各種類における事前調査と事後調査の比較に関して、ウィルコクソンの符号付順位検定を行った。表6は、その結果と効果量を示している。内発的動機だけに有意な差が見られ、効果量は中程度であった ($Z = -3.15, p = .001, r = .48$)。その他の動機の種類は、事前調査と事後調査に有意な差が見られなかった。無動機、取り入れ的調整、そして同一視的調整の効果量は小さかった (.12 ~ .26)。外的調整の効果量は .04 であり、小さい効果量の基準値よりも低かった。まとめると、事前調査と事後調査の間で統計的に有意な変化が見られたのは、内発的動機だけであった。

考察

まとめると、国際的イベントへのボランティアとしての参加が、本研究の参加者である日本人大学生の英語学習動機に影響を及ぼしていることが確認され、特に、内発的動機が事前調査よりも事後調査において統計的に有意に高まったことがわかった。つまり、英語そのものの有用性を認識したり、英語使用を通して英語や英語の

表5. フリードマン検定の事後比較 (ウィルコクソンの符号付順位検定)

動機の種類	1	2	3	4	5
<u>事前調査</u>					
1. 無動機	---	*	*	*	*
2. 外的調整	.63	---	ns	*	ns
3. 取り入れの調整	.56	.10	---	*	ns
4. 同一視的調整	.66	.54	.59	---	*
5. 内発的動機	.53	.27	.26	.60	---
<u>事後調査</u>					
1. 無動機	---	*	ns	*	*
2. 外的調整	.61	---	ns	*	ns
3. 取り入れの調整	.45	.19	---	*	*
4. 同一視的調整	.68	.61	.75	---	*
5. 内発的動機	.60	.42	.61	.50	---

注. 連関の強さとして、相関係数 r を対角下に示している。

* $p < .0007$

表6. 事前調査と事後調査の比較 (ウィルコクソンの符号付順位検定)

動機の種類	Z	p	r
無動機	-0.78	.448	.12
外的調整	-0.30	.771	.04
取り入れの調整	-1.72	.087	.26
同一視的調整	-1.57	.119	.24
内発的動機	-3.15	.001	.48

注. 連関の強さとして相関係数 r を示している。有意水準は $p < .007$ に設定されている。

学習そのものに興味が強まったことを示している。一方、統計的に有意ではなかったが、取り入れの調整による動機づけが事前調査から事後調査の間で弱まっていた。

本研究の結果は、事前調査と事後調査の間で英語学習動機が変容したことを示すものである。この結果は、動機づけを力動的に捉えようとする自己決定理論の考え方や、速水 (1995) や堀野・市川 (1997) の指摘を支持するものである。すなわち、動

機づけを静的な個人的特質と捉えるのではなく、変化しうる認知的状態と捉えることの妥当性を支持する結果であると考えられる。一方で、本研究の事後調査は、ボランティア活動終了後約2ヶ月経過してから実施されたことを考えると、本研究で見られた動機の変化は一時的なものではないことが示唆される。

本研究の結果によれば、短い期間であっても、国際的イベントへのボランティア参加が英語学習の動機づけを高めることに効果が見られた。表2によれば、参加日数に関する回答として、4日間が最も多く、前後の3日間及び5日間とあわせると、40人(90.9%)になる。国際的なイベントに数日間参加するだけで動機づけの変容に効果が見られたことは注目に値する。

本研究の質問票では、自己決定理論で想定している有能性、関係性、自律性の欲求の充足に関する質問項目を含めなかった。そのため、数日間の国際的イベントへのボランティア参加が動機づけの変容に効果を及ぼした理由について詳細な考察を加えることには限界がある。しかし、有能性の欲求について、英語使用時間の結果から述べたい。当初、実際に英語を使用する場面において、有能性への欲求が満たされ、その結果、動機づけが変容することを想定した。英語使用時間をみてみると、期間を通して1時間未満が最も多く(13人、29.5%)、2時間未満の人数を合わせると28名(63.6%)であり、4時間以上という回答は6人(13.6%)に過ぎない。つまり、ボランティアの期間中、学生の英語使用時間は必ずしも多くなかったといえる。したがって、自らが英語を用いるということよりも、英語が用いられている社会的環境にいたという経験が動機づけに影響を及ぼした可能性がある。すなわち、英語を使用する時間が少なかったことを考えると有能性への欲求の充足が単独で動機づけの変容をもたらしたとは考えるのは難しいと思われる。

終わりに

本研究の限界点を指摘しておきたい。まず、本研究の参加者は教育学部生であった。そのため、他の大学生に一般化できるとは限らないことに留意する必要がある。次に、本研究は、統制群を設けなかった。そのため、本研究で見られた変化は時間的な経過によるものという可能性もある。第3に、動機づけの質問票の項目数が、動機づけの種類によって異なる点が挙げられよう。特に、取り入的調整と同一視的調整の項目数が少なく、信頼性も低かった。第4に、SOへのボランティア参加が動機づけの変容をもたらした理由について、本研究では明らかにできなかった。最後に、SOのような国際的イベントへのボランティア参加をきっかけに高まった英語の学習動機によって学習者がどのような学習行動を取ったのかという点については、本研究の目的の範囲を超えるものであり、追究されなかった。しかし、堀野・市川(1997)が指摘するように、学習動機は学習方略の選択には強い影響を与え、さらに選択された学習方略と成績に強い相関が見られることを考えると、英語の学習動機と学習行動の関係を調べることは意義があると思われる。これらのことが明らかになれば、SOのような国際イベントへのボランティア参加をどのように英語学習プログラムに関わらせたり組み込んだりするとよいのかという示唆が得られると思われる。今後、これらの限界点を克服するために、研究の条件を変えたり、質問紙を修正しながら、追究する必要がある。

注

1. 英語論文の直接引用は、本論文の中で筆者による日本語訳である。
2. Hiromori (2006) によれば、学習者の熟達度によって、どの欲求が重要なかが異なることが示唆されている。本研究では、学習者の熟達度に関するデータを入力しなかったため、学習者の熟達度の要因は考察しなかった。しかし、SOへのボランティア参加は 3 つの基本的欲求をすべて満たす経験であると考えられるため、特定の熟達度にだけに影響を及ぼすものではないと考える。
3. 査読者により、「すべての質問が参加者が現在英語を学習しているという前提で質問されている」として、アンケートの内容的妥当性について指摘があった。質問項目の多くは「私が英語を学ぶ理由は」という文言で始まり、「私が英語を学んでいる理由は」ではない。回答時に英語を学んでいる参加者にとっては、「私が現在英語を学んでいる理由は」と解釈できる。一方で、回答時に英語を学んでいない参加者にとっては、「私が英語を学ぶとすれば、その理由は」と解釈できる。したがって、回答時に英語を学んでいない参加者であっても、質問票に回答することが可能であると考ええる。また、実際に英語学習の行動を伴っていない参加者の動機の状態を探ることは意義のあることであると考えている。しかし、査読者の指摘どおりの質問項目もあった（「私が英語を学ぶ理由は、英語が履修しなければならない授業科目だからである」と「これからもずっと英語を学び続けていきたい」）。そのような質問項目に対する回答は、分析から除外することにした。この点に関する査読者の指摘に感謝したい。
4. この点に関する査読者の指摘に感謝したい。

酒井英樹 (Sakai Hideki) は現在信州大学教育学部准教授である。主な研究領域は、第2言語習得におけるフィードバックの役割、注意配分、スピーキング・プロセスである。小池浩子 (Koike Hiroko) は現在信州大学教育学部准教授である。主な研究領域は、異文化間コミュニケーションである。

引用文献

- Dörnyei, Z. (1994). Motivation and motivating in the foreign language classroom. *The Modern Language Journal*, 78, 273-284.
- Dörnyei, Z. (2001). *Teaching and researching motivation*. Harlow, Essex: Pearson Education Limited.
- Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: Sage Publications.
- 速水敏彦. (1995). 外発と内発の間に位置する達成動機づけ. 『心理学評論』, 38, 171-193.
- 廣森友人. (2003a). 学習者の動機づけは何によって高まるのか—自己決定理論による高校生英語学習者の動機づけの検討. *JALT Journal*, 25, 173-186.
- 廣森友人. (2003b). 発達の視点に基づいた動機づけの検討: 高校生英語学習者の場合. *HELES (Hokkaido English Language Education Society) Journal*, 3, 71-81.

- 廣森友人. (2005). 外国語学習者の動機づけを高める3つの要因: 全体傾向と個人差の観点から. *JACET Bulletin*, 41, 37-50.
- Hiromori, T. (2006). The effects of educational intervention on L2 learners' motivational development. *JACET Bulletin*, 43, 1-14.
- 廣森友人・田中博晃. (2006). 英語学習における動機づけを高める授業実践: 自己決定理論の観点から. *Language Education & Technology*, 43, 111-126.
- Honda, K., & Sakyu, M. (2004). Multiple models of motivation for Japanese EFL learners – An investigation into concepts in different paradigms. *ARELE (Annual Review of English Language Education in Japan)*, 15, 41-51.
- 堀野緑・市川伸一. (1997). 高校生の英語学習における学習動機と学習方略. 『教員心理学研究』, 45, 141-147.
- Ikeno, O. (2003). Motivating and demotivating factors in foreign language learning: A preliminary investigation. *Ehime University Journal of English Education Research*, 2, 1-19.
- 鹿毛雅治. (1994). 内発的動機づけ研究の展望. *Japanese Journal of Educational Psychology*, 42, 345-359.
- 鹿毛雅治. (1995). 内発的動機づけと学習意欲の発達. 『心理学評論』, 38, 146-170.
- Kimura, Y., Nakata, Y., & Okumura, T. (2001). Language learning motivation of EFL learners in Japan: A cross-sectional analysis of various learning milieus. *JALT Journal*, 23(1), 47-68.
- 中田賀之. (2006). 「英語学習動機づけ」から「英語学習意欲」の研究への転換—研究対象領域、研究手法、研究目的の観点から—. *Language Education & Technology*, 43, 77-94.
- Noels, K. A., Pelletier, L. G., Clément, R., & Vallerand, R. J. (2000). Why are you learning a second language? Motivational orientations and self-determination theory. *Language Learning*, 50, 57-85.
- Ryan, R. M., & Deci, E. L. (2002). Overview of self-determination theory: An organismic dialectical perspective. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 3-33). New York: University of Rochester Press.

An Investigation into the Effect of Raw Scores in Determining Grades in a Public Examination of Writing

David Coniam

The Chinese University of Hong Kong

This article examines the effect on the grades assigned to test takers either directly through the use of raters' raw scores, or through the use of measures obtained through multifaceted Rasch measurement (MFRM). Using data from the Hong Kong 2005 public examination of writing, the current study examines how test takers' grades differ by comparing the results of grades from "lenient" raters against those of "severe" raters on the two systems for assigning grades—raw band scores and MFRM-derived scores. Examination of the results of a pair of raters indicates that the use of raw scores may produce widely different results from those obtained via MFRM, with test takers potentially disadvantaged by being rated by a severe rather than a lenient rater. In the Hong Kong English language public examination system from 2007 onwards, band scales are to be used extensively, as indeed they already are in many Asian countries. The article therefore concludes with a call for consideration to be given to how test takers' final grades may be derived from raw scores.

本研究は香港における公的試験のライティング・テストの採点に関する実証研究である。採点者の得点をそのまま使った場合と、多相ラッシュ・モデリング(MFRM)の得点を使った場合、成績の上でどのような違いがあるのかを調査したものである。香港で2005年度に実施された試験をデータとして使った。分析の結果、採点者の得点をそのまま使った場合には、より厳しい採点者によって受験者が不利を蒙る傾向があることがわかった。採点者の得点を使って最終成績をつける場合にはどうすればよいのかを論じて結論とした。

This article examines the use of raw scores obtained from the writing test of a public examination for Year 11 (the eleventh grade of schooling) test takers in Hong Kong. The current study draws on the methodology of Coniam (2005), who addressed an issue discussed by Weir (2005) on the notion of score validity, concerning the use of raw scores being an imperfect measure of test taker ability. Weir states "if FACETS is not being used in the evaluation of writing tests, I would want to know why not!" In the Coniam study, rater grade differentials on an oral test were investigated using novice raters. The current study extends the scope of the findings through data from a live Hong Kong public examination of writing using experienced raters.

With one major exception, rating scales are not a feature of English language public examination assessment in Hong Kong.¹ In the writing and oral public examinations, test takers are assessed using holistic, norm-referenced scales. As of 2007, the examination system in Hong Kong is, however, undergoing drastic changes in the English language elements (SCOLAR, 2003). This will involve the adoption of a standards-referenced, rather than a norm-referenced, approach to assessment, with scales and descriptors being used to rate test taker performance in English language examinations. In light of these changes to the Year 11 public examination, the purpose of the current study is to investigate how test takers' final grades differ depending on whether raw scores or Rasch-derived measures (Rasch, 1960) are used.

Raters and Raw Scores

In Hong Kong English language examinations, test takers' final grades are computed directly from raters' raw scores. While the latter may be adjusted for mean and standard deviation on the basis of correlations with other tests taken by the test takers, essentially the result is the raw score. The accuracy of the information obtained from raw scores has long been questioned, with the problems associated with their use having been discussed by a number of researchers. McNamara (1996, p. 122) presents a cogent discussion of some of the problems associated with the use of raw scores. Referring to studies by Linacre (1989) and Diederich, French, and Carlton (1961), he illustrates the variability in raw scores awarded to test takers, clearly stating that raw scores are "an unreliable guide to ability" (p. 118), and citing various reasons for this. He attributes, for example, variability in raters' assessment to a range of causes: rater (mis)interpretation of the rating scales and descriptors, rater freshness (or tiredness),

and interpersonal factors where raters respond positively or negatively (albeit unintentionally) to certain gender, race, or personality types. Research conducted by Hamp-Lyons (1989) suggests that raters respond to cultural differences in writing, which is in part attributable to their own cultural and experiential background. Vann, Lorenz, and Meyer (1991) relate raters' responses to their gender as well as their academic discipline. Vaughan (1991) illustrates how raters' reactions to different language features may result in essays being awarded different grades.

Linacre (1989) suggests that the above-mentioned issues (which may affect test taker performance) are *facets*, which can—or indeed should—be taken into account, and be modelled when assessing test takers in performance tests. This is especially the case with the latter type of test, where many more factors need to be considered. With fixed-response test items—for which a limited set of answers are possible—there are likely to be few extraneous factors to be taken account of.

Major changes to the system by which writing scripts are rated in Hong Kong Year 11 public examinations are imminent—one crucial change involving the move to using rating scales. Given this, using data from the 2005 Hong Kong Certificate of Education (HKCE) examination, the current study sets out to examine the use of raw scores in a writing test and how test takers' grades compare when rated by a severe as opposed to a lenient rater. The current study involves a comparison of the use of raw scores with scores derived from statistical procedures such as multifaceted Rasch measurement (where situational factors such as prompt difficulty or rater severity may be modelled and compensated for; see below) when calculating test takers' final grades.

To restate, the hypothesis being addressed in the current study is therefore that the use of raw scores may substantially disadvantage test takers who are rated by severe rather than lenient raters, with those test takers receiving lower final grades—a situation which in some examination situations may result in failure rather than success on a test.

The Hong Kong School and Examination System

Hong Kong's model of education, although currently undergoing substantial revision, is modelled on the British system. There are 6 years of primary school, and secondary school operates on a 5+2 model with students being banded, or streamed, on entry to secondary school. There are three broad bands of ability, with each band covering approximately 33% of the student ability range.

Hong Kong's major public examination is the Hong Kong Certificate of Education (HKCE) examination, administered by the Hong Kong Examinations and Assessment Authority (HKEAA) at the end of Secondary 5 (Grade 11). In 2005, the candidature for English language was 82,078 (Hong Kong Examinations and Assessment Authority [HKEAA], 2005). There are four papers in the English language HKCE—Writing; Reading; Oral; and Integrated Reading, Writing, and Listening. The HKCE Writing paper—the focus of the current study—offers test takers three prompts. They have to select one and are allowed 70 minutes in which to write in the region of 300 words. Overall grades awarded on the HKCE English language paper are A to C (credit), D and E (pass), F and U (Fail). Grade C and above are the crucial grades since the University of Cambridge accepts these as a GCSE level pass.²

Figure 1 (from the 2005 HKCE examination) presents the prompt around which discussion in the current article centres.

The Leisure and Cultural Services Department is planning to hold an international pop music festival in an open area very close to where you live. It has invited local residents to write letters expressing their views on the proposal.

Write a letter to the Department giving your opinion and explaining the benefits and/or problems of holding the festival. If you wish, you may refer to one or more of the following in your letter:

- noise levels
- entertainment value
- tourism
- large crowds
- hygiene and waste disposal
- possible performers
- opportunities for local musicians

Begin your letter, "Dear Officer," and sign it "R. Lee."

Figure 1. 2005 HKCE Writing Paper, Prompt 2

In the HKCE Writing paper, two raters assess each script independently with scripts currently pattern-marked on a single norm-referenced

9-point scale, with raters having to adhere to a specified pattern in terms of how many scripts can be allocated to a given point on the scale. Having to conform to a pattern mitigates, to an extent, the issue of severity since there are only so many high or low grades a rater may award. This changed, however, in 2007 when rating scales were adopted and raters were not constrained to a pattern.

Research Design

This section describes the data which made up the study and the methods used to analyse the data.

Data

The data used in the study were drawn from the live HKCE 2005 English language examination. Subsequent to the administration of the examination, 900 scripts (i.e., 300 scripts for each of the three prompts) were identified and extracted on the basis of the following three principles. First, that scripts should be drawn from the batches of markers with good statistics (i.e., good interrater reliability and a high correlation with other HKCE English language papers). Second, that scripts awarded the same grade by both markers should be selected since there would then be no differences between raters' raw scores. Third, that scripts selected should form a representative cross-section of ability across the whole candidature.

Nine markers then re-marked the three sets of 300 scripts. These were Hong Kong English teachers who had served as raters for the HKCE Writing paper for a number of years and had consistently achieved good rating statistics.

To prepare for the rating sessions, raters were first trained and their ratings standardised. After having familiarised themselves thoroughly with the new scales and descriptors, raters attended a training session where they rated a number of sample scripts illustrating different aspects of the scales and descriptors and a spread of ability across the 6-point scale. Because each script was double marked in line with standard HKEAA practice, 1,800 ratings were obtained for the current study; each rater assessed 200 scripts.

The subscales and descriptors used were those developed for the 2007 HKCE Writing Test (Note 2). The four subscales were:

1. Relevance and adequacy of content for purpose;

2. Accuracy and appropriacy of punctuation, vocabulary, language patterns;
3. Planning and organisation; and
4. Appropriacy of tone, style, and register; appropriacy of features for genre.

The subscales each had six levels, ranging from 1 (indicating the least able) to 6 (indicating the most able). For the subscales and descriptors, see HKEAA, 2007, pp. 104-105.

Methodology

As mentioned, the methodology in the current study involves a comparison of two composite scores. One of these was the rater's average of the four raw subscale scores. The second was obtained through multifaceted Rasch measurement (MFRM), a brief description of which will now be presented.

In classical measurement theory (CMT), test results cannot really be directly compared with one another. Consider for example, two Year 11 ESL classes. Last year's class scored 47% on their final exam; this year's class 43% on their (different) final exam. How are the two classes' scores to be compared? Is this year's class less able than last year's? Were the questions more difficult this year? Were the markers more severe in their judgements this year? We are not really in a position to answer any of these questions. Additionally, in CMT, test takers' results are not evenly spaced—despite the use of an apparently linear scale such as the percentage scale. Scores in the middle range are bunched together, while scores at the top and bottom end of the scale are disproportionately spread out (see Bond & Fox, 2007, pp. 24-26, for a cogent elaboration).

The use of the Rasch model enables all of these issues to be taken account of. First, in the standard Rasch model, the aim is to obtain a unified metric for measurement. This is not unlike measuring length using a ruler, with the units of measurement in Rasch analysis (referred to as logits) evenly spaced along the ruler. Logits are centered at zero, zero being the 50% probability represented by an "item" of average difficulty. Second, once a common metric is established for measuring different phenomena (test takers and test items being the most obvious), the phenomena can be examined and their effects controlled and compared. The result of using a Rasch model of measurement provides, in principle, independence from situational features (test takers, for example) in a particular test.

Consequently, results can be interpreted with a more general meaning. To return to the example in the above paragraph, test scores for different groups—such as last year’s and this year’s ESL classes—can be directly compared via Rasch measurement as the use of Rasch locates them on a single linear scale.

In MFRM, the measurement scale is based on the probability of occurrence of certain facets—in the current case, features associated with the rating of writing such as prompt difficulty, test taker ability, and rater severity levels. The phenomena or different situational factors can be explicitly taken into consideration and modelled in constructing the overall measurement picture.

While the focus in this study is on the rater, rater behaviour was examined in the context of the overall picture whereby it formed part of a three-faceted model of analysis, (i.e., raters, test takers, and prompts). The data presented in the paper is taken from the scores generated by the multifaceted Rasch analysis computer program FACETS (Linacre, 1994). In addition to logit measures, FACETS provides a “Fair Average” (see Linacre, 1997, p. 550, for details). The Fair Average is a more easily interpretable statistic in that logit values are converted back to the original rating scale, the 6-point scale in our case, rendering the output more easily interpretable by end-users. Because they are presented in the format of the original rating scale scores, the Fair Averages can be directly compared with the raters’ original raw scores. Such a comparison forms the cornerstone of the methodology in the current study.

For an accessible overview of MFRM, the manner in which it may be conducted, and its results interpreted, the reader is referred to Bond and Fox (2007).

Results and Discussion

The analysis in this section centers on an examination of the differences between test taker scores using the two different methods of arriving at a final score, (i.e., the average raw score of the four subscales compared with the FACETS-provided Fair Average). For illustrative purposes, results will be presented for one pair of raters only: the pair of raters who showed the widest degree of divergence in terms of severity levels.

First, however, Table 1 presents the results derived through MFRM for all the raters. In this Table, Column 5 presents the infit mean square statistic, which describes model fit, “fit” essentially being the difference between expected and observed scores. “Perfect fit” is defined as 1.0,

with an acceptable upper limit of fit stated as 1.3 (see Bond & Fox, 2007, pp. 285-286 for a discussion of limits of fit).

Table 1. Raters’ Results

Rater	Logit values	Fair Average	Model error	Infit mean square	Notes	
181	+2.19	2.53	0.06	0.98	most severe rater	paired with rater 142
123	+0.59	3.02	0.06	1.10		
106	+0.46	3.06	0.06	0.80		
171	+0.40	3.08	0.06	0.90		
183	-0.09	3.23	0.06	0.94		
110	-0.35	3.32	0.06	0.97		
142	-0.95	3.52	0.06	1.03		paired with rater 181
153	-1.03	3.55	0.06	0.90		
188	-1.23	3.62	0.06	1.13	most lenient rater	
Mean	0.00	3.22	0.06	0.97		
S.D.	1.01	0.32	0.00	0.10		

Note. RMSE .06 Adj (True) S.D. 1.01 Separation 16.18 Reliability 1.00 Fixed (all same) chi-square: 2409.9 d.f.: 8 significance (probability): .00

As infit mean square values in Table 1 indicate, all nine raters were well within acceptable degrees of fit. Looking at Column 2 (logit values), raters show quite a spread of severity, extending from the most severe rater at +2.19 logits (Rater 181), to the most lenient (Rater 188) at -1.23. The range of rater severity is consequently wide. Taking 3 standard errors as a delineator of a “statistically distinct” level (see Wright & Masters, 1982, p. 92), the Separation index of 16.18 indicates that raters are being separated into distinct levels of severity. The raters’ values in Column 2

are presented in logits. The Fair Averages are presented in Column 3 with logit values converted back to the original 6-point scale.

An analysis of the data will now be presented with regard to test takers' average band score compared with their Fair Average. The analysis presented centers on the pair of raters with the widest severity differential who rated the same test takers. These were Rater 181, with a measure of +2.19, and her partner Rater 142, with a measure of -0.95 logits. From the Fair Average scores in Table 1—where a lower score indicates a more severe rating—Rater 181 (whose Fair Average score is 2.53) can be seen to be one whole level more severe than Rater 142 (whose Fair Average score is 3.52). Their results are in bold type in Table 1 above.

It should be noted that Raters 181 and 142 co-rated 65 scripts: 22 test takers on Prompt 1, 27 on Prompt 2, and 16 on Prompt 3. The results and trends that emerge from the data hold good across all three prompts. To avoid overwhelming the reader with detail, however, only the largest data set (i.e., Prompt 2) is presented.

In Table 2 below, Column 2 provides the Fair Average. Two columns of data are then presented for each rater. The first column for each rater contains the rater's average raw band score from the four rating subscales; the second column presents the difference between the average raw band score and the Fair Average. A positive figure in a rater's second column indicates that the test taker would have received a higher (i.e., more lenient) score from that rater. A negative figure indicates a lower score, emerging from a more severe rating.

As can be seen from Table 2, the raters' tendency to severity or leniency is confirmed in the results that test takers would have received. Comparing the average raw scores against the Fair Averages, it can be seen that with Rater 181, all (100%) of her test takers would have received a lower grade. In contrast, only one (4%) of Rater 142's test takers would have received a lower grade, whereas 26 (96%) would have received a higher grade.

I would now like to explore further the extent to which the variation apparent in Table 2 above might be significant in determining a test taker's score on a test as a whole. As mentioned (in Note 1), band scales are only used on one test in Hong Kong—the English language teachers' Language Proficiency Assessment of Teachers (LPAT). On the test components that comprise the LPAT, test takers must reach level 3 of the 5-point scale on every scale, although they may still be awarded a pass with a 2.5 on one scale (Government of the Hong Kong Special Administrative Re-

Table 2. Prompt 2, Paired Raters 181 and 142

Test taker	Fair Average (FA)	Rater 181 (tendency to severity)		Rater 142 (tendency to leniency)	
		Average raw band score	Average raw band score minus FA	Average raw band score	Average raw band score minus FA
540	4.86	3.25	-1.61	6.00	+1.14
588	4.34	3.00	-1.34	5.25	+0.91
596	4.99	3.75	-1.24	5.75	+0.76
591	5.23	4.00	-1.23	6.00	+0.77
597	5.23	4.00	-1.23	6.00	+0.77
420	3.18	2.00	-1.18	4.00	+0.82
590	5.11	4.00	-1.11	5.75	+0.64
592	5.36	4.25	-1.11	6.00	+0.64
594	5.11	4.00	-1.11	5.75	+0.64
595	5.36	4.25	-1.11	6.00	+0.64
419	2.82	1.75	-1.07	3.50	+0.68
359	1.81	0.75	-1.06	2.50	+0.69
390	2.05	1.00	-1.05	2.75	+0.70
589	4.99	4.00	-0.99	5.50	+0.51
593	5.62	4.75	-0.87	6.00	+0.38
586	5.36	4.50	-0.86	5.75	+0.39
480	4.08	3.25	-0.83	4.50	+0.42
450	3.82	3.00	-0.82	4.25	+0.43
539	5.49	4.75	-0.74	5.75	+0.26
598	4.99	4.25	-0.74	5.25	+0.26
509	4.34	3.75	-0.59	4.50	+0.16
479	3.82	3.25	-0.57	4.00	+0.18
360	2.05	1.50	-0.55	2.25	+0.20
389	2.28	1.75	-0.53	2.50	+0.22
599	5.49	5.00	-0.49	5.50	+0.01
585	5.92	5.50	-0.42	6.00	+0.08
510	3.30	3.25	-0.05	3.00	-0.30

Note: (N=27)

gion, 2000). Obtaining two 2.5 level scores, or indeed any level 2 or lower score, results in an automatic failure grade being awarded on the LPAT. The writing test is regarded as one of the most demanding components of the LPAT (see Glenwright, 2002, for a discussion). The pass rate for the writing test is consistently one of the lowest across all five papers of the LPAT; in 2006, test takers achieved a proficiency attainment rate of 45.9% on this paper (HKEAA, 2006). Since the difference of half a band may therefore result in the difference between failing and passing, I would now like to explore this issue further.

Justification for half a band on the Hong Kong LPAT being taken as a determinant of “notable difference” can be seen to lie in the fact that the standard error of measurement (SEM) for the LPAT Writing Test is approximately 0.5 of a band (HKEAA, personal communication). This is comparable to the SEM for the IELTS Academic Writing module which, in 2005, was stated to be 0.37 of a band (IELTS, n.d.).

To underscore the importance of how half a band may stand as a “notable difference,” Table 3 provides a summary of the differences between the Fair Averages produced by MFRM and those produced from the two raters’ average raw band scores.

Table 3. Summary of Fair Averages Versus Raw Average Band Scores–Prompt 2

Leniency / Severity situation	Rater 181 (→ severity)	Rater 142 (→ leniency)
More lenient by half a band or more	0	14 (52%)
More lenient by less than half a band	0	12 (44%)
More severe by less than half a band	3 (11%)	1 (4%)
More severe by half a band or more	24 (89%)	0 (0%)
More lenient cases (total)	0 (0%)	26 (96%)
More severe cases (total)	27 (100%)	1 (4%)

(N=27)

As can be seen from Table 3, the two raters present almost a mirror image. If such results appeared on the LPAT, the consequences would be as follows. With half a band taken as criterial, 24 (89%) of Rater 181's test takers would have received a lower grade and potentially failed the LPAT whereas none of those rated by Rater 142 would have.

Conversely, for the test takers scored by Rater 142, 14 test takers (52%) would have been rated more than half a band higher, against none by Rater 181. If the half band score is crucial, over half of Rater 142's test takers might have been moved out of the potential failure zone, as against none of Rater 181's.

The implications of the differences between the two systems of rating are apparent: If a test taker were rated by a lenient rater such as Rater 142 as opposed to a severe rater such as Rater 181, the use of raw scores means that one test taker might "pass" the test while the other might well "fail." A discussion of this issue is provided by Coniam and Falvey (2001) in the context of the Hong Kong LPAT where simple raw scores are used to determine a final grade, and where a half-band score did, in certain cases, result in failure.

Conclusion

This study has examined the use of raw scores in the application of rating scales in the Hong Kong Certificate of Education (HKCE) 2005 Writing Test. The study has illustrated how the use of raw scores and measures derived through multifaceted Rasch measurement (MFRM) can produce markedly different results. The grades of two raters who assessed the same set of test takers were markedly different when the two methods of analysis were contrasted. Over half of the most lenient rater's test takers (52%) would have received a grade higher by half a band when this rater's raw scores were compared with MFRM-derived measures, with no test taker receiving a grade lower by half a band or more. In contrast, none of the most severe rater's test takers would have received a grade higher by half a band, although 89% would also have received a grade lower by half a band.

The current study has its limitations, however. The first of these lies in the fact that, to make its point, the study has been focusing on two extreme raters. An extension of the current study would possibly involve an examination of the "bigger picture" or how many test takers would have a different outcome (i.e., those who passed using raw scores, but failed using Rasch measures and vice versa) if MFRM had been used to

determine students' proficiency. It is in such a situation that the effects of rater variance really become apparent—when fair ratings are not provided and when students' lives are unfairly affected. The study has also focused essentially on test takers who are affected by severe raters and who are receive a lower grade than they may merit. The converse is also true: that using raw scores rather than Rasch measures awards some test takers higher grades than they deserve. Nonetheless, because more anguish is caused by test takers who fail when they should pass rather than vice versa, the focus in the current study is what it is.

Further, the current study has drawn on data from public examinations. Practical applications lie in the use of Rasch measurement (underpinned by an understanding of Rasch principles, Rasch, 1960) in school-based situations. However, convincing English language teachers of the value of certain statistics and getting them to use them represents something of a challenge. Popham (2006) comments, for example, on the temptation "to characterize any sort of test-related topic as 'too technical' for teachers" (p. 25). Nonetheless, it is achievable. The Centre for Assessment & Development (CARD) at the Hong Kong Institute of Education (<http://www.ied.edu.hk/card>) has a project running with about 100 primary and secondary schools in Hong Kong, with the objective of raising teachers' awareness of assessment targets and how targets may relate incrementally to other targets at higher levels. The project draws strongly on Rasch measurement principles, both as a technical tool as well as one that is delegated down to the teacher level with teachers using Rasch measurement in their evaluation of the tests they produce as an indicator of student progress and achievement. Between 2006 and 2007, CARD ran a total of 28 workshops (for more than 1,600 teachers) on Rasch measurement principles with hands-on practice in using Winsteps. Follow-up feedback indicated that individual teachers, and even whole schools, began to experiment with Rasch-based school assessment initiatives (see <http://www.ied.edu.hk/card>).

Given the results discussed in the current study, as Hong Kong moves towards adopting the use of scales and descriptors in rating test takers in its English language examinations when a standards-referenced approach to assessment is adopted in 2007, this issue of raw scores and the consequent disparity of results through rater severity is one which merits serious consideration. Given the fact that HKCE Grades A to C are recognised as a GCSE level pass, although a lower grade may not result in failure (as it can do with the LPAT), the consequence of being rated by a

severe rather than a lenient rater may make the difference between a test taker achieving a D rather than a C on the HKCE.

The use of band scales and descriptors are now the currently accepted method by which most speaking and writing tests are rated, with the practice being adopted in most countries across Asia. The implications from this small-scale Hong Kong study can therefore be extended beyond the Hong Kong context, and constitute an issue that needs to be considered by many educational and assessment bodies moving towards rating with scales and descriptors. The advantages of MFRM for analysing rating in speaking and writing tests are not new, as has been mentioned. The current study has attempted to underline the value of such a system in deriving test takers' results, suggesting that, if examination bodies adhere to a system whereby raw scores are the main determinant of a final grade, this may be doing test takers a disservice.

Acknowledgement

I would like to thank the Hong Kong Examinations and Assessment Authority for access to test takers' scripts from the 2005 HKCE English language examination.

David Coniam is a professor in the Faculty of Education at The Chinese University of Hong Kong, where he is a teacher educator, working with ESL teachers in Hong Kong secondary schools. His main publication and research interests are in language assessment, language teaching methodology, and computational and corpus linguistics.

Notes

¹ The one exception where scales and descriptors are currently used in a Hong Kong English language examination is the Language Proficiency Assessment of Teachers (LPAT) test for English language teachers (Government of the Hong Kong Special Administrative Region, 2000). The test consists of five papers. Of these, Speaking, Writing, and the Classroom Language Assessment test (a performance test conducted in a teacher's live classroom) are rated using scales and descriptors, with raw marks determining the final score. On the Speaking, Writing, and the Classroom Language Assessment test components, LPAT test takers must reach level 3 of the 5-point scale on every scale, although with a 2.5 on one scale one will still be awarded a pass (Government of the Hong Kong Special Ad-

ministrative Region, 2000). Obtaining any level 2 or lower score results in an automatic failure grade being awarded.

² Pass rates for the 2005 HKCEE were: Grades A–C: 10.5%; Grades A–E: 70.7% (Available at: http://www.hkeaa.edu.hk/doc/fd/2005cee/ceexamstat05_1.pdf).

References

- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental measurement in the human sciences*, (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Coniam, D. (2005). Raw scores as examination results: How far can they be relied upon? Paper presented at the Association of Language Testers in Europe Second International Conference, Berlin, 19-21 May 2005.
- Coniam, D., & Falvey, P. (2001). Awarding passes in the Language Proficiency Assessment of English Language Teachers: Different methods, varying outcomes. *Education Journal*, 29 (2), 23-35.
- Diederich, P. B., French, J. W., & Carlton, S. T. (1961). *Factors in the judgment of writing quality*. Princeton, NJ: Educational Testing Service.
- Glenwright, P. (2002). Language proficiency assessment for teachers: The effects of benchmarking on writing assessment in Hong Kong schools. *Assessing Writing*, 8 (2), 84-109.
- Government of the Hong Kong Special Administrative Region. (2000). *Language benchmark assessment for teachers—English language: Syllabus specifications explanatory notes, specimen questions with suggested answers, scales and descriptors*. Hong Kong: Government Printer.
- Hamp-Lyons, L. (1989). Second language writing: Assessment issues. In Kroll, B. (Ed.), *Second language writing* (pp. 69-87). Cambridge: Cambridge University Press.
- Hong Kong Examinations and Assessment Authority (HKEAA). (2005). *2005 HKCEE entry statistics*. Retrieved October 11, 2007, from http://www.hkeaa.edu.hk/doc/fd/2005cee/ceexamstat05_2.pdf
- Hong Kong Examinations and Assessment Authority (HKEAA). (2006). *Language Proficiency Assessment for Teachers (English language) 2006: Assessment report*. Retrieved October 11, 2007, from http://eant01.hkeaa.edu.hk/hkea/redirector.asp?p_direction=body&p_clickurl=otherexam%5Fbycategory%2Easp
- Hong Kong Examinations and Assessment Authority (HKEAA). (2007). *HKCEE English language examination report and question papers*. Hong Kong: Hong Kong Examinations and Assessment Authority.
- International English Language Testing Service (IELTS). (n.d.). *Test performance 2005*. Retrieved October 11, 2007, from <http://www.ielts.org/teachersandresearchers/analysisoftestdata/article234.aspx>

- Linacre, J. M. (1989). *Many-faceted Rasch measurement*. Chicago: MESA Press.
- Linacre, J. M. (1994). *FACETS: Rasch measurement computer program*. Chicago: MESA Press.
- Linacre, J. M. (1997). Communicating examinee measures as expected ratings. *Rasch Measurement Transactions*, 11(1), 550-551. Retrieved October 11, 2007, from <http://www.rasch.org/rmt/rmt11m.htm>
- McNamara, T. (1996). *Measuring second language performance*. New York: Longman.
- Popham, W. J. (2006). *Mastering assessment: A self-service system for educators*. New York: Routledge.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. The Danish Institute of Educational Research, Copenhagen.
- SCOLAR. (2003). *Action plan to raise language standards in Hong Kong*. Hong Kong: Government Printer.
- Vann, R., Lorenz, F., & Meyer, D. (1991). Error gravity: Faculty responses to errors in the written discourse of nonnative speakers of English. In Hamp-Lyons, L. (Ed.), *Assessing second language writing in academic contexts* (pp. 181-195). Norwood, NJ: Ablex.
- Vaughan, C. (1991). Holistic assessment: What goes on in the writer's mind? In Hamp-Lyons, L. (Ed.), *Assessing second language writing in academic contexts* (pp. 111-125). Norwood, NJ: Ablex.
- Weir, C. (2005, May). *A socio-cognitive approach to test validation*. Plenary presented at the Association of Language Testers in Europe Second International Conference, Berlin.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. MESA Press, Chicago.

A Comparative Analysis of the Japanese University Entrance *Senta Shiken* Based on a 25-Year Gap

Michael (Mike) Guest
University of Miyazaki

This study describes changes that have been made in Japan's National Center Examination for University Admissions (hereinafter *Senta Shiken*) by comparing the 1981 and 2006 versions of the test. An analytical outline of both tests was performed primarily with a top-down focus upon the categories of text type, topic and genre, task type, and skills required. Consideration was also given to the weighting of the various sections and tasks. The purpose of the study was not only to note new developments made in line with recent Japanese Ministry of Education, Culture, and Technology (MEXT) measures regarding university entrance exams and English education policy in general, but also to update the seminal work of Brown and Yamashita's (1995) analysis of Japanese university entrance exams and recommendations later made by Brown (2000). This study also briefly comments on positive washback that the *Senta Shiken* could have on high school English pedagogy in Japan.

本論文は1981年と2006年の大学入試センター試験を比較検討したものである。分析結果を基に、テキストの型、トピックとジャンル、タスクの型、測定対象となっている技能について変化しているのか否かを検証した。検証の主な目的は最近の大学入試の傾向を、英語教育全般に関する文科省の政策に照らして検討するものであった。さらにBrown & Yamashita (1995) による日本の大学入学試験に関するデータを再度検証した。

あわせて、Brown (2000)などで行われている提言の妥当性を再検討することも目的とした。さらにセンター試験が日本の高校英語教育に及ぼす好ましい影響についても簡単に触れた。

Background

Most readers are likely to be familiar with the centrally administered National Center Examination for University Admissions (hereinafter *Senta Shiken*). In the eyes of many this examination stands as the pinnacle of Japan's standardized education system. Each year over 500,000 examinees nationwide sit for this test, one which will have a great impact on determining which university exams, taken some weeks later at a second stage (*Niji*), they will have the best chance of succeeding in. Numbers will vary from individual university to university but many require that candidates obtain a certain score on the *Senta Shiken* before being allowed to sit for the *Niji* exam. Also, the *Senta Shiken* score is factored on a percentage basis (again varying by institution) into the total entry score set by any given university. However, recent changes have affected both the force and function of this exam. As both Mulvey (2001) and Mori (2002) note, Japan's low birthrate has now created a demographic in which there is nearly one seat for every university applicant, so competition is not as fierce nor as all-determining as it once was. Moreover, as mentioned by both Mori (2002) and Arai (1999), this has spawned alternative means of entering universities, such as recommendation systems, local quotas, and so forth. Trelfa (1998) adds that while there is educational standardization at the high school level, greater diversification and power given to local boards of education has weakened the uniformity of this standardization, such that the Japanese Ministry of Education, Culture, and Technology (MEXT) does not closely monitor the national guidelines at the local level, where boards of education have greater authority than ever to determine curricula.

Over the years, MEXT has implemented many changes in the *Senta Shiken* as well as recommendations made for individual university second-stage (*Niji*) exams—the exams which finally determine actual university entry (Monbukagakusho, 2000, 2003). A preliminary common entrance test (widely-known as the *Kyoutsuu Ichiji*) was originally established in 1979 under the leadership of the then Ministry of Education in order to create a unified, national standard that universities could use as a reference point. This was a result of a series of recommendations first made in 1971 by the Board of University Entrance Examination Improvement to alleviate the competition and stress that had hitherto surrounded the individual university entrance exams. In 1977, the National Center for University Entrance Examinations (*Daigaku Nyuushi Center*) was established to design and manage the *Kyoutsuu Ichiji*. One of the main criticisms of *Kyoutsuu Ichiji* was its inability to measure examinee abilities in analysis, creativity, and critical thinking (Kuroha, 1992). Hirezaki (1991)

added that the questions were too difficult, twisted, or obtuse for high school seniors.

Therefore, one of the main goals of the Interim Education Council (*Rinji Kyouiku Shingikai*) in 1984 was the improvement of university entrance examinations. This council was formed separately from the Ministry of Education as a quasi-private consulting body. In the Council's first report, in June, 1985, the establishment of another common test was recommended to replace the *Kyoutsuu Ichiji*. The new common test was proposed with characteristics different from the *Kyoutsuu Ichiji*, such as the improvement of multiple-choice task construction and alternative ways of scoring the exam. The new system began in 1991 under its current name *Daigaku Nyuushi Senta Shiken*.

Several revisions and guideline developments have taken place since. In both 1998 and 2000, MEXT set about reforming the content of the *Senta Shiken*—a process previously undertaken every ten years—with the goal of fostering general comprehension and analytical skills over and above those of memory or recognition. In 2003 the ministry announced its further intention to promote the development of “Japanese with communicative abilities” and recommended that individual university entrance exams try to reflect such a focus in terms of exam form and content. A listening component was established in 2005.

Research on the examination itself is surprisingly scant. References inevitably start with Brown and Yamashita's comprehensive (1995) review and critique of Japanese university entrance English exams, which has remained the seminal English work on the subject, even though the samples on that survey are now over a decade old. Since new approaches and directions have been implemented, one may ask whether these changes and implementations have moved the test away from its critical description, that of a poorly designed, discrete-item-based measure of grammatical minutiae and “testwiseness” made by “amateurs” (see Brown, 2002; Brown & Yamashita, 1995; and McVeigh, 2001), towards a test which comes closer to reflecting (or fostering) healthy pedagogical and educational practices.

Kikuchi (2006) updated Brown and Yamashita's research but focused only upon second-stage *Niji* examinations for prospective English majors at prestigious universities. Using the same categories of analysis as Brown and Yamashita, Kikuchi concluded that little has changed on these selected *Niji* exams, save for the emergence of, “a few new item types, such as summarizing reading passages or listening passages” (p. 90).

However, this study focuses solely on the *Senta Shiken*, since it is the only standardized nationwide English entrance exam and the most heavily weighted. It is developed through government agencies and therefore stands as a bellwether of national policy regarding English pedagogical content (more so than *Niji* exams which may reflect less uniform, more localized concerns and practices).

Ichige (2006) analyzed the 2005 *Senta Shiken*, concluding that “it can hardly be said that the current Center examination measures communicative ability appropriately” (p. 21). The validity of Ichige’s basic research rationale has been questioned by Guest (2007), who argued that the *Senta Shiken* was never meant to be a measure of communicative skills in the first place. Still, Ichige’s study is one of the very few attempts to measure how the *Senta Shiken* has changed in order to reflect new pedagogical trends or emphasis using an item-by-item analysis, a procedure also undertaken in the present study.

Brown (2000) followed up the 1995 research with a series of recommendations for improving the examination system. Among these were comments regarding economic and sociopolitical polity, which falls outside the concerns of this study, but these also included suggestions regarding test construct validity (several solicited from Hughes (1989)) such as:

- using a wide variety of samples,
- testing those abilities that one wants to develop and encourage,
- increasing the variety of examination formats,
- assessing higher order cognitive skills, and
- not limiting texts or tasks to academic fields.

While Guest (2006, 2007) has outlined many of the situational, environmental, and logistic factors limiting the scope of the *Senta Shiken* which render some of Brown’s original recommendations as impractical or implausible, there may well have been positive developments on the exam in the past several years that both reflect and foster healthy pedagogical approaches in the high school system based on such recommendations.

Thus, this research paper’s purpose is twofold. First, it seeks to analyze and compare a recent (2006) edition of the *Senta Shiken* with the 1981 version in order to highlight the developments that have taken place in the intervening years, especially in light of MEXT’s new focus. Secondly,

this paper seeks to confirm the realization of the above recommendations made by Brown (2000) and the subsequent possibility of positive washback onto high school English pedagogy created by changes in the exams instead of negative washback. Negative washback, in terms of fostering a grammar-translation methodology at the high school level, has been noted in past research (Bailey, 1999; Gorsuch, 1998). Mori (2002) further argues that a disjunction between high school and university entrance exam content "has essentially created the area of remedial education" (p. 42).

On the other hand, although many have questioned both the quality and quantity of washback (Guest 2000; Stout, 2003; Mulvey, 2001; Watanabe, 1996), the fact remains that perceptions of both the form and content of the *Senta Shiken*, whether accurate or not, still inform high school pedagogy and policy to a considerable degree.

Research Method and Design

For this study, the analytical template used in Brown and Yamashita's (1995) study has been avoided. This was a conscious decision made for several reasons. A judgmental rather than an empirical approach has been adopted because it is believed that an empirical approach does not do justice to analyzing or measuring certain important aspects of the *Senta Shiken*, nor is a purely empirical approach entirely sensitive to the environmental constraints that surround the test.

Here I must address the key points of test utility and construct validity. A test is considered valid only when the construct matches the stated purpose of the exam. But what is the purpose of the *Senta Shiken*? None has ever been publicly proposed but there is no doubting that the primary function of the English portion of the *Senta Shiken* is as a type of placement test, to a) determine student aptitude for academic study of English at the tertiary level in Japan, and b) to stratify examinees so that they make suitable choices in terms of deciding which second-stage university examinations to sit for. It is *not* an achievement test or a preparation test for utilizing English skills in real-world communicative settings.

Noting this about test utility and construct brings into question some of Brown and Yamashita's (1995) foci as well as those of Kikuchi (2006). A primary example can be found in their common concern with test difficulty and the resultant focus on utilizing readability scales. First, test difficulty as a whole can often be largely a matter of task difficulty and the type and number of skills demanded—areas that readability scales are

not able to measure. As a result, this analysis focuses more upon task type and skills required as categories.

Also, the designers of the *Senta Shiken* do not use any vocabulary or other content that is not mandated in the standardized high school curricula (Trelfa, 1998). Trelfa further explains that "scores...cluster nearer and nearer the top with each passing year and it becomes more difficult to discriminate among applicants. In order to distinguish among applicants, the Center Examination has had to create increasingly difficult questions. These questions are still based on the Monbusho curriculum, but involve, in the words of a test preparation manual, "the synthesis of several topics." In other words, because the *Senta Shiken* is norm-referenced, as long as examinees' scores can eventually allow them to be adequately stratified for placement in appropriate universities, "difficulty" might be said to be less of a factor in measuring construct validity.

This judgmental approach allows this study to question the categorization of certain discrete-point vs. integrative items. This is because although a question may be posed in multiple-choice format and thereby appear to be of the discrete-point variety, the skill required to complete the task correctly might well demand an integrative approach. Furthermore, there exists a large grey area as to what constitutes discrete-point vs. integration. Knowledge of an individual lexical item may be classifiable as "discrete-point" but knowledge of the function of said item within certain contexts would be "integrative." It seems that by taking a bottom-up approach in calculating such items both Brown and Yamashita's and Kikuchi's studies were thereby predisposed to categorize some integrative items as discrete-point.

This study also takes a *positive* view of utilizing a great variety of text-types, topics and genres, and tasks, a quality criticized by Brown and Yamashita. A wide variety of item types can measure a wider variety of skills, despite Brown and Yamashita's claims that such a test construct demands greater "testwiseness" on the part of examinees. And, in order to enhance reliability, texts, genres, and topics addressed in a well-rounded English test should not be reduced to single types or patterns which would favor a limited set of skills or particular types of learners.

In short, employing a top-down analysis, this judgmental study aims to focus upon the bigger picture. There is a greater concern for task type and genre, both of which influence the validity of an exam and yet are not widely treated in the background literature. This study also considers the skills demanded of the examinee, as opposed to the narrower concept of

item types (utilized by Brown & Yamashita, 1995), which tends to focus upon the layout of the question, rather than the actual task. The notion of “skills” refers to a variety of cognitive abilities which I believe are more accurate indicators of exactly what the tests are trying to measure. So, noting whether tasks ask examinees to predict, summarize, expound, extrapolate, interpret, infer, sequence, reconstruct, paraphrase, read for specific information, or read between the lines, none of which seem to have been specifically addressed in the previous analytical literature, is of greater interest.

By utilizing this different analytical design, we can gain a very different view of the current *Senta Shiken* in Japan, a perspective that is more sympathetic to the overall test purposes and functions—measuring Japanese students’ aptitudes for academic English in order to stratify candidates into appropriate universities. In taking this approach we do not intend to overturn the research of previous researchers but merely shed a different light on the exams, one which may provide us with a broader view to understand how positive changes have indeed appeared over the years.

The primary analysis of both exams will therefore be based upon four categories: a) text type, b) topic/genre, c) task type, and d) skills required. The scoring weight given to different items or sections, a crucial measure of test priorities, is also always noted.

Analysis of the 2006 *Senta Shiken* English Exam

This analysis of the 2006 exam is based on 2007 Kako Mondaisaku Tanki Kansei Ban available from Kyougakusha, Tokyo. (Note: The weight accorded to each section is represented as “pts.”)

Part 1 Reading: Total 200 pts. (Test total = 250 pts.)

Section 1: 16 pts.

- *Text:* Section a) individual sentences, b) casual conversation of nine turns
- *Topic/Genre:* a) none, b) casual conversation between friends regarding weather and umbrellas
- *Task type:* a) word accent (two questions), b) sentence stress (four questions).

- *Skills required:* a) knowledge of accent patterns, b) understanding of stress in verbal, interactive contexts

Section 2: 38 pts.

- *Text type:* a) various (sentence level), b) short four-turn conversations on various topics, c) various (sentence level)
- *Topic/Genre:* a-c) none consistent or specified
- *Task type:* a) discrete item slot and filler (10 questions), b) conversation slot and filler (three questions), c) sequencing/ordering at the sentence level (three questions double weighted)
- *Skills required:* a) discrete grammatical knowledge, b) norms of social interaction, c) syntax: sentence-level word order, d) specific lexical knowledge

Section 3: 34 pts.

- *Text type:* four paragraphs and one short essay of 150 words (seven questions total)
- *Topic/Genre:* expository magazine article type: image of elephants, throat microphones, gorilla behavior, the cause of allergies
- *Task type:* logical connector (discourse markers) slot and filler, reordering/sequencing
- *Skills required:* understanding rhetorical flow, understanding the role of logical connectors and transition phrases, coherently sequencing information

Section 4: 35 pts.

- *Text type:* 600-word essay plus graph (visual prompt)
- *Topics/Genre:* Expository magazine or journal essay: Attitudes toward volunteer work among different Japanese prefectures
- *Task type:* a) one heavily weighted question asking examinees to interpret an essay via a graph, b) general comprehension, rephrasing (four questions)
- *Skills required:* Applying information to visual prompt, interpreting data, paraphrasing

Section 5: 32 pts

- *Text type:* four-person extended conversation of 24 turns, with two large visual prompts
- *Topics/Genre:* casual conversation (contextualized) regarding description of a car, description of a task, and details on finding the car.
- *Task type:* a) slotting in common interactive phrases, b and c) applying information to two visual prompts, d) true / false
- *Skills required:* applying information to visual prompts, interpreting data, paraphrasing, understanding of conversational set phrases in context, understanding pragmatic force and uptake, displaying general comprehension

Section 6: 45 pts.

- *Text type:* one essay of about 600 words
- *Topics/Genre:* memoir, narrative on the author's change in perception of a neighbor
- *Task type:* a) five general comprehension questions, b) one three-part true / false question
- *Skills required:* holistic reading (general comprehension), summarizing, making inferences.

Part 2 Listening: Total 50 pts. (Test total = 250 pts.)**Section 1: 12 pts.**

- *Text type:* exchanges of two to four turns
- *Topics/Genre:* information on flight times, weather, planning a trip, and service encounters
- *Task type:* listening for specific information (three questions), listening to match visual prompts or illustrations (three questions)
- *Skills:* listening for specific information, making inferences from data, applying information to visual prompts

Section 2: 14 pts.

- *Text type*: seven short exchanges of two or three turns
- *Topics/Genres*: various
- *Task type*: listening in order to predict extended conversational responses
- *Skills required*: understanding of discourse flow, predicting

Section 3: 12 pts.

- *Text type*: a) three four-turn exchanges, b) one 10-turn extended conversation
- *Topics/Genres*: a) instruction, opinion, suggestion b) planning a ceremony
- *Task type*: a) three general comprehension questions, b) applying information to a chart
- *Skills required*: general comprehension, making inferences (pragmatic force and uptake), sequencing data to fit into visual prompt

Section 4: 12 pts.

- *Text type*: a) three short monologues: b) one extended monologue
- *Topics/Genres*: a) short formal speech, answering machine message, school announcement, b) narration on extreme weather experience
- *Task type*: general comprehension questions: a) one question for each of the three short monologues, b) three questions for the extended monologue.
- *Skills required*: a) listening for specific information, making deductions from data, b) holistic listening, paraphrasing

Additional Comments Regarding the 2006 Senta Shiken

Brown and Yamashita (1995) critically observed that *Senta Shiken* questions were all in multiple-choice format. However, the tasks demanded in these multiple-choice questions vary considerably and do not necessarily entail a discrete-point focus. Within a multiple-choice format, there is a

great difference between being asked (a) to correctly identify a discrete point from a text by “recognizing” it among the answers, and (b) choose the correct order of three sentences excerpted from that text as in the following:

#3B 1:

- a. While some people worry that a system like this will be able to read our minds, in its current stage it can understand only a few simple words.
 - b. The system is sensitive to pick up the “inner speech” we use when we are silently reading or thinking.
 - c. Now, NASA scientists have developed a more advanced system than throat microphones.
- Answer (choose the correct order) 1. A-B-C, 2. A-C-B, 3. B-A-C, 4. B-C-A, 5. C-A-B, 6. C-B-A

Likewise, the true / false questions are not simple binary questions that allow readers to skip reading the text and merely scan it for key word(s) in order to choose a T or F answer, but rather involve choosing one to three true or correct inferences, summarizations, or paraphrases from a list of six as in the following example:

#5d. Choose the two of the following that are true based on the text:

1. Owen’s sister has no children, but she keeps a teddy bear in her car.
2. Owen locked his sister’s car because it is not safe to park on the red level.
3. Since Owen’s car needs to be repaired, he borrowed his sister’s car that day.
4. The store where Owen works has plenty of space for employees’ things.
5. Jay, Yuki, and Ella will take Owen’s things to his sister’s car.
6. Owen did not buy very many things while he was shopping with his friends.

The current format for the reading section has been largely in place since 1998, as samples of older exams (*kakomon*) taken from the same source text reveal. *Kakomon* that are representative of the current test

format are usually provided in practice booklets. Very few questions or tasks from pre-1998 are listed in the *kakomon* sections. The various mock exams (*mogi shiken*) provided in practice booklets also conform to this post-1998 emphasis. The listening section was implemented in 2005.

No written English is required of examinees on this test (although so-called productive skills are required on most second-stage university tests). Given the massive number of candidates who take the Senta Shiken (again, over 500,000 in 2007 [Daily Yomiuri]), the requirement for absolute objectivity, and the demand for swift grading, the absence of productive tasks is not surprising.

Discussion of the 2006 Senta Shiken

The text-types and their weighting immediately reveal the 2006 test's priorities and focus. The reading section shifts gradually from a limited, discrete-item focus to extended, expository texts (word-sentence-paragraph-short essay). Another way of describing this is to say that the test moves from knowledge to skills to comprehension. It is noteworthy that in the reading section far more weight (almost 75%) is placed upon the more extended, comprehensive, integrated texts and tasks than upon discrete items and sentence-level or lower tasks (146 vs. 54 points). This is largely because the later extended-reading sections, with a large number of passage-dependent tasks, often have values of six or seven points per task/question as compared to 2 points for the majority of items in the discrete-item sections 1 and 2. This means that specific grammatical focus is evident in only about one-fifth of the test in total (and even within the section that falls under the rubric of "grammar" we can find a focus upon more communicative aspects of language, such as norms of social interaction and uptake). This makes the *Senta Shiken* decidedly not a "grammar test." The tasks on the 2006 test demand a variety of wide-ranging and comprehensive reading skills: making inferences, summarizing, recognizing themes, extrapolating information indirectly, knowing the functions of rhetorical signals and connectors, and applying sociopragmatic knowledge. In order to complete these varied tasks, comprehensive, holistic reading skills are required, as opposed to mere knowledge of English minutiae.

The topics and genres of the texts in 2006 are varied: journalistic essays on wide-ranging subjects, narratives, dialogues, opinion pieces, and so forth. Several include visual prompts. Having a wide variety of topics and genres increases the likelihood of the test content appealing to a

broader range of examinees' knowledge areas and interests and thereby increases the validity of the test by not catering to limited, specific knowledge or familiarity with certain content. Furthermore, the variety of task types provides a more accurate measure of wide-ranging English skills, with the added benefit that they would appeal to different learner types.

This variety of texts and tasks also applies to the listening section where we find monologues (speeches, narratives, announcements), dialogues (formal and casual, extended or short), and visual prompts, which demand a wide variety of skills such as listening for gist, listening for specific information, making deductions and inferences, and manipulating data-formats that might appeal to a wide variety of learner types.

Analysis of the 1981 Senta Shiken English Exam

A copy of the 1981 exam was obtained by special request from the National Library, Nagatacho, Tokyo. The 1981 test is divided into eight sections, all reading-based. The total Again, value is 200 points.

Section 1: 50 points

- *Text type*: 25 decontextualized sentences
- *Topics/Genre*: none identifiable
- *Task type*: fill-in-the-blank from multiple choices of four (25 questions)
- *Skills required*: discrete grammatical and/or lexical knowledge

Section 2: 15 pts.

- *Text type*: five short exchanges (two or three turns each)
- *Topics/Genre*: general conversation (various)
- *Task type*: fill-in-the-blank at the sentence/phrasal level; one question for each exchange
- *Skills required*: understanding pragmatic force, prediction, interactive and conversational cohesion

Section 3: 11 pts.

- *Text type*: single words
- *Topics/Genre*: none

- *Task type*: spot the different item from a multiple choice selection of four (11 questions)
- *Skills required*: knowledge of pronunciation at a phonemic level

Section 4: 24 pts.

- *Text type*: six decontextualized declarative sentences
- *Topics/Genre*: none identifiable
- *Task type*: sequencing, ordering
- *Skills required*: word order (at a syntactical level)

Section 5: 12 pts.

- *Text type*: one short conversation / one short essay
- *Topics/Genre*: personal conversation / education in England (expository article)
- *Task type*: reorder sentences into cohesive paragraphs (one weighty question in each paragraph)
- *Skills required*: organizing discourse, making predictions, understanding rhetorical connections

Section 6: 27 pts.

- *Text type*: three short paragraphs on very different topics in a magazine article style
- *Topics/Genre*: a) rising early (narrative), b) satellites, c) freedom of speech (expository paragraphs)
- *Task type*: a) paraphrasing proverbs (2 questions), b and c) true / false (one of 5five) summarizing questions
- *Skills required*: proverb knowledge, summarizing

Section 7: 21 pts.

- *Text type*: very short article of about 200 words
- *Topics/Genre*: birdsong (magazine / journal form)
- *Task type*: paraphrasing; summarizing
- *Skills required*: understanding rhetorical connections, paraphrasing, summarizing

Section 8: 40 pts.

- *Text type:* paragraph of about 250 words
- *Topics/Genre:* expository journal article on reading habits
- *Task type:* a) four true/false questions (one of five paraphrased statements based on the essay is true) b) one two-part question on summarizing the essay theme.
- *Skills required:* paraphrasing, summarizing

Discussion of the 1981 Senta Shiken

In the 1981 exam's extended reading and/or expository sections, the tasks invariably require examinees to paraphrase underlined passages or to look for specific information. This means that rather than comprehending rhetorical flow, cohesion, or the development of themes, one often needs only to read the area that contains the key phrase or word in order to answer the question (passage-independent). Therefore, many questions in 1981 that nominally appear to be integrative and holistic actually have a discrete-item focus.

The extended/expository reading texts in 1981 are also notable in that they seem to reflect the interests and reading habits (actual or idealized) of the test-makers themselves, invariably rather formalized magazine journal-based passages which often appear to be extracted from larger original texts. Having such a narrow content or stylistic focus is not evident on the 2006 exam.

It is also readily apparent that several tasks on the 1981 exam test knowledge of specific proverbs (Q# 6.1), and obtuse expressions. Q# 6.3's key phrase (referring to freedom of speech) reads:

"...the way to its attainment has lain through lakes of blood"

Some require specific *cultural* knowledge:

Q# 1.12 What they call "the first floor" in America is called "the _____ floor" in England.

1. base 2. ground 3. primary 4. second

Some employ rather stilted forms:

Q# 1.9 He knows little of math, _____ of chemistry.

1. as well as 2. still less 3. no less than 4. still more

Some texts lack clear contexts, which effectively invalidates the tasks:

Q# 2.2 *"Why don't you cut your cake in half?"*

"_____ . If I cut it in half I won't be able to tell which is the biggest."

1. *Because it is too big for me*
2. *All right, I'll cut it in half*
3. *Because I'd like to eat only a half of it*
4. *Because I'd like to eat the biggest piece last*

Many language forms are rather arcane ("*little boy*" being used as an address form), with some exchanges awkwardly contrived ("*I tore them up*"), both as noted in Q# 2.1:

Q# 2.1 *"I suppose you didn't see some papers on the table in my room, little boy?"*

"_____ . "What! You must be taught once and for all not to touch what does not belong to you."

- | | |
|---------------------------|-------------------------------------|
| 1. <i>Yes, I did.</i> | 2. <i>No, I didn't.</i> |
| 3. <i>I tore them up.</i> | 4. <i>I left them as they were.</i> |

The 1981 test also includes numerous poorly designed distractors (such that the correct answer can be deduced without referring to the actual text at all). No contextually barren passages nor such arcane/obtuse language were found on the 2006 test.

Earlier we mentioned that "difficulty" in terms of reading scales would not be a factor in this study. However, if one considers that arcane language forms, decontextualized passages, and unnatural interactive speech (qualities that standard reading difficulty scales often fail to measure) compound difficulty, then the 1981 has "difficulty" in spades.

Comparison of the 1981 and 2006 Senta Shiken

Several differences between the two tests become immediately evident. Most salient among these are:

- The lack of a listening component in 1981.
- On the 1981 test, 85 points (42% of the total) are given to questions that demand discrete, decontextualized knowl-

edge or skills at the sentence level or lower. Compare this to 54 points (21.6%) in the 2006 exam.

- The 1981 exam also shows much less variation in terms of text type. Five of the six expository texts in 1981 are of the magazine / journal article format, unlike the extended conversations, dialogues, narratives, visual prompts, and so forth encountered in the 2006 test.
- The 2006 texts are much longer. Whereas the 1981 exam has only one text of 250 words or more, the 2006 exam contains four, three of them nearing 500 words. This allows for greater contextualization and demands more holistic and integrative reading skills.
- While in the 2006 discrete-item sections there is a marked concern with features of coherence, such as logical connectors and discourse markers, the same sections in 1981 focus upon narrower units of discourse, such as prepositions and phrasal verbs.
- Of the 22 questions asked in the essay sections of the 2006 exam only four (17%) can be reasonably said to be answerable without comprehending wider contexts (largely passage independent). Of the 13 questions tied to similar texts in 1981, eight (63%) can be answered by reading only a small section of the text (largely passage dependent).
- The extended / expository texts in 1981 average only 2.4 questions per text (12 questions for five texts). The 2006 test not only averages 3.8 questions per text (15 questions for four main texts) but a greater variety of questions are asked, demanding a greater variety of reading skills—generally more holistic and integrative. This means that there is more “wasted” text in the 1981 test (section eight being the exception).

None of this is to say that the 1981 test was completely bereft of positive qualities, but when compared proportionately to the 2006 exam, the construct validity is palpably lower.

Conclusions

By many standard measures of test validity (utilizing a wide variety of texts, a multidimensional task focus, an emphasis upon higher level cognitive skills, fewer cases of arcane, decontextualized, obtuse and narrow-focus texts and tasks) the 2006 test is clearer the superior test. Moreover, the skills addressed on the test are also indicative of sound reading and listening pedagogy and appeal to a wide range of learner types and learning strategies.

In addition, the 2006 test realizes many of Brown's (2000) recommendations. For instance, a) there is a wide variety of samples and types; b) it tests those abilities that one would like to develop and encourage—that is, reading and listening for higher-level meanings and comprehension, rather than focusing on discrete items which would encourage a narrow item-by-item, bottom-up translation approach; c) there is a great variety of examination formats; d) the test puts a higher emphasis on assessing higher order cognitive skills; and e) the test's texts and tasks are not at all limited to academic fields.

If there is a washback effect from the *Senta Shiken* it should be a positive one. In order to enable them to succeed on the exam, students should be made to practice and develop a number of differing skills such as predicting, summarizing, expounding, extrapolating, interpreting, sequencing, reconstructing, paraphrasing, reading and listening for specific information, and reading and listening between the lines, all using texts of a variety of genres (narratives, information transactions, casual dialogues, scientific essays, personal essays, etc.). The listening section also demands a variety of developmental listening strategies since tasks and the skills required vary from making inferences to listening for gist and/or listening for specific information.

The 1981 *Senta Shiken* does conform to the critical view of a narrow, grammar-based, discrete-item test which does not adequately address more progressive or holistic educational strategies. However, this stands in sharp contrast to our findings regarding the 2006 exam. The various revisions and changes in MEXT's policies and practice regarding entrance exams do seem to be bearing some fruit and are not inconsistent with its broader educational aims nor with sound pedagogical and testing theory.

Michael (Mike) Guest is Associate Professor of English in the Medical School of Miyazaki University. He also serves as a section editor for the ETJ (English Teachers in Japan) Journal and is a regular columnist on EFL teaching for the national Daily Yomiuri newspaper. His interests include lexis, culture, and critical thinking, as well as testing.

References

- Arai, K. (1999). Articulation to shite no daigaku nyushi [University entrance examination as a way of articulation]. *IDE–Gendai no Koutou Kyouiku [IDE–Higher Education in Modern Society]*, 408, 13–19.
- Bailey, K. M. (1999). *Washback in language testing* (TOEFL Monograph Series No. 15). Princeton, NJ: Educational Testing Service
- Brown, J. D. (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. D. (2002). English language entrance examinations: A progress report. *Proceedings of the 1st Annual JALT Pan-SIG Conference*, (pp. 94–104.).
- Brown, J. D., & Yamashita, S. (1995). English language entrance exams at Japanese universities: What do we know about them? *JALT Journal*, 17 (1), 7–30.
- The Daily Yomiuri*. (2007, Jan. 20). p. 2
- Gorsuch, G. (1998). Yakudoku EFL instruction in two Japanese high school classrooms: An exploratory study. *JALT Journal*, 20 (1), 6–32.
- Guest, M. (2000). ...““But I have to teach grammar!”: An analysis of the role “grammar” plays in Japanese university English entrance examinations. *The Language Teacher*, 24 (11), 23–29.
- Guest, M. (2006). Teaching progressively...for the *Senta Shiken!* In K. Bradford-Watts, C. Ikeguchi, & M. Swanson (Eds.), *JALT 2005 Conference Proceedings*, 1182–1191 [CD-ROM]. Tokyo: JALT.
- Guest, M. (2007). A Response to Yoko Ichige’s “Validity of Center Examinations for Assessment of Communicative Ability”. *On Cue*, 15 (1), 28–32.
- Hirezaki, K. (1991). *Daigaku nyuushi–Kaikaku no kokoromi [University entrance examination: Attempts at its improvement]*. Tokyo: Iwanami Booklet.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge University Press.
- Ichige, Y. (2006). Validity of center examinations for assessment of communicative ability. *On Cue*, 14 (2), 13–22.
- Kikuchi, K. (2006). Revisiting English entrance examinations at Japanese universities after a decade. *JALT Journal*, 28 (1), 77–96.
- Kuroha, R. (1992). 1960 Nendai ikou no daigaku seisaku–Sono taiken teki seiri to kentou [Policy of higher education after the 1960s: Its pigeonholing and review through personal experience]. *Daigaku Kenkyu [University Studies]*, 81–92.

- McVeigh, B. (2001). Higher education, apathy and post-meritocracy. *The Language Teacher*, 25 (10), 29-32.
- Monbukagakusho (2000). *Daigaku nyuushi no kaizen ni tsuite* [How to improve the university entrance exams]. Retrieved August 21, 2007, from <http://www.mext.go.jp/b?Qmenu/shingi/12/daigaku/toushin/001102.html>
- Monbukagakusho (2003). *The course of study for foreign languages*. Retrieved August 21, 2007, from <http://www.mext.go.jp/english/shotou/index.htm>
- Mori, R. (2002). Entrance examinations and remedial education in Japanese higher education. *Higher Education*, 43, 27-42.
- Mulvey, B. (2001). The role and influence of Japan's university entrance exams: A reassessment. *The Language Teacher*, 25 (7), 11-15.
- Stout, M. (2003). Not guilty as charged: Do the university entrance exams in Japan affect what is taught? *ETJ Journal*, 4 (1), 1-7.
- Trelfa, D. (1998). The development and implementation of education standards in Japan. In *The Educational System in Japan: Case Study Findings*. Retrieved August 21, 2007, from <http://www.ed.gov/pubs/JapanCaseStudy/chapter2d.html>
- Watanabe, Y. (1996). Does grammar translation come from the entrance exams? Preliminary findings from classroom-based research. *Language Testing*, 13 (318-333).

Tracking and Targeting: Investigating Item Performance on the English Section of a University Entrance Examination over a 4-Year Period

Christopher Weaver

Toyo University

Yoko Sato

Tokyo University of Agriculture and Technology

This empirical study introduces population targeting and cut point targeting as a systematic approach to evaluating the performance of items in the English section of university entrance examinations. Using Rasch measurement theory, we found that the item difficulty and the types of items in a series of national university entrance examinations varied considerably over a 4-year period. However, there was progress towards improved test performance in terms of an increased number of items assessing different language skills and content areas as well as an increased number targeting test takers' knowledge of English. This study also found that productive items rather than receptive items better targeted test takers' overall knowledge of English. Moreover, productive items were more consistently located around the probable cut point for university admissions. The paper concludes with a detailed account of a number of probable factors that could influence item performance, such as the use of rating scales.

本研究は経験主義的立場に立ち、大学入試の英語の問題に用いられたテスト項目がよく機能したかどうかについて評価するための体系的アプローチとして、テスト項目が「母集団に的を絞れたか」、また「選抜ラインに的を絞れたか」という側面に注目する方法を導入する。ある国立大学の英語の入試問題についてラッシュ分析を行った結果、テスト項目の「項目困難度」と「項目の種類」は、四年間を通してかなり多様であった。しかし、多様な言語技能と内容領域を査定する項目の増加、受験者の英語力に的を絞った項目の増加という点において、テスト項目の機能性は改善の方向に向かっていた。また受容的能力(理解力)を問う項目に比べ、産出的能力(表現力)を問う項目の方が、受験者の総合的英語力の測定により的を絞ったテスト項目になっており、さらには入学者の選抜ラインと想定されるあたりに、より集中的に配置されていた。結論では、「評価(評定)尺度」の使用等、テスト項目の機能性に影響すると考えられる要因についても詳述する。

One of the most significant challenges facing university administrators and writers of the English section of any university entrance examination is how to ensure that the difficulty of the test items does not significantly vary from year to year. Guidelines issued by the Ministry of Education, Culture, Sports, Science, and Technology along with ministry-approved textbooks inform test writers about the type of English knowledge test takers should have mastered in junior and senior high school. Yet, the actual performance of entrance examination items designed to assess test takers' level of English knowledge remains largely unexamined. Although the English section has had a long-standing role in university admission policies, there have been relatively few empirical studies (e.g., Brown & Yamashita, 1995a, 1995b; Ito, 2005; Kikuchi, 2006) investigating its performance. This paper aims not only to contribute to this important area of second language assessment, but also to introduce a systematic approach to monitoring item difficulty that takes into consideration some of the special circumstances surrounding university entrance examinations in Japan.

Strategies to Monitor Item Difficulty

There are a number of ways in which item difficulty can be monitored. Often large-scale proficiency examinations such as TOEIC® and TOEFL® use item trialing. This technique involves adding a set of items to an examination not to assess test takers' level of English knowledge, but rather to determine the level of difficulty that test takers have with these items. Unfortunately, item trialing is usually not possible with university entrance examinations because on the same day examinations are given for a number of different subjects in addition to English, which in turn limits the number of test items that can appear in any one section. Adding a set of trial items would thus seriously reduce the number of items available to determine test takers' level of ability. Moreover, many universities publicize their entrance examinations and commercial publishers sell numerous books explaining previous examinations item by item. These materials become primary study materials for many prep schools and test takers. As a result, the function of trialed items when used in a future entrance examination may be reduced to simply assessing test takers' memorization skills. The combined effect of these factors thus prevents item trialing from being a practical means of monitoring item difficulty.

Conducting a small-scale trial before the actual administration of the examination is another means of determining item difficulty. This strategem involves recruiting a group of test takers, purportedly rep-

representative of the larger test taker population, to take the examination. Their responses would then provide test writers with estimates of item difficulty so that any needed adjustments could be made before the actual administration of the examination. Test security, however, renders this scheme a virtual nonstarter for many universities.

Another technique involves using a core set of items that reoccur on two different examinations. Using Rasch measurement theory, the difficulty estimates for this common set of items would anchor the estimates of difficulty for the remaining items (Wolfe, 2000). This approach also allows test writers to examine the degree to which item difficulty varies across the different examinations. Unfortunately, many of the same challenges that prohibit the use of item trialing also prevent the reuse of a core set of items.

Targeting Item Difficulty to Test Takers' Ability Levels

Targeting is an approach that evaluates an entrance examination according to the degree to which the difficulty of the test items overlaps with the test takers' level of ability. The amount of overlap between item difficulty and test taker ability can be determined using the graphical output from a Rasch analysis, commonly referred to as a Wright map (Wright & Stone, 1979). This graphical output is valuable because test takers' level of ability and test items' level of difficulty are placed upon the same scale of reference measured in logits. In order to provide a clear explanation of targeting, a simulated data set is used to illustrate what a poorly targeted examination looks like (see Figure 1).

Considering that many readers may be unfamiliar with Wright maps, a short explanation of how to interpret this graphical output is in order. In the middle of Figure 1, there is the logit scale with its values indicated on the far left side of the figure. Once again, logits define the common scale of reference regarding the test takers and the items on the examination. For this simulated data set, the logit scale starts at -1 logits and ends at 1 logit. By itself, a logit simply indicates the relative frequency of success over the relative frequency of failure (Smith, 2000). The logit scale thus needs a point of reference to become meaningful. The meaning of Figure 1 begins with the performance of the test takers on the examination. The resulting estimates of each test taker's ability, represented with a # sign, are shown on the left side of Figure 1. Ability in the context of this investigation is the test takers' knowledge of English as defined by the items on the English section of a university entrance examination. Test

takers' level of ability ranges from 0 logits to 1 logit. In other words, test takers located around 0 logits have less English knowledge than those located around 1 logit. The mean level of ability for these test takers is 0.5 logits, signified with the M marker. The S markers represent one standard distribution above and below the mean; while the T markers represent two standard distributions above and below the mean.

The right side of Figure 1 provides the second source of meaning for the logit scale. The different items on the examination, represented with a *, are placed along the scale according to their level of difficulty. For this simulated set of examination items, the range of difficulty starts at -0.08 logits and continues to 0.08 logits. Items located around -0.08 logits are less difficult, whereas items located around 0.08 logits are more difficult. The mean level of difficulty is 0 logits. Since the performance of the items is of primary interest, the standard practice is to set the starting point of the logit scale, 0 logits, at the mean for item difficulty. Once again the M, S, and T markers represent the mean for item difficulty, and the different standard distributions above and below the mean.

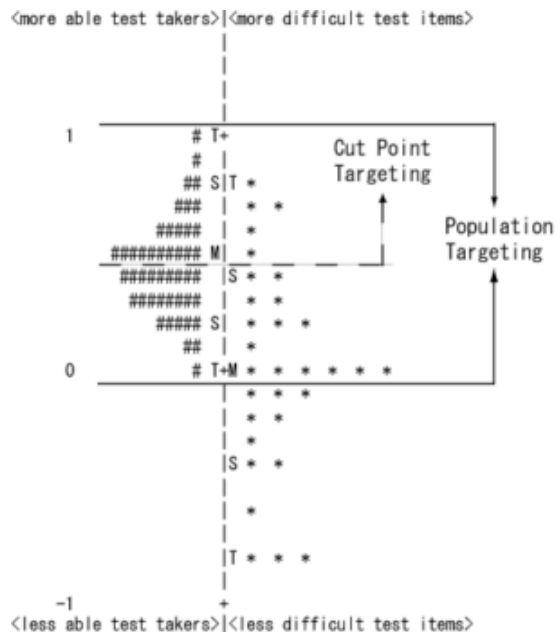


Figure 1. Wright map of a poorly targeted examination (simulated data)

Since the estimates of test takers' ability and the estimates of the item difficulty are placed upon the same scale, it is possible to compare the two directly. For example in Figure 1, a comparison of the mean for the test takers and the mean for the items produces a difference of 0.5 logits, revealing that the examination was very easy for most of the test takers. Another indicator of the ease of this examination is that the mean for item difficulty is located two standard distributions below the mean for test takers' ability. In other words, test takers with a level of English knowledge two standard distributions below the mean had a 50 percent chance of correctly answering almost half of the questions on the examination.

This imbalance between estimates of test taker ability and item difficulty does have an important implication. One's location on the logit scale is dependent upon the location of the test items. As a result, the estimate of test taker ability is more accurate when there are items in close proximity to that point on the logit scale. One of the advantages of Rasch measurement is that it provides an estimate of measurement error for every test taker and test item (Smith, 2001). Table 1 shows that for this poorly targeted examination, measurement error increases for test takers located at the higher ability levels. For example, the measurement error is four times higher for test takers located around 1 logit (0.24) than those located around 0 logits (0.06).

Table 1. Simulated test taker ability estimates and their accompanying estimates of measurement error

Test taker ability estimate	Standard error
1.00	0.24
0.86	0.20
0.78	0.19
0.68	0.17
0.58	0.15
0.40	0.12
0.34	0.11
0.22	0.09
0.13	0.08
0.00	0.06

One way to reduce the amount of measurement error is to increase the number of items that fall within the range of the test takers' ability. This concept is called targeting. By targeting the difficulty of items at the ability of the test takers, each item can provide the greatest amount of information. When test information is maximized, measurement error is minimized (Gershon, 2006). In the context of university entrance examinations, there are two types of targeting worthy of consideration. The first type involves targeting all who sit the examination. The focus here is having at least one test item located at each of the different ability levels of the test takers. This type of coverage ensures that the entire continuum of English knowledge is well defined and there is at least one item on the examination that test takers have a 50 percent chance of correctly answering. This type of coverage is called *population targeting*. Referring back to Figure 1, the population targeting is poor because too many are located at the lower levels of test takers' abilities in addition to 12 items that are below the ability level of any test taker. On the opposite end of the continuum, there are no items located around test takers who have an ability level one standard deviation above the mean. As a result, the exact location of these test takers is uncertain. Figure 2, in contrast, illustrates how items on an entrance examination can provide much better coverage of the test takers' abilities in an ideal situation.

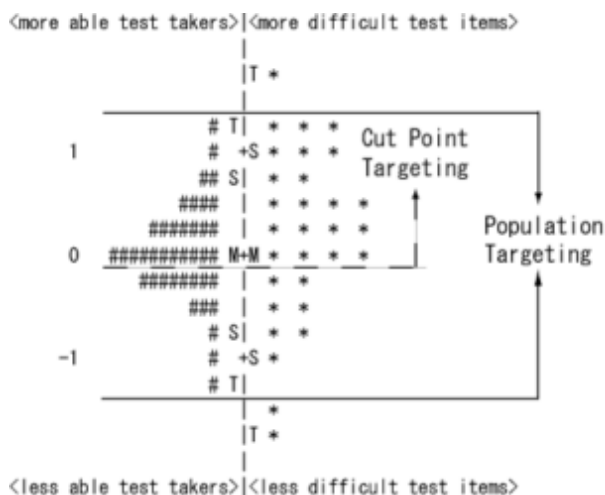


Figure 2. Wright map of a well targeted examination

The second type of targeting is of vital importance for entrance examinations. These types of tests are primarily used to make admission decisions, which in turn rely upon cut points to decide which test takers will be given an opportunity to attend the university. Such a decision becomes quite problematic when an examination has few or no test items located at the higher ability levels as shown in Figure 1. Considering the measurement error associated with poorly targeted examinations, it becomes imperative to have a group of items located around the probable cut point for admission decisions. This type of coverage is called *cut-point targeting*. The cut-point targeting in Figure 2 is a vast improvement over the population targeting, with 20 test items located around the probable cut point compared to the 5 items in Figure 1. Thus, the focus of cut-point targeting is to have a group of test items located around the probable cut point so that the items can accurately define test takers' level of English knowledge for the purpose of admission.

The ideal number or percentage of items located around the probable cut point depends upon the specific assessment needs of the university. For the purposes of the present investigation, the range of abilities where the probable cut point may fall is quite large, starting at the mean ability for test takers and extending to the most able test taker. The reason for this sizeable area is that different departments at the university have different cut points for admission decisions. As such, there is a need for a substantial number of items located around the multiple probable cut points in order to clearly define test takers' level of English knowledge.

Evaluating Targeting Over Time

Item performance across different administrations of the examination must be interpreted cautiously. Ideally, each entrance examination would feature a reoccurring set of core items that would anchor the item difficulty estimates for the other items. Since item security concerns often preclude this, population targeting and cut-point targeting are the most practical alternatives for the purposes of monitoring item difficulty. Yet, a key assumption is that the test takers' overall knowledge of English is relatively stable from year to year. This assumption, of course, is open to debate and thus should be factored in when evaluating item performance.

Research Questions

The purpose of this investigation is thus to demonstrate how population targeting and cut-point targeting can be used to monitor the difficulty of test items appearing on the English section of a university entrance examination over a 4-year period. The following research questions guided this investigation:

1. Which types of test items overlapped with test takers' knowledge of English?
2. Which types of items were located around the probable cut point for university admissions?

It is important to note that this investigation does not examine the relationship between the number of points allocated to different test items and test takers' performance. Although this is a very interesting area of research, which involves issues such as item weighting and rater effects (e.g., Myford & Wolfe, 2003), this investigation focuses upon item difficulty. As a result, it is important to clarify the relationship between item difficulty and the number of points allocated to an item. Item difficulty is defined as the proportion of incorrect responses a group of test takers have made on a particular test item. On the other hand, the number of points allocated to a particular item is a qualitative decision made (a) in advance by those who write and edit the examination; and, in certain cases, (b) afterward by those who grade the responses. In addition, there is not necessarily a direct relationship between the number of points allocated to a particular item and the level of difficulty that the item poses for test takers. While point allocations must be made before test takers sit the examination, the level of difficulty of items is not usually known until after the tests have been graded. Nevertheless, the levels of item difficulty found on previous examinations can inform decisions concerning point allocations for items to be used on future tests.

Method

A research team collected 1,996 test takers' examination responses from four consecutive administrations of the English section of an entrance examination from a national university located on the outskirts of Tokyo, Japan. This data was then submitted to a Rasch analysis implemented by WINSTEPS (Linacre, 2006) to produce the estimates of item difficulty for each examination.

Analysis

Classifying the Types of Test Items

In order to provide a more detailed account of item difficulty, the items were classified according to a) the skill or content area assessed by the test item (Brown & Yamashita, 1995a; 1995b), b) the characteristics of the test item, and c) the requirements of the test item (Gronlund, 1998). This level of detail led to initial disagreements amongst the research team concerning the exact characteristics of some test items. In these cases, the members of the research team discussed their differences until an agreed classification was reached. The following characteristics were specified:

Skills/Content Areas

The respective types of items were designed to assess the ability to:

- Translate a partial phrase or a sentence from English into Japanese.
- Describe a picture, illustration, table, or chart.
- Summarize what they have read in a reading passage.
- Comprehend or understand a reading passage or a written conversation.
- Inter meaning from a reading passage or a written conversation.
- Recognize different types of Narrative Structures such as cohesive devices.

Language of the Test Item Prompt

- Japanese or English

Language of the Test Takers' Responses (either receptively or productively)

- Japanese or English

Test Takers' Response to the Test Item

- Receptive: Test takers had to display their knowledge of English receptively (e.g., a multiple choice format).
- Productive: Test takers had to display their knowledge of

English productively (e.g., with a written response ranging from a word or a partial phrase to a paragraph).

The Source of the Test Takers' Response (Productive Items Only)

- Text-based: Test takers had to provide the correct response primarily using information appearing in an accompanying reading passage, illustration, table, and/or chart.
- Student-based: Test takers were required to provide the correct response primarily using their knowledge of English without direct assistance from an accompanying reading passage, illustration, table, or chart.

Item Format (Receptive Items Only)

- Multiple-choice: Test takers had to choose the correct response from a group of possible answers. The number of distractors accompanying the correct response is noted (i.e., CEERM2 means that this particular multiple-choice question had two distractors).
- Word bank: Test takers were required to choose a number of correct responses from a word bank of possible answers. The number of distractors accompanying the correct responses is also noted.
- True or False: Test takers were asked to indicate whether or not a statement was either true or false according to the accompanying reading passage, illustration, table, and/or chart.

This classification system produced a five-character tag for each productive test item. For example, a five-character tag of "TJJPT" refers to a test item that requires test takers to complete a Translation, the question prompt for the item was written in Japanese, test takers were required to write their response in Japanese (i.e., they translated a partial phrase or a sentence from English into Japanese), the item was Productive, and the information needed to complete the translation was found in an accompanying Text. Each receptive item has a six-character tag to account for the presence of distractors. For example, "CEERM3" refers to a test item designed to assess test takers' Comprehension skills, the question prompt for the item was written in English, the possible answers were written in English, the item was Receptive, the item format was Multiple-choice, and there were 3 distractors along with the correct response.

Results

The 4-Year Overall Performance

Table 2 shows the overall performance over a 4-year period, defined in terms of a) the number of items used in each examination, b) the number and the percentage of items that overlapped with test takers' level of ability (i.e., population targeting), and c) the number and the percentage of items that had a level of difficulty located around the probable cut point for university admissions (i.e., cut-point targeting). This information is also provided for items designed to assess test takers' productive and receptive knowledge of English.

Table 2. The overall performance of the items on the English section of the university entrance examination over a 4-year period

Items	Year 1		Year 2		Year 3		Year 4	
Total number of items	23		39		40		36	
Population targeted	16	70%	17	44%	30	75%	34	94%
Cut-point targeted	10	43%	9	23%	19	48%	12	33%
Productive items	12	52%	26	67%	24	60%	23	64%
Population targeted	11	92%	15	58%	15	63%	23	100%
Cut-point targeted	7	58%	8	31%	11	46%	9	39%
Receptive items	11	48%	13	33%	16	40%	13	36%
Population targeted	5	45%	2	15%	15	94%	11	85%
Cut-point targeted	3	27%	1	8%	8	50%	3	23%

The number of items on the English section ranged from the mid-30s to 40. The exception was Year 1 with 23 items. The percentage of items that targeted the ability level of the test taker population varied considerably: the highest was 94% in Year 4, the lowest was 44% in Year 2. In terms of items located around the probable cut point, the percentage varied from 48% in Year 3 to 23% in Year 2.

Over the period of four years, productive items composed over 50% of the items on the examination. Year 2 had the highest percentage of productive items with 67%. The percentage of productive items targeting the test taker population varied considerably, from 100% in Year 4 to 50%

in Year 3. Finally, the percentage of productive items located around the probable cut point was highest in Year 1 (58%).

Similarly, the receptive-knowledge items varied in terms of population targeting. The highest percentage occurred in Year 3 (94%). Year 3 also had the highest percentage of receptive items located around the probable cut point with 50%. The percentage of receptive items targeting the probable cut point in the other years was, however, significantly lower.

The Performance of the Productive Measures of Test Takers' English Knowledge

Table 3 shows the frequency of the different types of productive items used over a 4-year period. The frequency of these items' level of difficulty overlapping with test takers' knowledge of English (i.e., population targeting) and the frequency of these items' level of difficulty being located around the probable cut point (i.e., cut-point targeting) are also shown.

Over the 4-year period, there were a number of different types of productive items utilized. The most commonly occurring item types were TJJPT, IJJPT, SJJPS, and CEEPT. Generally, each examination featured a group of item types that composed the majority of productive items. In Year 1 the combination of SEEPT, TJJPT, SJJPS, and CJJPT items composed 76% of the productive items; in Year 2 CEEPT and CEEPS items combined for 69%; in Year 3 NEEPS and DEEPS items reached 71%, and in Year 4 CEEPT, CEEPS, and CJEPT items combined for 72% of the productive items.

In terms of targeting, the level of difficulty for the different types of productive items largely overlapped with the test takers' knowledge of English. The only exceptions were the SEEPT items in Year 2 and the SEEPS item in Year 3. The percentage of productive items located around the probable cut point was generally lower. Each examination had at least one type of productive item with a level of difficulty not located around the probable cut point: TJJPT, IJJPS, and IEEPT in Year 1; TJJPT, IJEPT, and SEEPT in Year 2; SEEPS and CEEPT in Year 3, and CJEPT in Year 4.

Table 4 shows the collective performance of the different types of productive items as well as their level of difficulty compared to the location of the probable cut point. The majority of item types, which were not located around the probable cut point, had a level of difficulty that was lower than the average ability of the test takers. In other words, these items did not pose a significant challenge for the test takers. The only

Table 3. The performance of productive items over 4 years

Items	Year 1						Year 2					
	Occurred		Pop. Targeted		Cut Point Targeted		Occurred		Pop. Targeted		Cut Point Targeted	
TJJPT	2	17%	1	50%	0	0%	3	12%	3	100%	0	0%
IJJPS	1	8%	1	100%	0	0%						
IJJPT	1	8%	1	100%	1	100%	1	4%	1	100%	1	100%
SJJPS	2	17%	2	100%	2	100%						
CJJPT	2	17%	2	100%	2	100%						
IJEPT							1	4%	1	100%	0	0%
CJEPT												
IEEPT	1	8%	1	100%	0	0%						
SEEPT	3	25%	3	100%	2	67%	3	12%	0	0%	0	0%
SEEPS												
CEEPT							8	31%	3	38%	3	38%
CEEPS							10	38%	7	70%	4	40%
NEEPS												
DEEPS												
Total	12	100%	11	92%	7	58%	26	100%	15	58%	8	31%

Items	Year 3						Year 4					
	Occurred		Pop. Targeted		Cut Point Targeted		Occurred		Pop. Targeted		Cut Point Targeted	
TJJPT	2	8%	2	100%	2	100%	1	4%	1	100%	1	100%
IJJPS												
IJJPT							1	4%	1	100%	1	100%
SJJPS	1	4%	1	100%	1	100%	2	9%	2	100%	1	50%
CJJPT												
IJEPT							2	9%	2	100%	2	100%
CJEPT							4	17%	4	100%	0	0%
IEEPT												
SEEPT												
SEEPS	1	4%	0	0%	0	0%						
CEEPT	3	13%	2	67%	0	0%	9	39%	9	100%	3	33%
CEEPS							4	17%	4	100%	1	25%
NEEPS	9	38%	3	33%	3	33%						
DEEPS	8	33%	7	88%	5	63%						
Total	24	100%	15	63%	11	46%	23	100%	23	100%	9	39%

two exceptions were four NEEPS items and one DEEPS item which had a level of difficulty that surpassed the ability of test takers who were located two standard deviations above the average test taker. In short, these items were quite difficult.

Table 4. The collective performance of the productive items

Items	Occurred	Cut-point targeted		Below cut point		Above cut point	
TJJPT	8	3	38%	5	63%	0	0%
IJJPS	1	0	0%	1	100%	0	0%
IJJPT	3	3	100%	0	0%	0	0%
SJJPS	5	4	80%	1	20%	0	0%
CJJPT	2	2	100%	0	0%	0	0%
IJEPT	3	2	67%	1	33%	0	0%
CJEPT	4	0	0%	4	100%	0	0%
IEEPT	1	0	0%	1	100%	0	0%
SEEPT	6	2	33%	4	67%	0	0%
SEEPS	1	0	0%	1	100%	0	0%
CEEPT	20	6	30%	14	70%	0	0%
CEEPS	14	5	36%	9	64%	0	0%
NEEPS	9	3	33%	2	22%	4	44%
DEEPS	8	5	63%	2	25%	1	13%
Total	85	35		45		5	

The Performance of the Receptive Measures of Test Takers' English Knowledge

Table 5 shows the frequency of the different types of receptive items used in the English section of the university entrance examination over a 4-year period. The frequency of these items' level of difficulty overlapping with test takers' knowledge of English and the frequency of these items' level of difficulty being located around the probable cut point are also shown.

Table 5. The performance of the receptive measures of test takers' English knowledge

Items	Year 1						Year 2					
	Occurred		Pop. Targeted		Cut Point Targeted		Occurred		Pop. Targeted		Cut Point Targeted	
SJERW0												
CEERW0							4	31%	0	0%	0	0%
CEERW1							6	46%	0	0%	0	0%
NJERW2												
IJERW3												
CJERW4												
CEERM2	6	55%	3	50%	2	33%						
CEERM3	5	45%	2	40%	1	20%	3	23%	2	67%	1	33%
CJERM3												
CJERTF												
Total	11	100%	5	45%	3	27%	13	100%	2	15%	1	8%

Items	Year 3						Year 4					
	Occurred		Pop. Targeted		Cut Point Targeted		Occurred		Pop. Targeted		Cut Point Targeted	
SJERW0							5	38%	5	100%	0	0%
CEERW0												
CEERW1												
NJERW2	5	31%	5	100%	1	20%						
IJERW3	2	13%	2	100%	1	50%						
CJERW4	2	13%	1	50%	1	50%	2	15%	2	100%	2	100%
CEERM2												
CEERM3							6	46%	4	67%	1	17%
CJERM3	2	13%	2	100%	2	100%						
CJERTF	5	31%	5	100%	3	60%						
Total	16	100%	15	94%	8	50%	13	100%	11	85%	3	23%

The 4-year period had two distinct patterns. During the first 2 years, the examinations items exclusively assessed test takers' comprehension skills. The next 2 years, however, featured a greater variety of receptive items that assessed other skills and content areas such as summarize, inference, and narrative structures in addition to test takers' level of reading comprehension. The number of receptive items was generally stable over the 4-year period with the exception of Year 3 with 16 receptive items, which coincides with a greater range of skills being assessed.

In terms of population targeting, the receptive items varied considerably over the 4-year period. Year 2 had the poorest coverage with only 2 out of 13 items located within the test takers' overall level of English knowledge, which was a significant drop from 5 out of 11 items in Year 1. Years 3 and 4 performed much better with only one receptive item in Year 3 and two items in Year 4 not targeting the test takers' overall English knowledge.

The percentage of receptive items located around the probable cut point also varied considerably over the 4-year period. Year 3 had the highest percentage with 53% followed by Year 1 (27%) and Year 4 (17%). Not surprisingly, Year 2 had the lowest percentage with only 8% of the receptive items located around the probable cut point. Table 6 shows that all of the receptive items not located around the probable cut point had a level of difficulty lower than the average ability level of the test takers. The only exception was two difficult CJERTF items in Year 3.

Table 6. The collective performance of the receptive items

Items	Occurred	Cut-point targeted		Below cut point	
SJERW0	5	0	0%	5	100%
CEERW0	4	0	0%	4	100%
CEERW1	6	0	0%	6	100%
NJERW2	5	1	20%	4	80%
IJERW3	2	1	50%	1	50%
CJERW4	4	3	75%	1	25%
CEERM2	6	2	33%	4	67%
CEERM3	14	3	21%	11	79%
CJERM3	2	2	100%		
CJERTF	5	3	60%	2	40%
Total	53	15		38	

Table 7 shows the performance of the different item formats (i.e., Multiple-choice, Word bank, and True or False items) as well as their performance according to the number of distractors. During the 4-year period, the receptive items were predominantly multiple choice items (22) or word bank items (26). These two item formats performed similarly in terms of population targeting with 59% of multiple choice items and 58% of word bank items targeting the test takers' overall knowledge of English. These two item formats, however, differed in terms of the percentage of items located around the probable cut point. Multiple choice items had 32% cut-point targeting compared to 19% for word bank items. Although True or False items were used only in Year 3, they performed quite well with 100% population targeting and 60% cut-point targeting.

Table 7. The performance of different item formats and their performance according to the number of distractors

Item Formats	Occurred	Population targeted		Cut-point targeted	
M	22	13	59%	7	32%
W	26	15	58%	5	19%
TF	5	5	100%	3	60%
W0	9	5	56%	0	0%
W1	6	0	0%	0	0%
W2	5	5	100%	1	20%
W3	2	2	100%	1	50%
W4	4	3	75%	3	75%
M2	6	3	50%	2	33%
M3	16	10	63%	5	31%

The use and performance of distractors in multiple choice and word bank items varied considerably during the 4-year period. Whereas the multiple choice questions had either two or three distractors the word bank items ranged from no distractors to four. In terms of population targeting, three distractors performed better than two for multiple choice items. For the word bank items, having no distractors or only one distractor resulted in poorer performances. In terms of the percentage of multiple choice format items targeting the probable cut point, two distractors

performed better than having three distractors. Word bank items, on the other hand, had better cut-point targeting with an increased numbers of distractors.

Discussion

Once again, the implications arising from the results must be considered carefully since the item difficulties over the four years are not anchored to a common set of items. The discussion then examines a number of factors that might underlie the performance of the productive and receptive items on the different examinations. This study focuses upon the characteristics of the different item types and does not take into consideration linguistic factors, such as vocabulary level or the level of readability which may also mediate the interaction between the test takers and the examination (see Weaver & Sato, 2008, for an example of this type of analysis).

Overall Performance of the English Section of the University Entrance Examination Over a 4-Year Period

This investigation reveals a considerable amount of variation from year to year. For example, Table 2 shows that the number of items is almost twofold between Year 1 and Years 2, 3, and 4. The initial increase of test items, however, did not necessarily improve performance. This finding is counter to conventional thinking that an increased number of items leads to improved test performance in terms of reliability (Traub & Rowley, 1991). Although the correlation-based reliability coefficient of the entrance examination increased by 0.04 from Year 1 to Year 2, the percent of population-targeted items fell from 70% in Year 1 to 44% in Year 2 despite an increase of 16 items. The additional items in Year 2 also did not help increase the number of items located around the probable cut point. In Years 3 and 4, the correlation-based reliability coefficient continued to increase by 0.05 each year. In addition, the percentage of items targeting the test taker population continued to increase to 75% in Year 3 and 94% in Year 4.

In terms of the percentage of items located around the probable cut point, Year 3 (48%) exceeded the level reached in Year 1 (43%), doing so in two distinctive ways. In Year 1, the productive items performed better than the receptive items. In Year 3, the performances of the productive and receptive items were more balanced.

Potential Factors Underlying the Performance of the Productive Items

Over the 4-year period, reading comprehension was the most commonly tested skill with 31 out of 40 items targeting the test takers' overall knowledge of English and 13 items located around the probable cut point (see Table 3). A factor that had a consistent influence on this type of item was whether or not test takers were required to respond in Japanese or English. The Japanese-response items (i.e., the two CJJPT items in Year 1) were more difficult than the items requiring responses in English (i.e., the four CJEPT items in Year 4). Table 3 shows that the CJJPT items were located around the probable cut point; in contrast, the CJEPT items were located below the mean ability level of the test takers, but still within the population target. One possible explanation for this difference is that requiring test takers to demonstrate their level of reading comprehension productively in English may be a relatively easy task since it requires test takers to identify what needs to be comprehended in reading text and transfer this information to their answer sheet. CJJPT items, on the other hand, require the additional steps of translating the information from the reading passage into Japanese as well as summarizing and synthesizing information from the reading passage. Another source of support for this explanation is a study that found that higher levels of cognitive load generally led to increased levels of item difficulty for reading comprehension questions used on a university entrance examination (Weaver & Romanko, 2005).

An interesting extension to this finding is the comparison between productive items with question prompts written in Japanese that required Japanese responses from test takers versus items with English question prompts requiring English responses. Table 3 shows that although English prompt/response items (59) occurred almost three times as often as Japanese prompt/response items (19) during the 4-year period, Japanese prompt/response items performed at a higher level. In terms of population targeting, 95% of the Japanese prompt/response items targeted test takers' overall level of English knowledge compared to 66% of the English prompt/response items. The difference between these two types of items also was apparent with cut-point targeting: 63% of the Japanese prompt/response items compared to 36% of the English prompt/response items. However, this finding should not be used as a justification for the use of Japanese prompt/response items. Rather this finding highlights a unique challenge that faces foreign-language-test writers. Table 4 shows that the majority of English prompt/response items were located below the mean ability level for the test takers and thus were within the realm of

their English knowledge. As a result, test writers need to design items that require more than identification and copying skills from test takers. However, Table 4 also shows that the difficulty level for one DEEPS item and four NEEPS items was beyond the ability level of the most able test taker in Year 3. In other words, these items designed to assess test takers' descriptive skills and knowledge of narrative structures were far too difficult and thus reveal the challenge of writing English prompt/response items located around the probable cut point.

Another interesting finding is that the productive items located around the cut point assessed a number of different skills and content areas over the 4-year period. Such variety not only helps to create a more comprehensive account of English knowledge, but also lends support to the argument that the examinations evaluate more than test takers' grammatical competence (e.g., Guest, 2000). It is hoped that this finding will have a positive washback effect on future test takers and their teachers: that a well rounded knowledge of English is important.

Potential Factors Underlying the Performance of the Receptive Items

The performance of the receptive items reveals an important rationale for tracking and targeting items. Table 5 shows that the receptive items in Years 1 and 2 focused exclusively on reading comprehension skills using English question prompts and English response choices. These items unfortunately did not provide significant amounts of information about the test takers' overall level of English knowledge, especially in Year 2 with only 15% of the items falling within the population target. Years 3 and 4, however, featured receptive items that assessed a larger range of skills and content areas and utilized a variety of question prompt/response choice formats. These changes resulted in an increased number of receptive items targeting the test takers' overall level of English knowledge. Year 3 also had the highest percentage of receptive items (50%) located around the probable cut point. The introduction of new types of receptive items, however, must be considered as a work in progress. For example, the five NJERW2 items in Year 3 successfully targeted the test-taking population, but had only one item located around the probable cut point. The five SJERW0 items in Year 4 also had a similar performance with good population targeting, but poor cut-point targeting. A systematic approach of tracking and targeting can provide test writers with vital information about how new types of receptive items performed in order to maintain or improve their performance in future entrance examinations.

Similar to the productive items, there were a number of factors that influenced the performance of receptive items. Although multiple choice and word bank items had similar amounts of success targeting the population of test takers, a greater percentage of multiple choice items were located around the probable cut point. This finding highlights an important design feature that differentiates these two types of item formats: whereas the possible answers for a multiple choice item are exclusive to one item, a number of different receptive items can share a common word bank. One implication of a shared word bank is that the number of possible answers decreases as test takers complete the different items. As a result, items that initially have a level of difficulty located around the probable cut point may become easier through a process of elimination. A means of circumventing this shortcoming is to design items so that possible answers can be used more than once. However, designing items so that alternative answers are a credible choice for multiple items can be a formidable challenge.

Another factor that influenced the performance of receptive items was the number of distractors accompanying the correct response. The influence of this factor, however, varied according to the item format. The number of distractors had a relatively consistent effect upon the performance of word bank items. Generally, an increased number of distractors led to higher percentages of population and cut-point targeting. An increased number of distractors in multiple choice items, on the other hand, resulted in better population targeting but poorer cut-point targeting. This finding provides partial support for the Shizuka, Takeuchi, Yashima, & Yoshizawa (2006) suggestion that traditional four-option multiple choice items can be reduced to three alternatives without sacrificing test performance.

Overall we found that relatively few receptive items were located around the probable cut point with the exception of Year 3. During the 4-year period, only 15 out of 53 receptive items reached this level of difficulty. This stands in contrast to the productive items. Productive items such as translations or written compositions usually utilize poly-chotomous rating scales. As a result, the productive items employing a multiple-point rating scale can provide partial credit for test takers' responses and thus define a larger range of English knowledge. For example, Weaver and Sato (2007) found that a 3-point rating scale was optimal for assessing test takers' grammatical competence set within a communicative situation. Receptive items, in contrast, score test takers' responses as either right or wrong. Dichotomous rating scales thus define

a very specific level of English knowledge. It is possible to have receptive questions that utilize a polychotomous rating scale where test takers receive credit for a response choice that is not entirely correct, but reveals attainment of some developmental stage on the way towards target-like use. Designing this type of receptive item, however, requires a great deal of planning and care. Moreover, test writers will need to be versed in processability theory (Pienemann, 1998) and the research concerning the developmental steps of different grammatical features such as negation (e.g., Batet & Grau, 1995), wh-question formation (e.g., Mackey, 1999), and relative clauses (e.g., Diessel, 2004).

Conclusion

The English section of a university entrance examination provides test writers with a multitude of challenges. In many cases, a strictly defined time limit combined with historical influences of how things are done govern the number and the types of items that appear. As such, a systematic approach focusing upon the performance of previous test items can provide test writers with an essential source of information. Tracking and targeting items allows test writers to gain a deeper understanding of how different factors potentially mediate item performance. Since most entrance examinations do not share a common set of items, test writers should be cautious when comparing different test performances over time. In other words, they should be continually on the outlook for consistent trends that appear over multiple administrations of the examination. Another focal point for test writers should be the importance of cut-point targeting in order to ensure the highest degree of measurement accuracy. In essence, the whole idea is to transform hindsight gained from previous item performances into foresight which can help improve future performance.

Acknowledgments

The authors would like to acknowledge the past and the present governing body of Tokyo University of Agriculture and Technology for their support of this research project. In addition, gratitude is extended to the many people who have contributed to this research in one way or another. This investigation was funded by a JSPS grant-in-aid of research ([平成17年度～平成18年度科学研究費補助金基盤研究(C) 課題番号17520371]).

Christopher Weaver is a lecturer at Toyo University. His area of research includes task-based instruction, individual differences, and psychometrics with a special focus on practical applications of Rasch measurement theory.

Yoko Sato is a Professor of English at Tokyo University of Agriculture and Technology. Her main research interest focuses upon a close textual analysis of poetry and drama, language testing, and vocabulary learning strategies. She is a cotranslator of Morton N. Cohen's *Lewis Carroll: A Biography* (1999).

References

- Batet, M., & Grau, M. (1995). The acquisition of negation in English. *Atlantis*, 17 (1), 27-44.
- Brown, J. D., & Yamashita, S. (1995a). English language entrance examinations at Japanese universities: 1993 and 1994. In J. D. Brown & S. Yamashita (Eds.), *Language teaching in Japan* (pp. 86-100). Tokyo: JALT.
- Brown, J. D., & Yamashita, S. (1995b). English language tests at Japanese universities: What do we know about them? *JALT Journal*, 17 (1), 7-30.
- Diessel, H. (2004). *The acquisition of complex sentences*. Cambridge: CUP.
- Gershon, R. (2006). Understanding Rasch measurement: Computer adaptive testing. *Journal of Applied Measurement*, 6 (1), 109-127.
- Gronlund, N. (1998). *Assessment of student achievement* (6th ed.). Boston: Allyn and Bacon.
- Guest, M. (2000). "But I have to teach grammar!": An analysis of the role "grammar" plays in Japanese university English entrance examinations. *The Language Teacher*, 20 (11), 23-29.
- Ito, A. (2005). A validation study on the English language test in a Japanese national wide university entrance examination. *Asian EFL Journal*, 7 (2), 91-117.
- Kikuchi, K. (2006). Revisiting English entrance examinations at Japanese universities a decade later. *JALT Journal*, 28 (1), 77-96.
- Linacre, J. (2006). *WINSTEPS Rasch measurement computer program* (Version 3.60). Chicago: Winsteps.com.
- Mackey, A. (1999). Input, interaction and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition*, 21 (4), 557-587.
- Myford, C., & Wolfe, E. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part 1. *Journal of Applied Measurement*, 4 (4), 386-422.
- Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Amsterdam: Benjamins.

- Shizuka, T., Takeuchi, O., Yashima, T., & Yoshizawa, K. (2006). A comparison of three- and four-point English tests for university entrance selection purposes in Japan. *Language Testing*, 23 (1), 35-57.
- Smith, E.(2000). Metric development and score reporting in Rasch measurement. *Journal of Applied Measurement*, 1 (3), 303-326.
- Smith, E., (2001). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. *Journal of Applied Measurement*, 2 (3), 281-311.
- Traub, R., & Rowley, G. (1991). Understanding reliability. *Educational Measurement: Issues and practice, National Council on Measurement in Education*, 10 (1), 37-45.
- Weaver, C., & Romanko, R. (2005). Assessing oral communicative competence in a university entrance examination. *The Language Teacher*, 29 (1), 3-9.
- Weaver, C., & Sato, Y. (2007). Jukensha no eigo-communication-noryoku-shikibetsu no tamen no hyouka-shakudo ni tsuite-bunseki-rei-Rasch model ga kitaisuru hyouka-shakudo no performance [The use of rating scales to differentiate test takers' English communicative competence: the Rasch model expectations of a well performing rating scale]. *Daigaku Nyushi Kenkyu Journal*, No.17, 135-142.
- Weaver, C., & Sato, Y. (2008). Estimating and determining the difficulty of English reading passages used in a university entrance examination. Manuscript submitted for publication.
- Wolfe, E. (2000). Equating and item banking with the Rasch model. *Journal of Applied Measurement*, 1 (4), 409-434.
- Wright, B., & Stone, M. (1979). *Best test design: Rasch measurement*. Chicago: Mesa Press.

Reviews

Language Policy, Culture, and Identity in Asian Contexts.
Amy B. M. Tsui and James W. Tollefson (Eds.). Mahwah,
NJ: Lawrence Erlbaum Associates, 2007. ix + 283 pp.

Reviewed by

R. A. Brown

Bunkyo University, Chigasaki, and Waseda University

This book addresses Asian policy responses to the challenges posed by globalization and the concomitant need for English, both as a medium of international communication and as the primary language of information technology. Tsui and Tollefson point out that language policy subsumes the three areas of language planning, language ideology, and language practice. The 14 articles in the book focus on these three themes. The chapters on Japan, South Korea, Malaysia, Singapore, and Cambodia address the extent to which policy manages the potentially deleterious impact of the spread of English on indigenous languages and the cultural values and belief systems they embody, while promoting its positive benefits. The chapters on Hong Kong, Brunei, and New Zealand examine the relations between language and the construction of cultural identities. The chapters on India, Nepal, Pakistan, and Bangladesh explore the role of English in environments where there has emerged local resistance to linguistic homogenization. Each chapter has its merits. However, I will focus on the chapter of most relevance for *JALT Journal* readers.

Kayoko Hashimoto's contention in the chapter "Japan's language policy and the lost decade" is that "Japanese education is designed to reduce the danger of dependency on the West by restricting the introduction of Western ideas to technical matters" (p. 26). Hashimoto argues that Japanese policy is intended less to promote the official goal of ensuring that "all Japanese acquire a working knowledge of English," than to promote "Japaneseness" (p. 27). This is not a new concern. Former U.S. Ambassador to Japan (1966-1969) U. Alexis Johnson recalled the Japanese Vice Minister of Education encouraging school children to "learn foreign languages but not learn them too well, because their souls were embodied in their knowledge of Japanese" (1984, p. 459). There appears to be something of this fear in current policy prescriptions. Languages embody

cultural beliefs and values, and influence behavior. Social distinctions that are encoded in one language may not be in another. English carries the worldview and national practices of the Anglo-American world and as such potentially endangers indigenous cultures. Thus, Japanese official policy seems ambivalent between wanting the benefits of English while fearing the consequences of “too much” English.

As of March 2007, virtually all Japanese elementary schools (96%) offer some English instruction, in the form of songs, games, greetings, self-introductions, or pronunciation drills. The instruction is provided by homeroom teachers and amounts to 14.8 hours per year. Approximately 95% of these teachers lack qualifications as language instructors. However, Education Minister Bunmei Ibuki has recently expressed the view that Japanese students should first perfect their speaking and writing skills in their mother tongue before tackling a foreign language (Most Elementary Schools, 2007), which seems to imply a subtractive theory of second language acquisition, such that the second language impedes development of the native language, precisely as suggested by the official quoted by Johnson above.

In view of the official ambivalence, it is difficult to imagine schools devoting the resources needed to overcome more immediate and practical difficulties, such as the lack of qualified teachers and limited classroom time. Hashimoto’s conclusion is that the Japanese government’s policy is in fact a successful response to the perceived challenge posed by English, and that broadly based competence in communicative English is not an official objective.

Interestingly, Yim Sungwon arrives at essentially the same conclusion regarding South Korea based on a content analysis of middle school textbooks, and many of the other articles suggest that official ambivalence toward English is not confined to Japan.

Tsui and Tollefson’s introductory chapter does a good job of putting the subsequent contributions in perspective. However, although a wide variety of Asian countries are surveyed, the situations in Thailand, Vietnam, and the People’s Republic of China are neglected. Despite this, overall the collection is a useful addition to the literature on language policy in Asia.

References

- Johnson, U. A. (1984). *The right hand of power*. Englewood Cliffs, NJ: Prentice-Hall.
- Most elementary schools teach English. (2007, March 3). *The Japan Times*, p. 2.

***Teaching English as a Foreign or Second Language: A Teacher Self-Development and Methodology Guide, (2nd Ed.).* Jerry G. Gebhard. Ann Arbor: The University of Michigan Press, 2006. xii + 268 pp.**

Reviewed by

Mieko Fukushima

Toyama University of International Studies

Jerry Gebhard offers a guided approach to the basic components of English language teaching in this book, which is intended primarily for preservice student teachers who are learning to teach EFL or ESL, but can also be used as a text for those in-service teachers who are on teacher development programs. The volume invites the reader not only to learn about teaching English but also how to explore teaching. Specific themes are: Part 1 (Chapters 1 to 3): Self-Development, Exploration, and Settings; Part 2 (Chapters 4 to 7): Principles of EFL/ESL Teaching; and Part 3: (Chapters 8 to 11): Teaching Language Skills.

Each chapter follows the same format. Starting with an insightful remark, typically from a prominent figure in language teaching, it has two major sections. The first section discusses a set of questions posed at the beginning. Almost all of them are either *what* or *how* questions. The second section offers groups of tasks for the reader's development as an EFL or ESL teacher. A variety of specific functions are implied in the tasks (e.g., becoming aware of one's own ideas and creating a lesson plan). The chapter ends with an updated list of recommended readings and notes in this second edition.

Part 1, including the teacher self-development tasks in each chapter, should be highly beneficial to those who are in contexts where pre- and in-service programs do not provide any systematic instruction on how teachers themselves might explore their classroom teaching. This may be the case with many Japanese teachers of English. There are some methods of teacher development that have been commonly practiced in official situations, such as the pervasive method of lesson demonstration and observation. However, I personally feel that it is rare to find language teachers who are learning such a range of methods for teacher self-development as those presented in this volume, unless they have been involved in a graduate school program or a research association.

The issues discussed in Part 1 include: seven key factors in teacher self-development (e.g., *"problem solving"* and *"exploration for exploration's sake"*), a comprehensive list of ways to explore teaching (which is related to the appendices containing a list of journals and of publishers), and a modeled cycle of actions involved in the self-observation method. Chapter 3, a new entry in this second edition, describes and discusses a variety of EFL and ESL settings around the world, with some reservations about this distinction. While presenting setting-by-setting descriptions, the chapter touches upon several factors highlighting differences between EFL and ESL and those within the ESL setting. Behind this lies the assumption that *"teaching English is context dependent"* (p. 55). Some setting-specific suggestions are also made in this chapter and other parts of the volume.

Three methods are utilized for the teacher self-development tasks: observation of oneself and other teachers, talking with other teachers, and writing in a teaching journal. These methods are not new to many teachers but are revitalized here in the form of guided tasks which would lead to ongoing exploration. Some values developed in the field of teacher education appear in the volume: cooperative work among teachers, non-judgmental attitudes toward teacher performance, and the teacher's own informed decisions.

Throughout the volume, discussions are fairly instructive and informative. They are largely built on teachers' work experiences, which I suppose derive from Gebhard's long-term association with students and teachers as a teacher educator. Thus the reader may get the feeling of being in a community of teachers. A fairly large portion of each chapter in Parts 2 and 3 is dedicated to common problems facing EFL/ESL teachers (e.g., *"How do I get students to use English in class?"*) and suggestions on possible solutions. Further, instructional ideas abound. Gebhard presents some of his experiences and teaching ideas developed as a teaching practitioner over many years (e.g., his personal experience of learning Japanese through developing relationships with people in a community in Japan). This feature should make his discussion more approachable and enjoyable to teachers.

As for the conceptual aspect of the volume, the first thing to note is that, in his endnotes, Gebhard makes transparent to the reader the way in which his thinking on a particular issue has been formed in relation to the sources he has drawn on. Second, the volume is a crafted work to be treasured in light of its concise, yet comprehensive and enriching manner of dealing with the basics of classroom instruction. Part 2 encompasses

four essential aspects of language teaching practice: teaching English as communication and interaction, classroom management, teaching materials/media/technology, and cultural issues (i.e., defining culture, adjusting to another culture, and cultural concepts to teach and ways of teaching them). Part 3 discusses the nature of each language skill (including “conversing” rather than speaking) and ways of teaching them. An integrated approach required for general-English types of classes can be explored with the help of Gebhard’s discussions on communicative classroom and language skills together with some of his recommended readings.

Finally, some conceptual developments in language education are incorporated, for example, the concepts of precommunicative and communicative activities coined by Littlewood in 1981 (Chapter 4), the active nature of listening (Chapter 8), and process writing (Chapter 11). The selection of those items seems to be just right for preservice teachers, and may be appropriate for in-service teachers as well on the assumption that many teachers are deprived of such developments in research and the literature.

All in all, *Teaching English as a Foreign or Second Language* by Jerry Gebhard is an excellent source of enrichment for those who are seeking to become better teachers of English. Appreciating the opportunity to get back to the basics, I would like to share this book with my undergraduate students in the secondary school teacher license program at my university and with those in-service teachers who have recently launched their teaching careers.

Reference

- Littlewood, W. (1981). *Communicative language teaching: An introduction*. Cambridge: Cambridge University Press.

***The Experience of Language Teaching.* Rose M. Senior.
Cambridge: Cambridge University Press, 2006. xiii + 301**

Reviewed by

Darren Elliott

Meijo University, Nagoya

As the orthodoxy slowly continues its shift from “teacher-fronted” to “student-centred” approaches, concepts of learner autonomy are becoming a well-established part of the mainstream and technology. This shift seems to offer students new directions for nontraditional learning. At the same time, job insecurity, low pay, and a lack of professional status are issues many teachers live with. Are teachers becoming marginalised in language education? In her broad longitudinal study, Senior places the teacher firmly at the centre of the learning process and formulates theories of what actually occurs in communicative language classrooms based on group dynamics and social-constructivism.

The book begins with a précis of the research methodology used in the study and an overview of the motivation behind the enquiry and the contexts in which it occurred. Senior utilises data from five interlinked studies comprised of extended interviews with over 100 teachers over 12 years following a grounded theory approach in which researchers use data to generate theory rather than gather information to support or refute hypotheses (Nunan, 1992). Grounded theory is often misunderstood, but the methodology and the thinking behind it are clearly explained here. For anyone considering embarking on a research project of this nature, Senior provides a valuable starting point and a number of helpful suggestions for further reading.

The author goes on to explore what motivates people to enter intensive language teaching training courses at the entry level (specifically the Certificate in English Language Teaching to Adults), and examines their early steps towards maturity and expertise as teachers of English. It soon becomes apparent that the study centres on a particular kind of language teaching career; indeed, the author explicitly points this out. Although the book specifically draws on the experiences of native English speakers teaching multilingual classes in a broadly “communicative” style in English-speaking countries, the author expresses a hope that it will resonate with teachers in other areas. In the central section of the book, Senior focuses on the social aspect of language teaching, discussing such topics

as classroom management, flexibility, rapport, and the development of learning communities within classes, effectively honing in on the ways in which teachers build and maintain group cohesion and positive environments. I believe that much of what is reported in the book will ring true in the wider teaching community, as the author hopes.

In two interesting and particularly pragmatic chapters near the end of the book, the author contextualises the conditions in which many language teachers work and considers how these might affect classroom practice. The rewards and frustrations of a career in English language education, as described by participants in the study, may well have the reader nodding in recognition. Senior also advises prudence in interpreting teachers' perspectives as research data. In particular, she sounds a note of caution to those attempting interviews for qualitative research; for example, interviewees may claim motivations which do not actually exist in an attempt to sound more professional.

In the final chapter, Senior proposes a model of classroom behaviour based on a balance between the social and the pedagogical. Although teachers often find it difficult to express exactly why they have taken certain actions in the classroom, and experienced teachers often appear to act intuitively, by analysing the reams of narrative generated over the course of the study the author is able to suggest "a teacher-generated theory of classroom practice" (p. 270). This theory is in a sense a localised solution which may have a general application. Senior returns to grounded theory to draw together her research and fits the data to existing theories, namely chaos theory and systems theory, in an attempt to find a match. She maintains that she finds her best fit in the work of van Lier and the metaphor of the classroom as a complex organism. This metaphor uses homeostasis in biological organisms—the process by which organisms adjust their functions in response to feedback—to represent the constant and often subconscious adjustments teachers make to maintain harmony and balance. But balance between what and what? Senior posits that social cohesion and effective learning are intertwined, and the flexibility of experienced teachers is intuitively informed by an understanding of group dynamics and pedagogy. A teacher will thus select the best course of action to serve both factors. This is perhaps described most pithily by the aphorism "Drawing the class together according to its social needs; Drawing the class along by attending to its learning needs; Drawing the class both together and along" (p. 280).

The book as a whole is well written and very readable. There are commentary boxes throughout which point to further reading, along with a

thorough bibliography. The structure of each chapter is topped and tailed with summaries of what came before and signposts of what comes ahead. This enables the reader to select pages of particular interest at leisure. However, I read the book from beginning to end and feel that doing so enhances the power of the final chapter in bringing together the complex anecdotal messages from classroom practitioners. This is a very valuable and enlightening study for teachers striving to develop an understanding of why they do what they do.

Reference

Nunan, D. (1992). *Research methods in language teaching*. Cambridge, UK: Cambridge University Press.

A Friendly Approach to English for Academic Purposes.
Sonia Sonoko Strain. Tokyo: Shohakusha, 2006. xiv + 246
pp.

Reviewed by

John Nevara

Kobe Gakuin University, Department of Law

The title, *A Friendly Approach to English for Academic Purposes*, nicely captures the hands-on, easy-to-understand essence of this book. Unlike other texts in the field of English for Specific Purposes (ESP) and its subfield English for Academic Purposes (EAP), this book focuses on the practical aspects of designing and implementing “an introductory academic English-as-a-foreign-language (EFL) seminar in the English Department of a small private university” (p. vii).

The author’s motivation for creating an academic EFL seminar arose from the junior year abroad program at her university. It often is the case that general English education courses—centered around skills such as general conversation, newspaper and short-essay reading (possibly with translation), and informal writing—neglect the acquisition of academic skills which might be useful for long-term study abroad.

Strain rectifies this problem by implementing a program for seven 1st-year university students on the different aspects of EAP, over a period of 26 eighty-minute classes. *A Friendly Approach to English for Academic Purposes* is her step-by-step case study of the program. As the author admits, the teaching time is too short to be ideal, but she claims the seminar to be a success after evaluating it using the methods of action research.

Since EAP "is concerned with all of the English communication skills which are required for study purposes in formal educational settings," quite a few skills need to be taught (p. 5). Strain divides the seminar into six modules: a) launching an academic life, b) interacting in academic discussions, c) lecture class simulations, d) academic English reading, e) academic English writing, and f) academic term papers. The modules are designed to be interconnected and progressive, so that each learned skill builds towards the next module. Typical of an EAP classroom, the skills involved in the above modules are not often practiced in general English courses. Furthermore, without these skills, students would not be able to fully participate in their year-abroad classes.

The text provides a detailed, chapter-by-chapter explanation of each module in practice. The general theoretical underpinnings for Strain's course design can be derived from R. R. Jordan's *English for Academic Purposes* (1997), among others. Strain does indeed strengthen her argument by applying previous scholars' research.

However, the tendency in reviewing a utilitarian book like Strain's is not to attack the theory behind her EAP seminar, but to find flaws in the specifics of her practical advice. For example, there are some teachers who would rather not follow Strain's example of using brainstorming in the classroom. Note-taking, a skill which Strain teaches in the lecture class simulations module, also may not seem worthy of so much emphasis, especially considering that a) more and more Japanese students are becoming particularly well-versed in this skill, and b) Western education focuses on the overall picture and general critical thinking skills rather than stressing specific details.

However, to be fair to the author, Strain does require the students to take notes on other students' presentations and then to use these same notes in writing a summary of each presentation, so the note-taking skill is rather nicely linked to other skills.

With a seminar of this length, of course there will be specifics to quibble over, yet there are also a few theoretical arguments which could be debated. While the author is usually scrupulous in her citations, there are

a few claims which go unsupported. For example, she states that Japanese students have a visual language orientation (p. 110). Furthermore, perhaps more disputably, she claims, "Plagiarism is not an easy concept for Japanese students to understand. One reason for this is that there is a cultural tradition of copying Buddhist sutra kanji characters one by one, as a way to meditate" (p. 178). This quote appears to be unsupported and an entirely conjectural personal opinion.

Another concern with the book is that the author occasionally makes claims without evidence. Perhaps this is a difficulty with action research methods that are based on an account of one class of seven students. For example, on page 102, she mentions that her "students engaged in the scanning activity seriously and seemed to enjoy the task." As another example, she states that during a mini-lecture phase, "a congenial sense of solidarity and serious co-learning prevailed" (p. 175). These statements, and other statements like these, may seem believable within the context of the text, but could benefit from some more specific examples and support.

Nonetheless, even considering the above "minor problems," this book is a very welcome addition to the growing EAP library. With few textbooks and teaching materials directed towards the EAP market, Strain's work can serve as a practical reference in helping educators design and implement their own EAP courses. Coupled with more general and theoretical texts such as Jordan's *English for Academic Purposes*, Strain's book gives enough advice to the teacher in Japan who is a novice to EAP. The text will also be appreciated by the more seasoned EAP educator, as Strain has clearly put much thought and effort into her course design.

Reference

- Jordan, R. R. (1997). *English for academic purposes: A guide and resource book for teachers*. Cambridge: Cambridge University Press.

***Connecting Speaking & Writing in Second Language Writing Instruction.* Robert Weissberg. Ann Arbor, MI: The University of Michigan Press, 2006. xiii + 171 pp.**

Reviewed by
Christian Perry
Hokkaido University

The title of this work proposes the union of two distinct skills, one social, the other ostensibly solitary. This disparity notwithstanding, Robert Weissberg asserts on page 2 that writing “is fundamentally a social phenomenon,” and by the end of the book, argues a convincing if concise case for infusing the writing classroom with dialogue.

Weissberg writes with the practitioner in mind, including relevant theory but emphasizing application. Numerous anecdotes and dialogs illustrate in concrete terms how discourse in the L2 writing classroom works. The organization and accessible prose lend themselves to the translation of ideas into execution.

Chapter 1, “An Introduction to Dialogue and Second Language Writing,” offers a succinct justification for linking speech with writing and shows how the communicative language teaching approach provides conceptual underpinnings for dialogue in the writing classroom.

Chapter 2, “From Talking to Writing,” lays out the theory that links speech to the development of literacy. Theorists such as Vygotsky have observed that L1 writing skills are an outgrowth of L1 oracy and “inner speech.” Weissberg argues that social interaction can likewise serve as the “springboard” for developing L2 writing skills. The last section of the chapter is devoted to guidelines for injecting dialogue into L2 composition classes.

Chapter 3, “ESL Writers and Speakers: A World of Individual Differences,” profiles three L2 learners with varying degrees of speaking and writing proficiency. Instead of undergoing a definable transition from oracy to literacy as L1 writers do, L2 learners may well (be made to) take up literacy before becoming adequately grounded in the spoken language. Irrespective of their oral skills, L2 writers are distributed along a “continuum of modality preference,” with speech on one end and writing on the other. In other words, some students favor speaking over writing, and vice versa. Teachers must therefore take into account the

dispersion of L2 learners along this continuum when designing tasks for the writing class. For example, students that favor oral activities should have a chance to exploit the modality of speech in order to drive their own writing process forward. Properties inherent in speech, such as coherence, can be transferred to written work. Weissberg includes a sample of talk-write tasks tailored to a range of students.

Chapter 4, "Beyond Teacher-Talk: Instructional Conversations in the Writing Classroom," explores teacher-directed, though not teacher-centered, whole-class interaction. It describes specific techniques for getting students to speak more, as well as more meaningfully, in service of writing assignments, then shows how unproductive student-teacher interaction common to many classrooms can be transformed into authentic dialogue, free of a set-in-stone preplanned itinerary. Students are afforded more say over the direction of the discourse, helping to nurture the discretion they need for developing their own written voice. The chapter also furnishes examples of the style of teacher questions and comments that can act as a catalyst for student involvement.

Chapter 5, "Conversations in the Writing Tutorial" (cowritten by Gina L. Hochhalter), describes how to establish productive tutorial sessions. One-on-one student-teacher interaction calls for an atmosphere less formal than that of the teacher-led classroom. Tutors must adopt a less authoritative demeanor, putting themselves on an almost-even footing with the student. Otherwise, the student becomes passive and refrains from contributing to the dialog, which then deteriorates into a private lecture. The chapter closes with recommendations on how to train tutors for effective dialogue.

Written comments on student work figure prominently in the interaction between student and teacher. Teacher commentary is usually corrective, evaluative, and delivered with a finality that precludes response from the student. Chapter 6, "Written Response as Dialogue," provides detailed suggestions for crafting appropriate feedback to students to extend this written form of dialogue. The chapter also examines the potential of electronic communication and online classes, which lend themselves to repeated exchange (e.g., teacher response to student writing, followed by student response to teacher response).

As readers peruse this book, they will undoubtedly have concerns about the feasibility of connecting L2 speaking and writing. In the final chapter, the author obliges his skeptics: taking a fresh hard look at the sort of dialogic stance he advocates, Weissberg questions the assump-

tions of the preceding chapters. He probes for shortcomings and risks in incorporating dialogue into writing instruction. This is an appropriate close to the book, since it underscores how the academic dialogue on this issue can never come to a close.

Connecting Speaking & Writing nicely complements other works in the Michigan Series on Teaching Multilingual Writers. Weissberg envisions the writing classroom as a community engaged in dialogue, and he has produced a compact volume on how to effect one.

***Motivation, Language Attitudes and Globalisation: A Hungarian Perspective.* Zoltán Dörnyei, Kata Csizér, & Nóra Németh. Clevedon, UK: Multilingual Matters, 2006. xiii + 205 pp.**

Reviewed by

Akiko Tsuda

Nakamura Gakuen University

Dörnyei, the principal author of this book, is one of the leading researchers on L2 motivation, having published many insightful works for language teachers and researchers. His work covers a variety of topics, ranging from advanced research on the attitudes of language learners (aimed at academics in linguistic fields) and hands-on material for practical use (aimed at language teachers and fledgling researchers). Some of Dörnyei's work has been translated into Japanese and has been well received among Japanese teachers and researchers.

This book is based on the results of the largest national survey on language attitude and motivation available to date. The survey, which involved over 13,391 13 to 14 year old language learners in Hungary, was carried out on three successive occasions: 1993, 1999, and 2004. Five foreign languages were targeted: English, German, French, Italian, and Russian in six language communities in Hungary. This ensured that data were gathered from each region of the country. The results of the survey were not confined to the European environment, but had wider implications concerning changes in attitude, motivational dynamics, and the globalization of language.

If the reader is interested in the application of research on language learners' motivation, two other books of Dörnyei's are also highly recommended. His *Motivational Strategies in the Language Classroom* (2001) has been an especially helpful guide and provides many useful hints to motivate learners in the class. Additionally, Dörnyei's *Questionnaires in Second Language Research: Construction, Administration, and Processing* (2003) has been an indispensable step-by-step guide to designing questionnaires for my own research.

There are seven chapters in the book including the introduction, summary, and conclusion, as well as 16 appendices. In Chapters 1 and 2, the authors describe the theoretical background and basic information about the surveyed country, Hungary. These chapters make the book accessible not only for specialists but also nonprofessionals who may lack knowledge of Hungary or the study of language globalization, language attitudes, and language learning motivation. Chapters 3 and 4 focus on the main findings of their survey. Chapter 3 deals with attitudinal and motivational dimensions and their changes over the 12-year period. Chapter 4 highlights the three modifying factors in language attitudes and motivation: gender, geographical location, and school instruction. Chapter 5 introduces a new motivation theory known as the "L2 Motivational Self System," which consists of three main components: "Ideal L2 Self," "Ought-to L2 Self," and "L2 Learning Experience." Chapters 6 and 7 analyze two issues: the motivational profiles of language learners and the effects of intercultural contact on language. Finally, the summary focuses on: a) the globalization of language, b) motivation and the self, and c) intercultural contact.

The survey presents detailed findings on the relationships between motivation, language attitudes, and globalization. It has caused me to rethink my current teaching situation. For example, language classrooms are no longer homogeneous. Each class I am currently teaching includes *ryugakusei* (students from overseas) and / or *kikokushijo* (Japanese returnee students). Further, the number of fellow language teachers from overseas working in Japan has been increasing. In addition, I have observed that the cultural content of course material used in English language classes has gradually shifted from Anglo-American English to World Englishes. These dramatic changes sometimes amaze those of us who are teaching in real-world classroom situations. This survey, therefore, provides me with insightful perspectives for my own classes in Japan. As a result, my own teaching environment has been globalized.

Despite the fact that Hungary is geographically distant from Japan, we who live in the era of globalization can learn a lot from this book, which is full of examples of surveys that have been carried out with a long-term perspective. The book can lead us to reassess foreign language education and its policy in Japan. It is recommended reading for individual foreign language teachers and researchers as well as language curriculum developers and policymakers serving as decision makers for language programs in school systems.

References

- Dörnyei, Z. (2001). *Motivational strategies in the language classroom*. Cambridge: Cambridge University Press.
- Dörnyei, Z. (2003). *Questionnaires in second language research: Construction, administration, and processing*. London: Lawrence Erlbaum Associates.

Motivation and Experience in Language Learning. **Yoshiyuki Nakata. Bern: Peter Lang, 2006. 278 pp.**

Reviewed by
Phillip Barkman
Asia University

Motivation and Experience in Language Learning began as a Ph.D. dissertation with the aim of showing that language learning motivation and experiences in language learning are clearly connected. Nakata uses qualitative and quantitative research approaches to help identify links between motivation and experience. The book consists of 11 chapters that can be divided into four sections. Chapters 1 to 3 deal mainly with the issue of motivation and look at it from different angles. Chapters 4 to 6 touch on theories in educational and cognitive psychology, especially as to how they relate to motivation and autonomy. Chapters 7 to 10 present an analysis of two studies and in Chapter 11, Nakata draws a summary of the discussion including pedagogic implications. Each chapter concludes with a summary of what was discussed in that chapter.

In the first section, the groundwork is laid for the rest of the book. Nakata begins in Chapter 1 by reflecting back on his own experience at different levels of his own language learning experience. He recounts times when he was really motivated at language learning as well as times when he was not motivated at all. With motivation being the central topic of the first section, great effort is used in Chapter 2 to try to identify what motivation is and if it can be defined at all. It is in this chapter that Nakata reveals that what this book is based on is the concept of social constructivism. Finally, in Chapter 3, the broad topic of motivation is narrowed down to primarily looking at language learning motivation. Attention is given to a brief history of language learning motivation and then to more recent studies that have tried to address its complexity.

In the second section on educational and cognitive psychology, Chapter 4 specifically takes up language learning motivation from an educational standpoint. Several theories are discussed to help the reader understand where the research in this book is going. These theories include the goal theory, self-determination theory, and the concept of learner autonomy. The discussion of these theories directly introduces the reader to Chapter 5, which attempts to provide a clearer picture of the language learning process as seen through a social constructivist view of learning and theories of social interaction. These theories help explain motivation in a deeper sense, which is very important when applying it to the context of Japanese English learners. Chapter 6 is completely dedicated to the Japanese learning context. Nakata explains that Japanese philosophy of thought is Confucian in nature, which is why Japanese students approach education and learning English the way they do. These aspects include the avoidance of public shame, respect for the teacher, making an effort, being persistent, and loyalty to the group (p. 152).

The third section, which includes Chapters 7 to 10, guides the reader from discussions in theory to recent research. In Chapter 7, this transition takes form. The first part of the chapter introduces the crucial elements which must be included in research of this kind. The last part of the chapter then outlines and explains the procedure that will be undertaken in the following chapters. Chapter 8 begins with a pilot study including a detailed account of the participants, materials, and results. The two research goals are a) to gain insight into the construct of foreign language motivation among first year university non-English-major EFL learners who study English as a part of their degree requirements, and b) to use the findings as a framework for an appropriate teaching approach with a sample of these learners (p. 203). The result of this pilot study formed

the basis of the more detailed research in the next chapter. Chapters 9 and 10 present the essence of Nakata's research. Chapter 9 explains the procedure of the project in great detail. Chapter 10 is a well-documented account of each of the five final participants involved in the research. The research focuses on the developmental process of the intrinsic motivation of each learner. The study employed both interviews and closed questionnaires to reveal past experiences in English language learning as well as to uncover changes in motivation. While this book is a good read for researchers who are currently employing these techniques, educators who are looking for practical suggestions and concrete ideas will find few in this book. In summary, it can be seen that some students like or dislike studying English for a variety of reasons. I respect the effort and time given to this research, but one would have to say that many changing factors which influence a student's motivation from day to day are out of an educator's control. It is very difficult if not impossible to accurately show the complexity of these findings in a graph or summary.

Chapter 11 serves as the conclusion to the book, and lists seven pedagogic principles. Nakata concludes by saying that this book may have only introduced the study of motivation and language learning and that there are still gaps to fill. Finally, Nakata provides a detailed bibliography, appendices of the materials he used, and two indexes—one on themes and the other of authors mentioned.

This book has inspired and challenged me to look at my own teaching in new ways. With reference to the numerous factors that influence students, how much do classroom dynamics play a role in motivation? Critically reflecting on my own teaching, I ask myself if there are things I say or do that leave a student with positive or negative impressions of English language learning.

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 practical and theoretical articles and research reports on second and foreign language teaching and learning in Japanese and Asian contexts. Submissions from other international contexts are accepted if applicable to language teaching in Japan. Areas of particular interest are:

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: (1) full-length articles, (2) short research reports (*Research Forum*), (3) essays on language education or reports of pedagogical techniques which are framed in theory and supported by descriptive or empirical data (*Perspectives*), (4) book and media reviews (*Reviews*), and (5) comments on previously published *JALT Journal* articles (*Point to Point*). Articles should be written for a general audience of language educators; therefore statistical techniques and specialized terms must be clearly explained.

Guidelines

Style

The *JALT Journal* follows the *Publication Manual of the American Psychological Association*, 5th edition (available from APA Order Department, P.O. Box 2710, Hyattsville, MD 20784, USA; by e-mail: <order@apa.org>; from the website: <<http://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://www.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 be not more than 10 pages in length. *Perspectives* submissions should not be 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 750 words. All submissions must be typed and double-spaced on A4 or 8.5"x11" paper with line spacing set at 1.5 lines. The author's name and identifying references should appear only on the cover sheet. Authors are responsible for the accuracy of references and reference citations.

Submission Procedure

Materials should be sent as an e-mail attachment in Rich Text Format (preferred) or by post. Postal submissions must include a clearly labeled floppy disk or CD-ROM and one printed copy. Please submit materials to the appropriate editor indicated below.

Materials to be submitted—all of this can be sent by e-mail in one attached RTF file (preferred) or by post. If by post, be sure to include all of the material on a floppy disk or CD-ROM.

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 and e-mail address.
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).
7. Authors of **accepted** manuscripts must supply camera-ready copies of any diagrams or figures and a copy of the manuscript (RTF or ASCII).

Evaluation Procedures

Submissions will be acknowledged within one month of their receipt. All manuscripts are first reviewed by the editorial board to insure 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 three months. The first author of a published article will receive one copy of the *Journal* and 10 off-prints with the option to order more at a set rate (contact JALT Central Office for price details). Review authors receive one copy of the *Journal*.

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. We regret that manuscripts or computer disks cannot be returned. 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, and Point to Point Submissions

Please send submissions in these categories or general inquiries to:

jj-editor@jalt-publications.org

Ian Isemonger; Incoming *JALT Journal* Editor

Perspectives

jj-editor2@jalt-publications.org

Darren Lingely, *JALT Journal* Associate Editor

Reviews

The editors invite reviews of books, tests, teaching systems, and other publications in the field of language education. A list of publications that have been sent to JALT for review is published monthly in *The Language Teacher*. Please send submissions, queries, or requests for books, materials, and review guidelines to:

jj-reviews@jalt-publications.org

Yuriko Kite, Ph. D; *JALT Journal* Reviews 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

Yoshinori J. Watanabe, Ph.D; *JALT Journal* Japanese-Language Editor

Address for Inquiries about Subscriptions, Author Reprints, 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)

E-mail: jco@jalt.org URL: <http://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 (5th 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。図表をいれる場合は、JALT Journalのページの サイズに合わせて作成して下さい(縮小コピー可)。。

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
- 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
- MS-Word で保存した本文(マッキントッシュ使用の場合はrtf ファイルで保存したもの)(執筆者は無記名のこと)
- 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: 編集委員会が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:

〒102-8554 上智大学 外国語学部 言語学副専攻
JALT Journal |日本語編集者 渡部良典
電話・Fax: 03-3238-4243
jj-editorj@jalt-publications.org

JALT Journal 第30巻 第1号

2007年4月20日 印刷
2007年5月 1日 発行
編集人 スティーブ・コーンウィル
発行人 スティーブ・ブラウン
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