

# jalt journal

The research journal of  
the Japan Association  
for Language Teaching

Volume 35 • No. 1 • May 2013



全国語学教育学会

Japan Association for Language Teaching

¥950 ISSN 0287-2420

# JALT Journal

Volume 35 • No. 1

May 2013

## Editor

Melodie Cook  
*University of Niigata Prefecture*

## Consulting Editor

Darren Lingley  
*Kochi University*

## Reviews Editor

Greg Rouault  
*Doshisha Women's College of Liberal Arts*

## Associate Editor

Anne McLellan Howard  
*Miyazaki International College*

## Consulting Editor

Greg Sholdt  
*Kobe University*

## Japanese-Language Editor

Ken Urano  
*Hokkai-Gakuen University*

## Editorial Board

William Acton  
*Trinity Western University, Canada*

David Aline  
*Kanagawa University*

David Beglar  
*Temple University—Japan Campus*

James Dean Brown  
*University of Hawai'i, Manoa, USA*

Charles Browne  
*Aoyama Gakuin University*

Yuko Goto Butler  
*University of Pennsylvania, USA*

Christine Pearson Casanave  
*Temple University—Japan Campus*

Eton Churchill  
*Kanagawa University*

Steve Cornwell  
*Osaka Jogakuin College*

Neil Cowie  
*Okayama University*

Anne Ediger  
*Hunter College, City University of New York, USA*

Peter Gobel  
*Kyoto Sangyo University*

Tim Greer  
*Kobe University*

Michael Guest  
*Miyazaki University*

Yuri Hosoda  
*Kanagawa University*

Ian Isemonger  
*Kumamoto University*

Yuriko Kite  
*Kansai University*

Setsuko Mori  
*Kinki University*

Tim Murphey  
*Kanda University of International Studies*

Cynthia Quinn  
*Kobe University*

Carol Rinnert  
*Hiroshima City University*

Gordon Robson  
*Showa Women's University*

Hidetoshi Saito  
*Ibaraki University*

Hideki Sakai  
*Shinshu University*

David Shea  
*Keio University*

Tamara Swenson  
*Osaka Jogakuin College*

Donna Tatsuki  
*Kobe City University of Foreign Studies*

Deryn Verity  
*The State University of Pennsylvania*

Yoshinori J. Watanabe  
*Sophia University*

Sayoko Yamashita  
*Meikai University*

Jack Yohay  
*Seifu Gakuen, Osaka*

**Additional Readers:** Nick Zhiwei Bi, Jim Elwood, Eric Hauser, Aek Phakiti, Tomohito Hiromori, Yo In'nami, Keita Kikuchi, Rie Koizumi, Ken Tamai, Yosuke Yanase. Akiko Takagi

**JALT Journal Production Editor:** Aleda Krause

**JALT Journal Proofreading:** Amy Brown, Susan Gilfert, Satoshi Inoue, Shusaku Kida, Aleda Krause, Joseph Sheehan, Alan Stoke, Jack Yohay

**JALT Publications Board Chair:** Darren Lingley

**JALT Journal Layout & Design:** Malcolm Swanson

**JALT Journal on the Internet:** <http://jalt-publications.org/jj/> **Website Editor:** Theron Muller

- 3 In this Issue
- 4 From the Editor

## Articles

- 5 Using Rasch Analysis to Create and Evaluate a Measurement Instrument for Foreign Language Classroom Speaking Anxiety — *Matthew T. Apple*
- 29 Teacher Cognition and Grammar Teaching in a Japanese High School — *Masateru Nishimuro and Simon Borg*
- 51 EFL Teachers' Learning: Transitional Experiences from an Overseas Teacher Education Program to Japanese School Settings — *Yuka Kurihara*
- 73 学内開発プレイズメントテスト得点解釈と使用の妥当性の評価について [Evaluating Validity for In-House Placement Test Score Interpretations and Uses] — 熊澤孝昭 (*Takaaki Kumazawa*)

## Research Forum

- 101 The Role of Think-Aloud and Metacognitive Strategies in L2 Meaning-Inference During Reading — *Megumi Hamada and Chaehee Park*

## Reviews

- 126 *The Developing Teacher* (Duncan Foord) — Reviewed by John Bankier
- 129 *Can Do Statements in Language Education in Japan and Beyond: Applications of the CEFR* (Maria Gabriela Schmidt, Naoyuki Naganuma, Fergus O'Dwyer, Alexander Imig, and Kazumi Sakai, Eds.) — Reviewed by Carol Begg
- 132 *English in Japan in the Era of Globalization* (Philip Seargeant, Ed.) — Reviewed by Theodore Bonnah
- 135 *Thinking About Language Teaching: Selected Articles 1982-2011* (Michael Swan) — Reviewed by Robert James Lowe
- 138 *Towards an Understanding of Language Learner Self-Concept* (Sarah Mercer) — Reviewed by James Rock
- 141 *Innovating EFL Teaching in Asia* (Theron Muller, Steven Herder, John Adamson, and Philip Shigeo Brown, Eds.) — Reviewed by Colin Skeates

## JALT Journal Information

- 145 Information for Contributors (English and Japanese)

# 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, all in Japan, along with 24 special interest groups (SIGs) and three forming SIGs. 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 semiannual research journal; *The Language Teacher*, a bi-monthly periodical containing articles, teaching activities, reviews, and announcements about professional concerns; and the annual *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 600 papers, workshops, colloquia, and poster sessions. Each JALT chapter holds local meetings and JALT's SIGs provide information and newsletters on specific areas of interest. JALT also sponsors special events such as workshops and conferences on specific themes, and awards annual grants for research projects related to language teaching and learning. Membership is open to those interested in language education and includes 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 <[www.jalt.org](http://www.jalt.org)>.

### JALT National Officers, 2013-2014

*President:* ..... Kevin Cleary  
*Vice President:* ..... Nathan Furuya  
*Auditor:* ..... Caroline Lloyd  
*Director of Treasury:* ..... Oana Cusen  
*Director of Records:* ..... Roehl Sybing  
*Director of Program:* ..... Steve Cornwell  
*Director of Membership:* ..... Judith Buzz Green  
*Director of Public Relations:* ..... Ted O'Neill

### Chapters

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

### Special Interest Groups

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

### JALT Central Office

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

# In this Issue

## Articles

We are pleased to publish four feature articles in this issue of *JALT Journal*. Our first contribution comes from **Matthew T. Apple**, who introduces Rasch analysis to applied linguistics literature in an original way. Our second article, written by **Masateru Nishimuro** and **Simon Borg**, offers an illuminating look at high school English teaching practices and will be very helpful for those teaching at the university level (particularly expatriates) to more fully understand their students' English educational backgrounds. **Yuka Kurihara**, who is known for her work on outsourced teacher education for Japanese teachers of English, introduces further work in our third article, this time using Vygotsky's sociocultural theory as a conceptual framework. In the fourth article, **Takaaki Kumazawa** reports on an attempt to validate the use of an in-house test for placement purposes employing an argument-based validity framework. As the author points out, validation is important even for low-stakes tests like those used as placement tests for university general education.

## Research Forum

In this section, **Megumi Hamada** and **Chaehee Park** examine the role of think-aloud as it pertains to lexical inferencing. While their research focuses on Korean learners of English, their findings are applicable to the Japanese context.

## Reviews

This issue features six book reviews. In the first one, **John Bankier** highlights a book of activities with universal appeal for teachers exploring professional development. In our second review, **Carol Begg** covers an edited collection of papers on can do statements and the application of the CEFR in Japanese contexts. The third review, by **Theodore Bonnah**, examines an edited title addressing how globalization impacts on how English is taught and used in Japan. Next, **Robert James Lowe** reviews a select collection of Michael Swan's papers on language teaching methodology that have been published over the past 30 years. In the fifth review, **James Rock** reports on a longitudinal case study into foreign language learner self-concept. **Colin**

**Skeates** closes out this section with another edited title, this one looking at diversity in teaching in Asian EFL contexts.

## From the Editor

Publication of this May issue marks my first as Editor of *JALT Journal*. While **Darren Lingley's** shoes have been a challenge to fill, I hope and believe he has prepared me well for the privilege of undertaking this task. I am happy to say that Darren is not leaving *JJ* entirely; he joins us as a Consulting Editor, taking over from Ian Isemonger, who has shown great devotion to *JALT Journal* as editor, reviewer, and author, and who has also given me help and advice. Ian will be staying in the *JJ* family as a member of the EAB. I am pleased to have the assistance of **Anne McLellan Howard** as Associate Editor; she brings a wealth of experience and, more importantly, a sense of humour, to the post. Words can't express how grateful I am to the members of the Editorial Advisory Board and additional readers who referee submissions with, in Darren's words, "just the right blend of scholarly expertise, sensitive criticism, and peer encouragement." I am especially indebted to those EAB members (you know who you are) who are willing to accept a steady stream of review requests and process them in a timely and good-humoured manner. I also offer my thanks to all the authors who have chosen *JALT Journal* as the home for their work; without them, we are nothing. Last, but not least, I must also thank **Aleda Krause** and her *Journal* production team who have been ever-so-gently breaking me in. I am enjoying my tenure as Editor of *JALT Journal* and promise to do my utmost to keep up the high standards set by my predecessors.

Melodie Cook

# Articles

## Using Rasch Analysis to Create and Evaluate a Measurement Instrument for Foreign Language Classroom Speaking Anxiety

Matthew T. Apple  
*Ritsumeikan University*

Despite the existence of skill-specific anxiety instruments measuring reading anxiety, writing anxiety, and listening anxiety, there is still no single measurement instrument specifically designed to measure levels of speaking anxiety. This research had two purposes. The first was to provide for classroom-based foreign language teachers and researchers an example of the advantages of Rasch model analysis, the use of which is increasing in first-language educational contexts for measurement instrument creation and validation. The second purpose was to create and evaluate an instrument for measuring foreign language speaking anxiety within the classroom in an EFL learning context, in which few native speakers of the language are available for interaction. Using data from a sample of Japanese university students ( $N = 172$ ), the Rasch model identified misfitting items and examined the construct validity of a 20-item questionnaire to measure levels of Foreign Language Classroom Speaking Anxiety (FLCSA).

リーディング、ライティング、リスニングといった特定のスキルに関する不安を測定する手段はあるが、スピーキング不安のレベルを測定する手段は現在のところ存在しない。本研究の目的は二つある。一つ目は、教室で教える外国語教師や研究者にラッシュモデル分析の利点の例を示すことである。ラッシュモデルによる分析は、第一言語の教育環境において、測定手段の作成やその妥当性を高める目的でますます使用されるようになっている。本研究の二つ目の目的は、

母語話者とのやりとりの機会がほとんどない「外国語としての英語」を学ぶ教室において、外国語スピーキング不安を測る手段を作成し、評価することである。日本の大学生(N=172)のデータを使用し、外国語スピーキング不安(FLCSA)の度合いを測定する20項目からなる質問紙からラッシュモデルにより不適当な項目を割り出し、構成概念妥当性を検証した。

**A**nxiety has attracted a considerable amount of attention in SLA research (MacIntyre, 1999; Scovel, 2001; Young, 1999). Although some SLA researchers have insisted that L2 anxiety was primarily due to poor language skills in a student's L1, the so-called "linguistic coding deficit hypothesis" (Sparks & Ganschow, 1991, 1995; Sparks, Ganschow, & Patton, 1995), other researchers have persuasively argued that anxiety negatively affects cognitive processes and thus is not a result but a cause of performance (Gardner & MacIntyre, 1993; MacIntyre, 1995a, 1995b; MacIntyre & Gardner, 1991). Some researchers (e.g., Brown, Robson, & Rosenkjar, 2001; MacIntyre & Gardner, 1994) indicated the possible benefits of "facilitative" anxiety. Whether facilitating or debilitating, anxiety is a critical concern for foreign language instructors in the classroom setting.

Before dealing effectively with anxiety in the classroom, foreign language instructors need to measure levels of anxiety among their learners. Anxiety itself was originally conceived of as a stable trait variable that holds true in any social situation (Eysenck, 1970; Taylor, 1953), but quickly became categorized into two forms by Spielberger (1983): trait anxiety (constant across contexts) and state anxiety (changeable according to circumstances). Foreign language anxiety is generally considered a state-related, situation-specific anxiety that arises only within foreign language contexts (MacIntyre, 1999; 2007). As such, questionnaire instruments<sup>1</sup> designed to measure foreign language anxiety typically focus on anxiety as it occurs within the foreign language classroom, the most salient learning context for language learners.

Quantitatively oriented foreign language anxiety studies typically use short two-, four-, and six-item Likert-type questionnaire instruments (e.g., MacIntyre, Babin, & Clement, 1999; MacIntyre & Charos, 1996; MacIntyre, Noels, & Clément, 1997) or the 33-item Foreign Language Classroom Anxiety Scale (FLCAS) that measures a combination of communication apprehension, fear of negative evaluation, and test anxiety (e.g., Elkhafai, 2005; Gregersen, 2007; Gregersen & Horwitz, 2002; Horwitz, Horwitz, & Cope, 1986; Mak, 2011; Matsuda & Gobel, 2004). Skill-specific measurements of foreign language anxiety created through correlation to previous questionnaires include reading anxiety (Saito, Garza, & Horwitz, 1999), writing



anxiety (Cheng, Horwitz, & Schallert, 1999), and listening anxiety (Kimura, 2008). However, there is still no foreign language anxiety questionnaire instrument designed specifically to measure classroom speaking anxiety. SLA anxiety researchers have noted that language anxiety development seems particularly linked to the fear of speaking in front of one's peers (Cohen & Norst, 1989; MacIntyre, 1999; Price, 1991), making the lack of a foreign language speaking anxiety questionnaire all the more surprising.

### **Drawbacks to Traditional Statistical Methods of Measuring Anxiety**

Many previous instruments measuring foreign language anxiety contained items related to various aspects of anxiety. Many researchers claim to have "validated" responses gathered from such instruments through the use of Cronbach's alpha reliability estimates, factor analyses, and correlation to data obtained from other questionnaires. However, the use of existing foreign language anxiety measurement instruments without detailed examination of the data scores generated for specific student samples may have resulted in four potentially misleading assumptions.

The first assumption is that each item in a questionnaire contributes equally to the measurement of the construct<sup>2</sup>, regardless of whether some items are easier to agree or disagree with than others (Bond & Fox, 2007, p. 120). For example, when many instructors want to determine the level of anxiety of their students, they often assume that the "3" of one item means a "3" on another item and that by simply adding raw Likert scores, the resulting combined anxiety score will accurately represent the amount of anxiety that each student experiences. A second, related assumption is that all items share equal endorsement difficulty; that is, that all items are equally easy or difficult to agree with. However, answering "agree" to one item may actually be more difficult than answering "agree" to another item. Adding raw scores from Likert-type categories treats the items as having the same endorsement difficulty, potentially leading to erroneous interpretations of the data (Wolfe & Smith, 2007a). Thus, scores obtained from summing Likert-scale responses may misrepresent the anxiety present in the classroom.

The third assumption is that existing questionnaire instruments measure unidimensional constructs. A measurement instrument must measure a single construct before it can be deemed reliable (Thurstone, 1931; Wright, 1999); construct validity is thus a prerequisite of reliability. SLA researchers whose questionnaire instrument items have not been checked for construct unidimensionality cannot be certain that the data obtained from

the questionnaire instrument represents the psychological construct that the researcher intended the item to measure. In other words, while foreign language anxiety may indeed be a multidimensional *concept*, each individual *construct* that comprises foreign language anxiety (i.e., reading, testing, listening, speaking) should produce scores with demonstrable unidimensional properties. This is true even when using exploratory factor analysis (EFA), a procedure based on correlational analysis, because EFA is designed to find multiple constructs within data and cannot determine the construct validity for individual constructs (Wright, 1996a). Thus, prior to any analysis of questionnaire instruments that consist of multiple constructs, each individual construct should be created and evaluated independently to determine the ability of items to measure the intended construct.

A final assumption is that Likert-type categorical data are true interval data. The treatment of categorical data raw scores as interval data rather than ordinal data may help explain why results from questionnaire studies have proved difficult to reproduce across different samples (Wright, 1999). An alternative method of analysis, which transforms Likert-type categorical data from ordinal data into interval data from which a linear construct can be mathematically extracted, is the Rasch model (Rasch, 1960).

### **The Advantages of Rasch Model Analysis**

The Rasch model is a unidimensional measurement model that uses respondents' positive endorsements (responding "agree") or negative endorsements ("disagree") of questionnaire items to calculate the relationship between the amount, or *levels*, of the construct present in each of the respondents (called "persons") and the *endorsability difficulty level* of questionnaire items, that is, the degree to which respondents find it difficult to agree or disagree with items. This relationship is expressed in log-odds, or *logits* (Embretson & Reise, 2000). Data are fit to the Rasch model by mathematically transforming raw scores on items into interval data and then by placing both persons and items on the same linear scale for comparison (Bond & Fox, 2007; Wright, 1999).

Analysis of Likert-type categorical data using the Rasch Rating Scale model (Andrich, 1978) offers several advantages over traditional analysis. The first advantage is the use of *fit* to demonstrate the quality of both persons and items measured by the hypothesized construct. By identifying misfitting person responses and items that do not contribute to the construct being measured, Rasch model analysis can assist the researcher in revising the

questionnaire instrument in order to provide a more accurate estimation of the construct (Wolfe & Smith, 2007a). A second advantage is that, whereas a typical analysis, such as Cronbach's alpha reliability estimates, only shows the consistency of person responses (Sijtsma, 2009), Rasch analysis provides reliability figures both for person responses and for questionnaire items. Additionally, Rasch model analysis uses *separation*, which shows the number of different groups within the sample and the number of different item difficulty levels (Fisher, 1992; Wright, 1996b).

A third advantage is that the Rasch model uses the concept of *measure* to indicate the level of the construct within each questionnaire respondent as well as the item *endorsability difficulty level*, in other words, the degree to which respondents find it difficult or easy to agree with items (Smith, E. V., 2001). Higher person measures indicate greater amounts of the construct present in respondents, and higher item measures indicate items that are more difficult to endorse. A fourth, related advantage is the use of an *item-person map* to display the relative levels of the construct for each person and relative endorsability difficulty level of each item on a single scale. The levels of construct for each questionnaire respondent and the endorsability difficulty level of each questionnaire item are displayed side by side on the same logit scale for easy visual comparison (Wilson, 2005, p. 96).

Finally, the use of Rasch principal components analysis (PCA) of item residuals can demonstrate the degree to which all items demonstrate coherence to a single dimension, thus satisfying the one-construct, one-dimension criterion of measurement theory (Thurstone, 1931). Rasch PCA of the item residuals provides an estimation of internal construct validity by examining not only the items that load onto the hypothesized construct but also the error left over from extracting the construct from the data (Waugh & Chapman, 2005; Wright, 1996a). The use of Rasch analysis to evaluate the data obtained from questionnaires can thus provide the researcher with a wealth of information about the quality of the questionnaire items that the researcher is using to measure latent psychological traits such as foreign language speaking anxiety in the classroom.

## Purpose of This Study

In this paper I seek to demonstrate the advantages of using Rasch analysis for the creation and evaluation of a measurement instrument specifically aimed at the construct of foreign language classroom speaking anxiety. Although the sample in this study is based in Japan, it is hoped that the result-

ing questionnaire will become a beneficial measurement resource for others as well.

There were two research goals for this study:

1. To examine how SLA instructors and researchers can use Rasch model analysis to create and evaluate Likert-type category questionnaire instruments, and
2. To evaluate the degree to which the levels of foreign language classroom speaking anxiety in a typical four-skills EFL classroom in Japan can be measured with a questionnaire instrument that has been analyzed and evaluated using Rasch analysis.

## Materials and Methods

### *Participants*

There were initially 172 participants in this study. All were 1st-year students in a large private university in western Japan in six intact four-skills English classes that met twice per week for 90 minutes and were taught by native speakers of Japanese. The average class size was 30; the smallest was 25 and the largest was 38, which is representative of typical Japanese university EFL classrooms. Eighty of the participants were economics majors, 69 were engineering majors, 22 were English literature majors, and one participant did not give a major. There were 49 females, 122 males, and one participant who did not indicate a sex.

### Measurement Instrument

As a sample of how researchers can use Rasch analysis to create and evaluate questionnaires, in this study I examined the Foreign Language Classroom Speaking Anxiety Scale (FLCSAS). The original FLCSAS consisted of 20 items related to typical speaking situations within a communicative English classroom in Japan (see Appendix). Items had been previously piloted with a smaller sample ( $N = 116$ ) and modified prior to the analysis in this study (Apple, 2008). Items that measured participants' speaking anxiety in different communicative situations in the foreign language classroom (Items 2, 4, 6, 8, 13, 14, 16, 17, 18, 19, and 20) were based on items from the Personal Report of Communication Apprehension (PRCA-20; McCroskey, 1977, 1978). Items concerned with speaking to the teacher (Items 3, 5, 7, 10, and 15) as well as general speaking anxiety items (Items 1, 9, 11, and 12) were based on items from the FLCAS (Horwitz et al., 1986).

The questionnaire items were translated into Japanese by a native speaker of Japanese and back-translated by another native speaker of Japanese to confirm accuracy of statement. A 6-point Likert-type category scale was employed with the end points of the scale labeled (1 = “Strongly disagree” and 6 = “Strongly disagree”). Other points were not labeled, and there was no middle or neutral option.

### ***Analysis Procedure***

The questionnaire instrument was implemented during class time in the middle of the spring semester. Data were analyzed with Winsteps 3.63 software (Linacre, 2006) using the Rasch Rating Scale model for categorical data (Andrich, 1978). Rasch analysis consisted of Rasch person and item fit analysis, Rasch item-person maps, Rasch PCA of item residuals, and disattenuated person measures correlational analysis.

#### *Person Fit and Item Fit Analysis*

To determine the fit of persons and items to the construct, Rasch analysis produces both *infit* and *outfit* statistics, which have two forms: one unstandardized (*mean squares*) and one standardized (*z-scores*) (Linacre, 2002). Infit statistics are weighted to give more information about persons who are at or near item endorsability difficulty levels; that is, questionnaire respondents whose probability of endorsing items is similar to the difficulty of endorsing items, thus giving insight into the item quality. Outfit statistics are not weighted and are more easily affected by respondents who find items too easy or too difficult to endorse (i.e., statistical outliers), and thus provide less useful information about the majority of questionnaire respondents. Researchers typically pay more attention to *infit* in the interests of determining the quality of items as they apply to the majority of respondents (Bond & Fox, 2007, p. 57).

A mean-square fit statistic of 1.0 indicates perfect fit. Person and item responses that misfit the model may be the result of carelessness, response set answering, or item bias and may need further examination to determine whether they are contributing to the construct as intended (Wolfe & Smith, 2007b, p. 211). For this study, misfit was defined as less than 0.50 mean-squares or -3.0 z-scores, or greater than 1.50 mean-squares or 3.0 z-scores, based on the recommended criteria of Linacre (2007).

The questionnaire instrument’s reliability was estimated using four statistics: (a) person reliability (to determine the consistency of person

responses), (b) person separation (to estimate the ability of the instrument to separate participants into different levels of the construct), (c) item reliability (to estimate how well the items cohered), and (d) item separation (to estimate the ability of the participants to distinguish between items measuring different levels of the construct) (Wright & Masters, 2002).

### *Item-Person Maps*

Item maps were requested as part of item fit analysis to provide a visual representation of person and item locations on the construct. Persons are located along the linear scale based on the level of construct (i.e., the amount of construct present in individual respondents), while items are located based on their endorsement difficulty level (i.e., degree of difficulty of answering “agree” or “disagree”). A person has a 50% chance of endorsing an item located at the same level of the construct on the opposite side of the vertical line (Bond & Fox, 2007). Items located above the person are more difficult to endorse and items located below the person are easier to endorse.

### *PCA of Item Residuals*

A Rasch PCA of item residuals was conducted to examine the unidimensionality of the construct in two ways. First, a unidimensional construct should account for at least 50% of the total variance in the data. Second, the principal contrast, the residual errors left over after the linear construct is extracted, should represent uncorrelated error with an eigenvalue of less than 3.0 and less than 10% of the variance (Linacre, 2007). If the principal contrast has a greater eigenvalue and variance than these criteria, there may be an additional, unwanted construct present in the data that the items were not meant to measure. Item residuals on the main dimension of the construct are termed “positively loading items,” while items on a possible secondary dimension are termed “negatively loading items” (Wright, 1996a).

### *Disattenuated Person Measures Correlational Analysis*

Disattenuation refers to the process by which the Rasch model takes into account the error of measurement when transforming raw, ordinal scores from questionnaire item responses into true interval measures. The transformed responses are referred to as *disattenuated person measures*, which can be used in the place of raw scores in correlational analysis. The correlational analysis of disattenuated person measures serves two functions.

The first is to verify whether the removal of items that misfit the model will adversely affect the ability of the questionnaire instrument to measure the level of the latent construct of questionnaire respondents (Smith, R. M., 2000). A strong correlation ( $r > .7$ ) indicates that the items are measuring the same construct. The second is to provide a further verification of construct validity by correlating person measures from positively and negatively loading items in the Rasch PCA of item residuals (Smith, E. V., 2002). A strong correlation ( $r > .7$ ) suggests construct unidimensionality and thus supports claims of construct validity.

## Results

### *Person Fit Analysis*

The data acquired from the questionnaire instrument were input into the Rasch model and Rasch person fit analysis was conducted to determine whether all participant responses fit the model's expectations. The Rasch reliability of person responses was estimated at .88, with a Rasch person separation of 2.88. Based on the fit criteria, the responses of 20 persons were found to systematically misfit the model on all questionnaire items. Examination of the misfitting person responses showed the existence of set response patterns and repeated extreme response scores, indicating possible carelessness or lack of seriousness in answering the questionnaire. Because extreme scores add measurement error and adversely affect item fit and unidimensionality of construct, the misfitting person responses were excluded from further analysis, making an adjusted  $N$ -size of 152. The data were input into the Rasch model again; further examination of person fit indicated no misfitting person responses. The revised Rasch person reliability was .90 and the person separation was 3.02. The Rasch person separation of above 3.0 indicated the ability of the measurement instrument to stratify person responses into at least four separate groups, or levels, across the construct.

For comparison, traditional descriptive statistics were calculated based on raw response scores (Table 1). Cronbach's alpha was calculated at  $\alpha = .93$ , indicating that Rasch person reliability was a more conservative estimate.

**Table 1. Descriptive Statistics for Items Measuring the Foreign Language Classroom Speaking Anxiety Construct**

Item	Item description	<i>M</i>	<i>SD</i>
A1	I'm worried that other students in class speak better than I do	3.29	1.57
A2	I feel nervous speaking in front of the entire class	4.03	1.51
A3	I tremble when the teacher is about to ask me a question	2.26	1.21
A4	I am reluctant to express my opinion in a group	2.50	1.25
A5	I'm worried about making mistakes when I speak with the teacher	2.73	1.46
A6	I'm worried that my partner speaks better English than I do	2.69	1.42
A7	I am reluctant to ask the teacher a question	2.57	1.48
A8	I start to panic when I speak with a classmate in a pair	2.03	1.14
A9	I dislike speaking entirely	2.09	1.27
A10	I'm worried that the teacher will think my speaking is no good	2.37	1.37
A11	I'm worried about making mistakes while speaking	2.58	1.52
A12	I feel nervous when I can't express my opinion	3.17	1.57
A13	I'm afraid my partner will laugh when I speak with a classmate in a pair	1.89	1.09
A14	I'm worried about making mistakes when I speak with a partner	2.02	1.21
A15	Answering a teacher's question in class is embarrassing	2.14	1.22
A16	Speaking in a group of classmates makes me feel self-conscious	2.26	1.27
A17	I feel tense when I have to speak with a classmate in a pair	2.02	1.10
A18	I start to panic when I have to speak in a group	1.96	1.08
A19	I'm afraid that others in a group discussion will laugh if I speak	2.95	1.62
A20	I can feel my heart pounding when it's my turn to speak in a group	1.82	1.07

*Note.* A Likert scale from (1) *Strongly disagree* to (6) *Strongly agree* was used.



### Item Fit Analysis

After person fit analysis was concluded, Rasch item fit analysis was conducted (Table 2). The Rasch item reliability was .98, demonstrating the strong coherence of the items. The Rasch item separation was 6.41, indicating that the study participants were able to distinguish approximately six different levels of the construct as measured by the items. Item fit statistics showed that several items (Items 1, 2, 7, 13, 15, 18, and 19) were well above the  $\pm 3.0$  z-score criterion; however, only three of these were also outside the mean-squared criteria of 0.5 and 1.5 (Items 1, 7, and 19).

**Table 2. Rasch Item Analysis for Items Measuring the Foreign Language Classroom Speaking Anxiety Scale**

Item	Measure	Error	Infit		Outfit	
			MNSQ	Infit ZSTD	MNSQ	Outfit ZSTD
A16	.90	.11	1.30	2.1	1.25	1.5
A13	.72	.10	.59	-3.7	.53	-3.7
A19	.57	.10	.47	-5.2	.43	-5.0
A14	.50	.09	.79	-1.8	.73	-2.0
A18	.47	.09	.55	-4.3	.53	-4.0
A9	.46	.09	1.27	2.0	1.29	1.9
A8	.41	.09	.91	-.7	.80	-1.5
A15	.38	.09	.66	-3.1	.63	<b>-3.0</b>
A3	.21	.09	.93	-.6	.99	.0
A17	.18	.09	.81	-1.7	.81	-1.5
A10	.03	.08	.90	-.8	.89	-.9
A4	-.07	.08	1.01	.2	1.10	.9
A7	-.07	.08	1.72	5.2	1.90	5.8
A11	-.16	.08	.94	-.5	.87	-1.1
A6	-.33	.08	1.02	.2	1.06	.6
A5	-.35	.08	1.00	.1	1.09	.8
A20	-.65	.08	1.29	2.5	1.30	2.4
A12	-.79	.08	1.00	.0	1.01	.1
A1	-.82	.08	1.52	4.1	1.66	4.9
A2	-1.61	.08	1.46	3.7	1.45	3.5

*Note.* MNSQ = mean-squared; ZSTD = standardized z-scores. Measures are in Rasch logits. ZSTD misfit is indicated by boldface; MNSQ misfit is indicated by italics.  $N = 152$ .

The items misfitting both sets of criteria were removed and the analysis was run again. However, there was no significant difference in person separation (before = 3.02; after = 2.98) or item separation (before = 6.41; after = 6.68), meaning in practical terms that the range of persons and items across the construct remained essentially the same. Both person and item reliability remained unchanged. To examine changes in construct validity, a correlational analysis of disattenuated person measures was conducted with person measures obtained from before removing misfitting items and from after removing the items. The two sets of person measures were strongly correlated,  $r = .99$ ,  $p < .01$ , signifying that the misfitting items could be removed without adversely affecting the measurement. For the time being, misfitting items were retained for further confirmation in the PCA of residuals analysis.

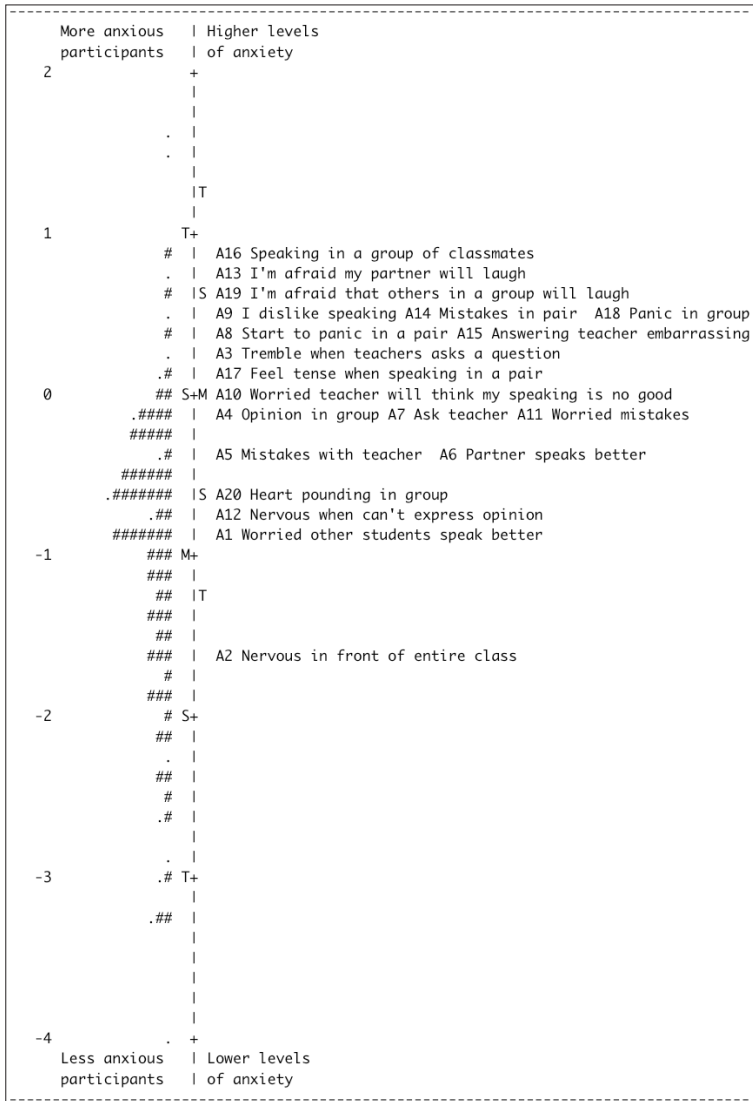
### ***Item-Person Map***

An item-person map was obtained as a visual representation of person and item difficulty levels along the construct (Figure 1). The most difficult item to endorse was Item 16 ("Speaking in a group of classmates makes me feel self-conscious," Rasch item difficulty measure = .90), and the item easiest to endorse was Item 2 ("I feel nervous speaking in front of the entire class," Rasch item difficulty measure = -1.61).

To give a clearer visual picture of levels of the construct for each study participant, the item-person map was reconfigured to show individual participants along the linear scale (Figure 2). The item-person map in Figure 2 shows items on the left side and individual study participants minus the 20 misfitting persons who were removed, labeled 1 through 172, on the right side. Even though the mathematical mean of the person ability estimates was slightly lower than Item 1 ("I'm worried that other students in class speak better than I do") on the left side of the linear scale, the median of persons was the same level as Item 20 ("I can feel my heart pounding when it's my turn to speak in a group"), suggesting that Item 20 was the best item to distinguish between low and high levels of foreign language classroom speaking anxiety among the participants.

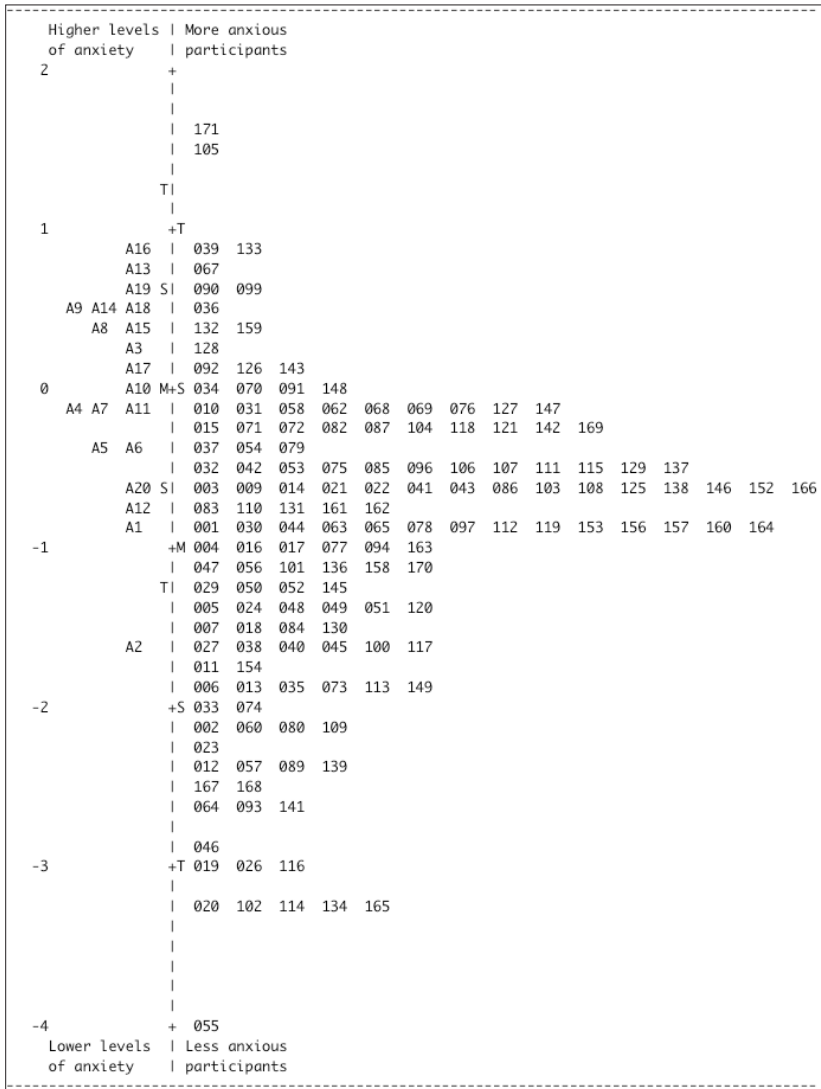
### ***PCA of Item Residuals***

Following person and item fit analyses, a Rasch PCA of item residuals was conducted for all 20 items (Table 3). Results indicated that 71.0% of the variance (eigenvalue = 49.1) was explained by the construct. The principal



**Figure 1. Item-Person Map for Items Measuring the Foreign Language Classroom Speaking Anxiety Scale, Including Misfitting Items**

*Note.* Each # represents two participants. Items more difficult to endorse and thus indicative of higher levels of anxiety are near the top of the scale. M = mean; S = one standard deviation from the mean; T = two standard deviations from the mean.



**Figure 2. Item-Person Map for Persons Along The Foreign Language Classroom Speaking Anxiety Scale**

*Note.* Study participants with higher levels of anxiety are located near the top of the scale. M = mean; S = one standard deviation from the mean; T = two standard deviations from the mean.

contrast explained 3.9% of the variance (eigenvalue = 2.7). The variance accounted for by the construct was above the minimum criterion of 50%, and the variance accounted for by the correlated residuals in the principal contrast were below the criterion of 10%, indicating that the items demonstrated strong construct validity. Disattenuated person measures from items that loaded positively and negatively onto the principal contrast were correlated,  $r = .74, p < .01$ , as an additional support of construct validity.

**Table 3. Rasch Principal Components Analysis of Item Residuals of the Principal Contrast to the Foreign Language Classroom Speaking Anxiety Scale**

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
A19	<b>.68</b>	.57	.47	.43
A18	<b>.64</b>	.47	.55	.53
A17	<b>.55</b>	.18	.81	.81
A13	<b>.52</b>	.72	.59	.53
A14	<b>.52</b>	.50	.79	.73
A8	.32	.41	.91	.80
A15	.28	.38	.66	.63
A11	.28	-.16	.94	.87
A16	.21	.90	1.30	1.25
A2	<b>-.50</b>	-1.61	1.48	1.45
A4	<b>-.40</b>	-.07	1.01	1.10
A7	-.35	-.07	1.72	1.90
A1	-.26	-.82	1.52	1.66
A12	-.16	-.79	1.00	1.01
A6	-.15	-.33	1.02	1.06
A9	-.13	.46	1.27	1.29
A20	-.10	-.65	1.29	1.30
A5	-.08	-.35	1.00	1.09
A3	-.07	.21	.93	.99
A10	-.01	.03	.90	.89

*Note.* Measures are in Rasch logits. *Positive loading* items above the line indicate the main construct identified by the Rasch model and *negative loading* items below the line indicate a possible secondary dimension to the main construct. Item loadings above .40 are in boldface. MNSQ = mean-squared.

Examination of the item loadings on the principal contrast in Table 3 revealed that five items (Items 13, 14, 17, 18, and 19) had high positive loadings above .40, and two items (Items 2 and 4) had high negative loadings above -.40. Because Item 19 ("Others in a group will laugh...") had been previously identified as a misfitting item, there was a possibility that the item was exerting undue influence on other positive loading items in the construct. Therefore, the item was temporarily removed and the PCA of item residuals was conducted again with the remaining items. However, the variance accounted for by the construct was not improved (before = 71.0%; after = 70.3%). The other misfitting items (Item 1, "Others students speak better..." and Item 7, "Reluctant to ask the teacher...") were similarly examined. The variance accounted for was improved in both instances; after the removal of Item 1, the variance accounted for was 73.6%, while with Item 7 removed, the variance accounted for was improved to 74.3%.

The PCA of item residuals indicated that three other teacher-related items (Items 3, 5, and 10) had extremely low loadings of -.07, -.08, and -.01, respectively, indicating that the items did not strongly cohere to other items measuring the construct. When all five of the teacher-related items (Items 3, 5, 7, 10, and 15) were removed, the resulting variance accounted for improved (77.0%, eigenvalue = 50.3), and the principal contrast accounted for 3.9% of the variance (eigenvalue = 2.6). The disattenuated person measures from positively and negatively loading items were correlated again and the result was stronger ( $r = .89$ ) than prior to removing the teacher-related items ( $r = .74$ ). Thus, Rasch analysis indicated that teacher-oriented items could be removed to improve construct validity for the study sample (Table 4).

## Discussion

At first glance, the descriptive statistics in Table 1 seem to match the Rasch model analysis results; for example, Item 2 had the highest mean score ( $M = 4.03$ ), and thus was the easiest with which students agreed. In the Rasch analysis, Item 2 was likewise the easiest to endorse (Rasch item difficulty measure = -1.61). However, while Item 20 had the lowest mean score in the descriptive statistics ( $M = 1.82$ ) and thus was the most difficult item with which to agree, Item 20 was not the most difficult to endorse in the Rasch analysis. In fact, the results of Rasch model analysis showed Item 20 to be relatively easy to endorse (Rasch item difficulty measure = -.65). The most difficult item for students to endorse, as measured by Rasch analysis, was Item 16; in the descriptive statistics, this item was relatively difficult for students to agree with ( $M = 2.26$ ). This demonstrates how a reliance on mean

**Table 4. Rasch Principal Components Analysis of Item Residuals of the Principal Contrast to the Foreign Language Classroom Speaking Anxiety Scale Without Teacher-Related Items**

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
A19	<b>.67</b>	.60	.47	.43
A18	<b>.64</b>	.50	.51	.48
A17	<b>.52</b>	.53	.81	.79
A13	<b>.52</b>	.76	.60	.55
A14	<b>.47</b>	.21	.75	.74
A8	.31	.44	.89	.78
A11	.30	-.15	1.02	.92
A16	.24	.95	1.40	1.28
A2	<b>-.51</b>	-1.65	1.55	1.52
A1	-.35	-.83	1.52	1.64
A4	-.34	-.06	1.07	1.20
A6	-.26	-.32	1.02	1.04
A12	-.24	-.81	1.02	1.00
A20	-.23	-.66	1.28	1.30
A9	-.11	.49	1.30	1.32

*Note.* Measure is in Rasch logits. *Positive loading* items above the line indicate the main construct identified by the Rasch model and *negative loading* items below the line indicate a possible secondary dimension to the main construct. Item loadings above .40 are in boldface. MNSQ = mean-squared.

scores to judge which items are the best indications of levels of a psychological variable within a participant population may be potentially misleading.

Likewise, although the questionnaire data had a high Cronbach's alpha of .93, this figure was not helpful in determining to what degree the items measured the construct, or to what degree the persons were ranged along the construct. Researchers can use Rasch analysis to take into account measurement error, item location, person location, and fit statistics to better determine the degree to which speaking anxiety levels exist for individual students as well as to determine which questionnaire items were the best indicators of speaking anxiety. Thus, the first research goal, of examining how researchers can use Rasch analysis to create and evaluate questionnaire instruments, has been demonstrated: Using Rasch analysis of the example questionnaire data in this paper, we were able to determine per-

son fit, item fit, the degree to which items contributed to the construct, the amount of variance accounted for by the items, and the degree of construct unidimensionality. The Rasch analysis results provide valuable information that a researcher could use to revise such a questionnaire instrument prior to further implementations with other learner population samples.

The second research goal, regarding the measurement of different levels of foreign language classroom speaking anxiety among the study participants, was also reached. Rasch person separation (3.02) indicated the presence of four levels of anxiety within the sample, while the Rasch item-person map indicated the location of individual persons along the construct. While the identification of three items (Items 1, 7, and 19) as misfitting the intended construct on both infit mean-squared and standardized z-scores provided a useful starting point for the examination of construct validity in the PCA of item residuals, an examination of Item 7 proved most useful, as subsequent removal of all five teacher-related items in the PCA of item residuals resulted in more variance accounted for by the construct. Additionally, the location of three teacher-oriented items (Items 5, 7, and 15) next to items of equal endorsement difficulty level on the item-person map in Figure 1 indicated item redundancy.

Finally, the item fit (Table 2) indicated only a .03 logit difference in item difficulty measure between a fourth teacher-oriented question, Item 3 (“I tremble when the teacher is about to ask me a question”), and Item 17 (“I feel tense when I have to speak with a classmate in a pair”). Thus, the Rasch analysis indicated that the five teacher-related items did not contribute as much to the construct as other items and could be eliminated to improve measurement of the construct.

By way of comparison, Cronbach’s alpha would slightly decrease ( $\alpha = .92$ ) if one were to remove the teacher-related items, leading to the potentially erroneous conclusion that the items should be kept. The use of Rasch model analysis in this paper demonstrates that over-reliance on Cronbach’s alpha reliability estimates of internal consistency, in which more items and more participants equals better reliability, does not necessarily lead to a better questionnaire (Nunnally, 1978; Sijtsma, 2009).

## **Conclusion**

Results from this study demonstrate that using Rasch analysis can identify items that may be inappropriate for the sample population. The results indicated that the questionnaire performed adequately as a measure of foreign



language classroom speaking anxiety. Results obtained from Rasch item fit analysis, item-person map, and Rasch PCA of item residuals revealed that the questionnaire items regarding communication with the teacher did not contribute to the intended construct. From these results it can be inferred that, for students in the sample, questionnaire items concerning direct interaction with the teacher were not as salient as other items pertaining to communication among classmates for determining speaking anxiety levels as measured by the questionnaire.

The analyses also indicated room for improvement before future implementations. The measurement instrument presented in this study as an example of how Rasch analysis results can assist researchers in creating and evaluating their own questionnaire instruments was part of a pilot study for a larger study (Apple, 2011), in which 11 of the original 20 items were retained (see Appendix). A shorter version was also used in two recent studies (Apple, Falout, & Hill, 2012, in press). In each case, items from the FLCSAS were reevaluated for the participant sample to determine the effectiveness of items to measure the theoretical construct of foreign language classroom speaking anxiety. When implementing a previously created and tested measurement instrument, researchers should keep in mind that instruments do not “have” a certain reliability; in that sense, questionnaire instruments should always be evaluated to check the fit of the items to the construct and the fit of the participants to the questionnaire. Rasch analysis provides researchers with a full set of evaluation tools to help them improve existing questionnaires and create their own if needed.

## Notes

1. The terms *questionnaire instrument* and *questionnaire measurement instrument* are preferred to *questionnaire* in this paper. In the quantitative measurement tradition, any method of obtaining and evaluating responses of persons that “express their achievements, attitudes or personal points of view” is a measurement instrument (Wilson, 2005, p. 5). Instrument is thus a more precise term than questionnaire when attempting to collect, measure, and summarize participant responses in a questionnaire survey study.
2. The term *construct* is used in this paper rather than *factor* to represent a “theoretical concept that explains observable behaviors and refers to assumed latent (unobservable) characteristics of respondents” (Wolfe & Smith, 2007a, p. 106).

## Acknowledgements

The author would like to thank the university language instructors who assisted in data collection and David Beglar of Temple University who gave much statistical advice regarding the Rasch model.

*Matthew Apple* has an MFA (University of Notre Dame) and an MEd and EdD (Temple University) and has taught English at various educational levels in Japan since 1999. His research interests include CALL and individual differences in SLA and his teaching interests include academic writing and presentation.

## References

- Andrich, D. (1978). Scaling attitude items constructed and scored in the Likert tradition. *Educational and Psychological Measurement, 38*, 665-680.
- Apple, M. (2008, May). *Developing an anxiety Rasch*. Paper presented at the PanSIG 2008 Conference, Kyoto.
- Apple, M. (2011). *The Big Five personality traits and foreign language speaking confidence among Japanese EFL students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3457819)
- Apple, M., Falout, J., & Hill, G. (2012). The L2 motivational selves of technical college students. In H. Terai (Ed.), *Proceedings of the 6th International Symposium on Advances in Technology Education* (pp. 189-194). Retrieved from [http://www.isate2012.jp/dl/iste2012\\_web\\_proceedingsv2.pdf](http://www.isate2012.jp/dl/iste2012_web_proceedingsv2.pdf)
- Apple, M., Falout, J., & Hill, G. (in press). Exploring classroom-based constructs of EFL motivation for science and engineering students. In M. Apple, D. Da Silva, & T. Feller (Eds.), *Foreign language motivation in Japan*. Bristol, UK: Multilingual Matters.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Erlbaum.
- Brown, J. D., Robson, G., & Rosenkjar, P. R. (2001). Personality, motivation, anxiety, strategies, and language proficiency of Japanese students. In Z. Dörnyei & R. Schmidt (Eds.), *Motivation and second language acquisition* (pp. 361-391). Honolulu: University of Hawai'i.
- Cheng, T.-S., Horwitz, E. K., & Schallert, D. L. (1999). Language anxiety: Differentiating writing and speaking components. *Language Learning, 49*, 417-446.
- Cohen, Y., & Norst, M. J. (1989). Fear, dependence, and loss of self-esteem: Affective barriers in second language learning among adults. *RELC Journal, 20*, 61-77.

- Elkhafaifi, H. (2005). Listening comprehension and anxiety in the Arabic language classroom. *The Modern Language Journal*, 89, 206-220.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Eysenck, H. J. (1970). *The structure of human personality* (3rd ed.). London: Methuen.
- Fisher, W. (1992). Reliability statistics. *Rasch Measurement Transactions*, 6, 238.
- Gardner, R. C., & MacIntyre, P. D. (1993). A student's contribution to second language learning. Part II: Affective variables. *Language Teaching*, 26, 1-11.
- Gregersen, T. (2007). Breaking the code of silence: A study of teachers' nonverbal decoding accuracy of foreign language anxiety. *Language Teaching Research*, 11, 209-221.
- Gregersen, T., & Horwitz, E. K. (2002). Language learning and perfectionism: Anxious and non-anxious language learners' responses to their own oral performances. *The Modern Language Journal*, 86, 562-570.
- Horwitz, E. K., Horwitz, M. B., & Cope, J. A. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70, 125-132.
- Kimura, H. (2008). Foreign language listening anxiety: Its dimensionality and group differences. *JALT Journal*, 30, 173-195.
- Linacre, J. M. (2002). What do infit, outfit, mean-square, and standardized mean? *Rasch Measurement Transactions*, 16, 878.
- Linacre, J. M. (2006). *Winsteps* (Version 3.63) [Computer software]. Chicago: Winsteps.com
- Linacre, J. M. (2007). *A user's guide to WINSTEPS: Rasch-model computer program*. Chicago: MESA.
- MacIntyre, P. D. (1995a). How does anxiety affect second language learning? A reply to Sparks and Ganschow. *The Modern Language Journal*, 79, 90-99.
- MacIntyre, P. D. (1995b). On seeing the forest through the trees: A rejoinder to Sparks and Ganschow. *The Modern Language Journal*, 79, 245-248.
- MacIntyre, P. D. (1999). Language anxiety: A review of the research for language teachers. In D. J. Young (Ed.), *Affect in foreign language and second language learning: A practical guide to creating a low-anxiety classroom atmosphere* (pp. 24-45). Boston, MA: McGraw-Hill.
- MacIntyre, P. D. (2007). Willingness to communicate in a second language: Understanding the decision to speak as a volitional process. *The Modern Language Journal*, 91, 564-576.
- MacIntyre, P. D., Babin, P. A., & Clément, R. (1999). Willingness to communicate: Antecedents & consequences. *Communication Quarterly*, 47, 215-229.

- MacIntyre, P. D., & Charos, C. (1996). Personality, attitudes, and affect as predictors of second language communication. *Journal of Language and Social Psychology, 15*, 3-26.
- MacIntyre, P. D., & Gardner, R. C. (1991). Language anxiety: Its relation to other anxieties and to processing in native and second languages. *Language Learning, 41*, 513-534.
- MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning, 44*, 283-305.
- MacIntyre, P. D., Noels, K. A., & Clément, R. (1997). Biases in self-ratings of second language proficiency: The role of language anxiety. *Language Learning, 47*, 265-287.
- Mak, B. (2011). An exploration of speaking-in-class anxiety with Chinese ESL learners. *System, 39*, 202-214.
- Matsuda, S., & Gobel, P. (2004). Anxiety and predictors of performance in the foreign language classroom. *System, 32*, 21-36.
- McCroskey, J. C. (1977). Oral communication apprehension: A summary of recent theory and research. *Human Communication Research, 4*, 78-96.
- McCroskey, J. C. (1978). Validity of the PRCA as an index of oral communication apprehension. *Communication Monographs, 45*, 192-203.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Price, M. L. (1991). The subjective experience of foreign language anxiety: Interviews with highly anxious students. In E. K. Horwitz & D. J. Young (Eds.), *Language anxiety: From theory and research to classroom implications* (pp. 101-108). Englewood Cliffs, NJ: Prentice-Hall.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago.
- Saito, Y., Garza, T. J., & Horwitz, E. K. (1999). Foreign language reading anxiety. *The Modern Language Journal, 83*, 202-218.
- Scovel, T. (2001). *Learning new languages: A guide to second language acquisition*. Boston, MA: Heinle and Heinle.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*, 107-120.
- Smith, E. V. (2001). Evidence for the reliability of measures and validity of measurement interpretation: A Rasch measurement perspective. *Journal of Applied Measurement, 2*, 281-311.
- Smith, E. V. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement, 3*, 205-231.

- Smith, R. M. (2000). Fit analysis in latent trait measurement models. *Journal of Applied Measurement, 1*, 199-218.
- Sparks, R. L., & Ganschow, L. (1991). Foreign language learning differences: Affective or native language aptitude. *The Modern Language Journal, 75*, 2-16.
- Sparks, R. L., & Ganschow, L. (1995). A strong inference approach to causal factors in foreign language learning: A response to MacIntyre. *The Modern Language Journal, 79*, 235-244.
- Sparks, R. L., Ganschow, L., & Patton, J. (1995). Prediction of performance in first-year foreign language courses: Connections between native and foreign language learning. *Journal of Educational Psychology, 87*, 638-655.
- Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y)*. Palo Alto, CA: Consulting Psychologists.
- Taylor, J. A. (1953). A personality scale of manifest anxiety. *The Journal of Abnormal and Social Psychology, 48*, 285-290.
- Thurstone, L. L. (1931). Measurement of social attitudes. *The Journal of Abnormal and Social Psychology, 26*, 249-269.
- Waugh, R. F., & Chapman, E. S. (2005). An analysis of dimensionality using factor analysis (true-score theory) and Rasch measurement: What is the difference? Which method is better? *Journal of Applied Measurement, 6*, 80-99.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Erlbaum.
- Wolfe, E. W., & Smith, Jr., E. V. (2007a). Instrument development tools and activities for measure validation using Rasch models: Part I – Instrument development tools. *Journal of Applied Measurement, 8*, 97-123.
- Wolfe, E. W., & Smith, Jr., E. V. (2007b). Instrument development tools and activities for measure validation using Rasch models: Part II – Validation activities. *Journal of Applied Measurement, 8*, 204-234.
- Wright, B. D. (1996a). Comparing Rasch measurement and factor analysis. *Structural Equation Modeling, 3*, 3-24.
- Wright, B. D. (1996b). Reliability and separation. *Rasch Measurement Transactions, 9*, 472.
- Wright, B. D. (1999). Fundamental measurement for psychology. In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement: What every psychologist and educator should know* (pp. 65-104). Mahwah, NJ: Erlbaum.
- Wright, B. D., & Masters, G. N. (2002). Number of person or item strata. *Rasch Measurement Transactions, 16*, 888.

Young, D. J. (1999). *Affect in foreign language and second language learning: A practical guide to creating a low-anxiety classroom atmosphere*. Boston, MA: McGraw-Hill.

## Appendix

### ***The Foreign Language Classroom Speaking Anxiety Scale (FLCSAS)***

- I'm worried that other students in class speak better than I do.\*
- I feel nervous speaking in front of the entire class.\*
- I tremble when the teacher is about to ask me a question.
- I am reluctant to express my opinion in a group.
- I'm worried about making mistakes when I speak with the teacher.
- I'm worried that my partner speaks better English than I do.\*
- I am reluctant to ask the teacher a question.
- I start to panic when I speak with a classmate in a pair.\*
- I dislike speaking entirely.
- I'm worried that the teacher will think my speaking is no good.
- I'm worried about making mistakes while speaking.\* †
- I feel nervous when I can't express my opinion.\* †
- I'm afraid my partner will laugh when I speak with a classmate in a pair.\*
- I'm worried about making mistakes when I speak with a partner.\* †
- Answering a teacher's question in class is embarrassing.
- Speaking in a group of classmates makes me feel self-conscious.
- I feel tense when I have to speak with a classmate in a pair.\*
- I start to panic when I have to speak in a group.\*
- I'm afraid that others in a group discussion will laugh if I speak.\* †
- I can feel my heart pounding when it's my turn to speak in a group.\* †

*Notes.* Items marked with an asterisk (\*) represent items included in the final long version of the construct (Apple, 2011). Items marked with the symbol † represent items included in the short version (Apple, Falout, & Hill, 2012, in press).

# Teacher Cognition and Grammar Teaching in a Japanese High School

Masateru Nishimuro

*Kwansei Gakuin Senior High School, Nishinomiya*

Simon Borg

*School of Education, University of Leeds, UK*

In this paper we examine the relationship between teachers' practices and cognitions in teaching grammar. A qualitative analysis of classroom observations and interviews with three experienced Japanese teachers of English highlights both the key features of these teachers' pedagogies in teaching grammar and the cognitions underpinning their work. Their practices were characterized by detailed teacher-fronted explanations of grammatical forms, metalinguistic explicitness, and the use of the L1. These practices were underpinned by the teachers' beliefs about the value of grammar, though there was little evidence that the teachers' pedagogical choices were consciously informed by current theoretical or methodological ideas in the field of L2 teaching. Rather, the approach to grammar adopted by the teachers had a strong experiential basis and was influenced by a desire to motivate and promote the well-being of their students as well as by contextual factors such as learners and colleagues.

近年、教師認知の観点からCommunicative Language Teaching (CLT) の導入に関わる考察が多くなされているが、本稿は日本の英語教育の主な関心であり続ける文法指導にその焦点をあてた。日本の高校に勤める経験豊かな3人の日本人英語教師各々に授業観察と観察前後のインタビューからなる質的調査を実施し、文法指導の実践とその実践の根拠となる彼らの教育的信条や認知との関係性を探った。日本語を用いた教師主導の文法形式の説明とメタ言語学的明示性に象徴された彼らの実践は、文法指導に価値をおく彼らの信条が根拠となっていたが、その指導法を採用する判断においては第二言語(外国語)教育分野の理論、方法論を基にしたという根拠はほぼなかった。むしろ彼らの指導法は、教師自身の学習・指導経験が大きな基礎を

なし、学習者の動機を高め彼らに人間として良く成長してほしいという願い、学習者や同僚といったコンテキスト要因に影響されていた。

**T**eacher cognition—defined as “what teachers know, believe, and think” (Borg, 2003, p. 81)—has in the last 15 years become a major area of research in the field of L2 teaching. This work has provided insights into the mental lives of teachers as well as into how cognition and context interact in shaping the instructional decisions teachers make. As a result of this research, it is widely acknowledged that it is not possible to understand what teachers do without also understanding their cognitions. Various domains of L2 teaching have been examined from a teacher cognition perspective (Borg, 2006); amongst these, grammar teaching, the focus of this paper, has received the most attention.

### **Teacher Cognition and Grammar Teaching**

Research into grammar learning has a long history (see the comprehensive review in Ellis, 2008) and while it is now acknowledged that some degree of formal instruction of grammar supports the L2 learning process (Nassaji & Fotos, 2011), precise and universally applicable answers to many of the questions teachers face in teaching grammar remain unavailable, as Ellis (2006) explains. Examples of such questions are whether grammar should be approached deductively or inductively (Gollin, 1998), whether grammar is best taught integratively or as a separate component of the foreign language learning syllabus (Borg & Burns, 2008), and whether knowledge of grammatical terminology enhances learning. Teachers also face questions about the nature of the grammar practice activities they should employ (Nitta & Gardner, 2005) and about the medium (i.e., L1 or L2) through which explicit grammar work is best handled. To complement research on the learning of L2 grammar and in response to the lack of definitive guidance it has provided to the above issues, a teacher cognition perspective has more recently been applied to the study of L2 grammar teaching.

Several studies in a range of international contexts illustrate this perspective. For example, drawing on a number of studies in Hong Kong, Andrews (2007) found that teachers' own knowledge about grammar plays a significant role in the instructional decisions they make when teaching it, while Borg (2001), working with a teacher in Malta, showed that it was not only teachers' actual knowledge of grammar that influenced their teaching, but also how confident they felt about this knowledge (less confident teachers



were found to minimize explicit grammar work). In relation to whether grammar is best taught separately or integrated with other language skills, Borg and Burns (2008) reported that 84% of an international sample of 176 teachers disagreed that “grammar should be taught separately” from other skills (p. 466). This body of research has also shown that how teachers approach grammar will be shaped by the interaction of their beliefs about language learning, their views about their learners’ needs and wants, and contextual factors such as the availability of time (e.g., Farrell & Lim, 2005). As a result of such interactions, what teachers do in the classroom may not necessarily reflect their beliefs about how grammar should be taught (see, e.g., Basturkmen, Loewen, & Ellis, 2004). It is therefore essential in studies of teacher cognition to examine not only teachers’ theoretical beliefs (e.g., Jean & Simard, 2011; Schulz, 1996) but also how these are mediated in practice by the contexts in which teachers work (e.g., Pehissa & Tragant, 2009, which shows how context impacts on the teaching of L2 grammar).

There has been little specific qualitative research into teacher cognition and grammar teaching in Japan. There have been, however, a number of studies into teachers’ implementation of communicative principles (e.g., Gorsuch, 2000, 2001; Kikuchi & Browne, 2009; Nishino, 2008, 2011; Sakui, 2004, 2007; Sato & Kleinsasser, 2004; Taguchi, 2005). These are relevant to the current study in that they provide evidence that teachers of English in Japan continue to value explicit grammar work despite policy and teacher training initiatives aimed at increasing the frequency of communicative activities in the language classroom. For example, Gorsuch, through a survey of 876 Japanese teachers of English, concluded that while they mildly approved of communicative activities, there were also “potent impediments working against teachers actually using such activities in their classrooms” (Gorsuch, 2001, Conclusion section, para. 1). A recent analysis by Underwood (2012) highlighted a similar phenomenon, although he also questioned the extent to which teachers’ perceptions of some of these impediments were accurate. Via a questionnaire and focus group interviews with 16 senior high school teachers of English in Japan, this study suggested that the teachers held positive beliefs about the integration of grammar with communication-oriented teaching. These beliefs, though, were tempered by concerns (and misconceptions, the author suggested) about the negative impact that a communicative approach to grammar would have on students’ performance on university entrance examinations. Many teachers in this study thus seemed to believe that such examinations emphasized discrete-point grammatical knowledge much more than an analysis of the examinations themselves would suggest.

Given such insights and the continuing prominence of grammar teaching in debates about English language teaching in Japan, further qualitative research in this context is needed to develop deeper understandings of both what teachers do and the reasons behind their pedagogical decisions in relation to grammar. The study we report here responds to this need.

Note: For the sake of conciseness, and corresponding to such factors as *practices* and *behaviours*, any identifiable teacher knowledge or awareness is herein referred to as a *cognition* (plural: *cognitions*).

## The Study

### Research Questions

Informed by the literature discussed above, this study examined the following questions:

1. What practices characterize the grammar teaching of three teachers of English in a Japanese high school?
2. What cognitions underpinned these practices?
3. What contextual factors shaped the teachers' work?

Grammar teaching here was defined as instruction which focuses on the morpho-syntactic features of English.

### Context

This study took place in a private boys' high school in a large Japanese urban area. Approximately 90% of the school's graduates enter the university with which it is affiliated by way of self-recommendation (based on their academic and extra-curricular achievement reports for the previous 3 years) rather than via a more customary high-stakes entrance examination. This is significant here because university entrance examinations—a source of pressure the teachers in this study were free from—have been cited in studies of Japanese teachers of English as a major constraint on their practice (e.g., Gorsuch, 2001). There were approximately 900 students in the school together with 16 Japanese English teachers (seven of whom worked part-time) and three native-speaking English Assistant Language Teachers. The school followed the national curriculum and enjoyed a good academic reputation in the region. The textbooks used for teaching English provided both grammar and skills-oriented work and were locally written, government approved, and chosen by the English department in the school. They were in Japanese and contained some pair work and group work activities after every five lessons.

## Participants

The participants were three experienced male Japanese EFL teachers. Two criteria determined their choice: first, when contacted prior to the study, they confirmed that grammar teaching was a central facet of their work; and second, they were interested in contributing to the study. Table 1 describes the participants (using pseudonyms) and provides information about the lessons they were observed teaching in this study. As this shows, in Mr. Tanaka's and Mr. Endo's classes, teaching grammar was the main focus of the lessons, while grammar was a secondary focus in Mr. Matsuda's reading lessons. All three teachers used officially approved coursebooks but were not obliged to follow any particular teaching methodology.

**Table 1. Participants in the Study**

Pseudonym	Age	Years of teaching experience	Lessons observed
Matsuda	50	25	Reading for Year 12 (45 students)
Tanaka	50	27	English II*: Level-based English writing for Year 11 <To infinitive> (20 students)
Endo	47	25	English I*: English grammar for Year 10, <will/be going to, perfect tense> (45 students)

\*In this school, English I is taught to Year 10 students and is divided into two subjects: Reading and English Grammar. English II is taught to Year 11 students and is divided into Reading and level-based English writing.

## Data Collection

The study employed a qualitative methodology, which was appropriate given our goal of understanding practices and cognitions in some depth and from teachers' own perspectives. Our choice acknowledged the position that "studies which employ qualitative strategies to explore language teachers' actual practices and beliefs will be more productive (than, for example, questionnaires about what teachers do and believe) in advancing our understanding of the complex relationships between these phenomena" (Phipps & Borg, 2009, p. 388). A case study approach (Duff, 2008) was employed, as

this allowed us to examine each teacher's practices in detail and individually before moving on to consider the findings collectively. For each case, there were three phases of data collection. First, each teacher participated in a pre-observation interview. They were then observed teaching two regular lessons of 45 minutes before finally participating in a post-observation interview. All data were collected in the 2010-11 academic year.

### *Pre-Observation Interviews*

The pre-observation interview aimed to elicit background information about each participant that was felt to be relevant in understanding their approach to grammar teaching. The issues covered were the teachers' educational background, experience of learning and teaching English or grammar, self-perceptions of their English proficiency and knowledge about grammar, school culture, and their general views about ELT and grammar teaching. Some of the questions were adapted from those in existing research on teacher cognition and grammar teaching, namely Borg (1998), Borg and Burns (2008), and Schulz (2001). The interviews were conducted in Japanese, audio recorded, and lasted on average 90 minutes. The semi-structured format (for a discussion see, for example, Dörnyei, 2007) meant that while questions were set in advance, there was flexibility for additional questions to be asked in response to issues emerging in the course of the conversation. Appendix A gives examples of the pre-observation interview questions.

### *Classroom Observations*

As noted above, each teacher was observed teaching two regular lessons of 45 minutes each. As Borg (2006) suggests, classroom observation "has a central role to play in the study of language teacher cognition by providing a concrete descriptive basis in relation to what teachers know" (p. 231). Thus, the observations provided direct evidence of teachers' practices in teaching grammar. The observations took place in a normal classroom setting, and the observer was a nonparticipant in the lesson (i.e., observing as a researcher and not interacting with the teacher or students). A video camera at the back of the classroom recorded each lesson and field notes were also made. All lessons were conducted in Japanese and the episodes presented here are English translations.

### *Post-Observation Stimulated Recall Interviews*

Data collected through the observations were analysed in order to generate questions for a second interview with each teacher. This was a stimulated recall interview using transcribed extracts of the observed lessons and materials (i.e., coursebooks and handouts) to prompt a discussion of the teachers' approach to grammar and of factors shaping it (for recent examples of the use of stimulated recall interviews in the study of language teacher cognition, see Kanno & Stuart, 2011; Kim, 2011; Kuzborska, 2011). Participants were asked to comment on aspects of their teaching and, in line with the purposes of qualitative stimulated recall interviews, they were given "the opportunity to verbalize their thinking, in a relatively free and open-ended manner" (Borg, 2006, p. 210). These interviews were conducted (again, in Japanese) as soon as possible after the observed lesson. They were recorded and lasted on average 90 minutes. Appendix B lists examples of the questions asked during the post-observation interviews.

Overall, the database for this study consisted of approximately 9 hours of recorded interview data and 4.5 hours of recorded classroom observation data, together with the materials used in the lessons. All phases of the study were informed by an awareness of ethical principles in educational research, such as informed consent and confidentiality (see, for example, Denscombe, 2002).

### ***Data Analysis***

All recorded data were transcribed, then translated into English by the first author. The pre-observation interviews were analysed qualitatively and profiles of each teacher developed using the background information provided in those interviews. The observational data were then analysed for each teacher to identify key instructional episodes (Borg, 1998) in their work. A key instructional episode was defined as one that illustrated an aspect of the teacher's work in teaching grammar. In identifying such episodes, key themes in the literature on grammar teaching, highlighted earlier, provided a set of categories for describing practice. For example, observations were analysed for evidence of the ways in which teachers introduced and practised grammar and of the extent to which the L1 and grammatical terminology were used. As a result of this analysis, the key characteristics of each teacher's work in teaching grammar were identified. The post-observation interviews were analysed according to established principles for working with qualitative data (see, e.g., Newby, 2010). Through repeated readings

of the data, key themes in the teachers' rationales for their practices were identified and categorized, following a largely inductive process (i.e., where categories emerged from the data) although factors highlighted in previous research on teacher cognition and grammar teaching also provided some direction to the analysis. Insights from the pre-observation interviews were also drawn on here in constructing, for each teacher, an account of how he taught grammar and why he did so in particular ways. These individual accounts were then compared and key crosscutting themes in them identified.

## **Findings**

In the curriculum followed by the teachers, grammar was addressed through 2 hours of designated grammar lessons each week. The teachers did not express any opposition to this approach and shared the belief that grammar instruction plays a vital role in an EFL environment such as Japan. While they acknowledged the need for their students to learn how to use English communicatively, they considered grammar to be the basis of communication and thus felt the need to make it a primary and explicit focus of their work. Endo, for example, explained in his first interview that he regularly told his students that grammar was absolutely necessary for both accurate and fluent use of English. As we illustrate below, the teachers' observed practices reflected these beliefs about the centrality of explicit grammar teaching.

### *Grammar Teaching Activities*

In terms of the activities the teachers utilized in teaching grammar, there was a pronounced preference for the presentation of grammar rather than for activities where the grammar was practised. Collectively, some 80% of the lessons observed were spent on teacher-led grammar presentations (in one case, Endo spent 35 minutes of the 40 available explaining grammar, while in another Tanaka's explanation took up 30 minutes of a lesson). Tanaka explained the prevalence of explanations in this manner:

It [the explanation] may be too long. However, in fact, I need to cover what other levels are doing for mid/end-of-term exams. . . . To cover those, I cannot exceed 3 hours for one unit. For level 1 students, they cannot do exercises without explanation . . . . I cannot waste my time and should not let students sleep . . . and I need to keep pace with other classes. That's why I give

a lot of explanations but cannot do enough exercises. (Tanaka, Interview 2, 5 Dec 2010)

Tanaka's concerns about keeping up with his colleagues highlight a contextual factor in his approach to teaching grammar. This was also noted by Matsuda when he explained that "I select grammar to explain based on what's in the coursebooks, so as to keep pace with other teacher's classes" (Matsuda, Interview 2, 5 Dec 2010), which was important to do because all students in the same year did the same mid-term and final examinations. Endo also acknowledged that he gave long lectures on grammar; but, echoing Tanaka's comments, he said that it was inevitable due to the limited time available to cover the syllabus. He was hopeful, though, that during his explanations the students were engaged in thinking about grammar:

I do give lectures, but I try to let them think and use their brains while listening . . . It is like a Socratic method or playing catch between teachers and students. So the role of the teacher is to throw *why* questions to students to let them discover the rules. (Endo, Interview 2, 5 Dec 2010)

The teachers also cited a lack of time as the reason why they did not employ communicative activities (e.g., involving pair work) in their lessons; for example, Endo said, "No, it's impossible to use such activities. I know they are absolutely necessary but physically impossible because time is limited" (Endo, Interview 2, 5 Dec 2010).

Matsuda gave students a little more time to work on exercises than the other two teachers. However, during the activities, he translated English into Japanese sentence-by-sentence, and checked the meaning of words and key grammar items. In justifying this approach, his views were that

we need to check the meanings of each sentence, to help those students who cannot get good grades, to help them keep up with the lessons. Otherwise, they will lose a sense of accomplishment . . . We need to care for such students. (Matsuda, Interview 2, 5 Dec 2010)

A concern for helping the less able students was a strong theme in his account of why he focused explicitly on explaining grammar:

One of the features of a Japanese coursebook is that it has a lot of grammar items. When I think of the faces of the students

who cannot do well in the exam, I feel I need to explain things to them. When I think about examinations, I cannot leave them stranded . . . I must help them review so that they can recall what they learned before. (Matsuda, Interview 2, 5 Dec 2010)

### *Analysing Grammar*

The detailed analysis of grammar was a key feature of the work of Endo and Tanaka. Because Matsuda was observed teaching grammar in the context of a reading lesson, he spent comparatively less time than the other two teachers talking about grammar in class but still explained grammar in detail after the reading work had been completed. Episode 1 illustrates how Tanaka approached the analysis of grammar:

#### Episode 1: Presenting *to-infinitive*

T: In the last lesson, I explained how to use *to-infinitive* in noun or adjective phrases. But today, what we will learn is for example . . .

And then he writes *He went to England* on the blackboard.

T: This *to* shows what?

S: Direction.

T: Right. Mr. Nishimuro went to England.

Then he adds *to study* next to the sentence.

T: What we will learn today is this *to*. *To* plus a verb. What comes after *to* is a verb. What does this *to* mean?

S: In order to.

T: Right. This *to* refers to a goal. Because this *to study* modifies the whole sentence, we call this the adverbial usage of *to-infinitive*. (Tanaka, Observation 1, 5 Jul 2010)

In this extract, grammar is being presented through decontextualized sentences (i.e., individual sentences not part of a longer stretch of discourse). This was typical of the work of Tanaka and Endo and both teachers shared the belief that analysing grammar in this manner allowed students to learn it more effectively. For example, Endo explained that



to obtain the knowledge of the grammar, students will get confused if there is a context . . . . In the beginning, students should only focus on the form and know why this is past tense or present tense. And once they understand, it's better to read the text and check how they are used. (Endo, Interview 2, 5 Dec 2010)

Such comments imply a belief that the form of grammar needs to be mastered before its use can be learned, a position which is at odds with the view that encountering grammar in context is beneficial to learning as it allows students to establish clear links between form and meaning.

Episode 1 also illustrates how Tanaka led the analysis of grammar by asking students a series of closed questions—this analysis was conducted orally and via teacher-class interaction. Endo's approach to grammar analysis reflected a similar approach, as Episode 2 shows:

#### Episode 2: Presenting present perfect

E: I'm sure you have learned present perfect in junior high school. What was the form?

Ss: *Have* plus a past participle.

He writes *have (has) + p.p.* on the blackboard.

E: And in junior high, you learned that the present perfect is used to mean things being done, result, experience, and continuity. But, it's not important to remember those things. But knowing the world of present perfect is important. Before we do that, what do you think about these two sentences, *I finished lunch* and *I have finished lunch*? What is the difference? Could you translate these?

Ss: One student answers—*Watashi wa hirugohan wo tabeta.*

E: Yes, in Japanese, we say them in the same way, but in English, there is a difference. An important thing is to grasp the image that native speakers have. The present progressive is like a movie and presents a sense of energy, the past tense is like a photograph. Then what's present perfect?

(He then explains that the present perfect tense indicates an image of *approaching* to the present, and distributes a handout and reads it with students.) (Endo, Observation 2, 5 Dec 2010)

Endo liked to compare structures and to make students aware of the way that native speakers of English interpreted the target structures (often, as in this case, using metaphors as part of his explanation). On the importance of comparing structures, Endo explained that

for example, in junior high schools, they don't teach the difference between *will* and *be going to*. In high schools, I expect students not only to know the rules as rules, but also I want them to know how native speakers grasp the images or the world of grammar. . . . Because students do have basic knowledge of grammar, I want students to discover the rules using what they have got like in mathematics. (Endo, Interview 2, 5 Dec 2012)

Endo's comments echo those made earlier by Tanaka about encouraging students to discover rules, though as we see from the extract above Endo's approach to presenting grammar involved discovery that was highly teacher led (as opposed to forms of discovery grammar work where learners are deeply engaged in analyzing and discussing language and reaching their own conclusions—see, for example, Tomlinson, 2007).

### *L1 and Terminology*

In the lessons observed, grammar teaching was strongly characterized by the use of the L1 and (particularly in the work of Tanaka and Endo) of grammatical terminology (the teachers used terminology in Japanese, not in English—for example, *to-futeishi*, for *to-infinitive*). Endo's view on the widespread use of the L1 in teaching grammar was that "It was too unnatural to use classroom English. It confuses students who are low in proficiency . . . . The most important goal in this lesson is to understand the logic of English grammar" (Endo, Interview 2, 5 Dec 2010).

Tanaka also justified his use of the L1 with reference to students' level and motivation:

It may depend on the students' level but for level 1 students, it's better to use Japanese, because they can understand it the most. Moreover, these lower level students hate English and do not study hard. Students tend to sleep in class or don't even open the coursebook . . . . To attract them to English, I must go off on tangents and tell interesting stories relating to the target grammatical points, but to do so, they won't understand if I use English. (Tanaka, Interview 2, 5 Dec 2010)

Matsuda felt that using English in his reading classes was not relevant to the goal of enabling students to do well in reading comprehension tests: “The target skill of the lesson is to read English . . . students feel no necessity to use English for instruction because it won’t be on the exams. Their goal at this stage is to get good grades” (Matsuda, Interview 2, 5 Dec 2010).

Grammatical terminology was another salient feature of the classroom talk during Endo and Tanaka’s lessons (see Episodes 1 and 2 above). Matsuda’s use of terminology was less frequent but was nonetheless evident, as the next episode shows:

### Episode 3: A participle construction

(After reading the text, checking the meanings of the passage and listening to the CD, Matsuda introduced the target grammar.)

M: Look at line 10, *Just lying around all day, it opens its cute eyes if it is patted, and begins to wave its flippers*. What is this construction [i.e., *lying around all day*]?

Ss: A participle construction.

M: Yes, the subject is omitted, and this *it* refers to what?

Ss: A robotic seal.

M: That’s right. So the subject of this clause is also *a robotic seal*. (Matsuda, Observation 2, 5 Nov 2010, p. 3)

The three teachers held varying perspectives on the value of terminology. Tanaka and Endo both agreed it was important for learners, but for different reasons. Tanaka’s view was that “for level 1 students, I make them pay attention to the structure and usage of grammatical items because they don’t even understand what adjectives are . . . They should be taught thoroughly” (Tanaka, Interview 2, 5 Dec 2010). He believed that a knowledge of terminology was essential for lower level learners to understand grammar. Endo also valued the use of terminology, but he explained, “I use them [grammatical terms] to give students a sense of security. There are students who like terminology to remember grammatical items. Students can check if their knowledge of grammar is correct so that they can feel secure” (Endo, Interview 2, 5 Dec 2010). Endo was not suggesting that knowing terminology improves grammar learning; rather, he believed that students derive a kind of psychological reassurance from learning the terms. Matsuda’s views

about terminology were less positive than the other two teachers: “I don’t think they need to understand terminology. Some may be useful for students who are good at English, or take entrance exams outside the school or who are interested in English. I know that most students struggle with those words” (Matsuda, Interview 2, 5 Dec 2010). Nonetheless, he did not avoid the use of terminology in his classes.

## **Discussion**

Our first research question focused on the nature of teachers’ practices in teaching grammar. Contemporary views of L2 learning (see, for example, Storch, 2010) recognize a role for explicit grammar work in the context of a broader communicative framework. Relative to this position, the grammar teaching practices highlighted in this study—characterized by discrete grammar lessons, extended decontextualized and metalinguistically rich teacher-led analysis conducted largely in the L1, with an emphasis on logic (e.g., equating grammar with mathematics) and limited opportunities for meaningful grammar use—would be considered somewhat traditional. It is not our goal here, though, to criticize the pedagogies of the three experienced teachers who participated in this study; rather, our aim is to understand factors that interactively shape the approach to grammar that the teachers adopted.

Our second research question was to ascertain which cognitions underpinned the teachers’ practices. All three teachers shared the view that a knowledge of grammar was central to L2 learning and that grammar thus needed to be an important and explicit facet of the teaching of English. Such a view is not necessarily inconsistent with contemporary views of grammar in SLA research, though there was no evidence here that any of the teachers had been influenced by an awareness of such views; in fact, one noteworthy facet of the accounts they provided for their approaches to grammar was the absence of any reference to formal theory or methodological concepts in L2 teaching (which was also a feature of the teacher accounts of grammar teaching reported in Borg & Burns, 2008, and Eisenstein-Ebsworth & Schweers, 1997). The bases for the teachers’ classroom practices seemed to be in fact largely experiential. One contributing factor here was their own histories as learners of English and it is highly likely that the way they were taught was not fundamentally different to their own practices today (see, for example, Nishino, 2011, who found that Japanese teachers said they had had few chances to experience communicative activities in English lessons when they were high school students). The experiential basis of how the

teachers taught grammar was also informed by their professional experiences over many years in a system that has traditionally promoted formal grammar learning and by their accumulated knowledge of the students and the institutional context in which they work. One general conclusion from this study is that teachers' pedagogies had a very strong experiential basis, with very limited evidence that theoretical or methodological principles played a role in determining what the teachers did. We are not suggesting that the teachers lacked an awareness of such principles, but it is clear that in explaining the rationales for their pedagogical choices the teachers did not seek to justify their work in terms of theories of language learning and teaching.

In discussing the reasons for their approach to grammar, the teachers did articulate a number of beliefs. Some were specifically about language learning, such as that grammar is central to L2 learning, that grammar is best learned from first and out of context, and that a knowledge of terminology can support grammatical understanding. Many of the teachers' beliefs, though, related to learning more broadly. Thus, the teachers felt that learner motivation may be harmed if they have to struggle to understand the teacher (and for this reason the L1 was widely used). They also felt a duty to give lower level learners additional support so that they may experience a sense of achievement (and this is why lengthy grammatical explanations were provided). Effective rapport with students was also valued (and this was another reason for using the L1). A belief in learning through *discovery* was also cited to justify teacher-led whole class elicitation of grammar rules (as we noted earlier, though, this notion of discovery does not reflect what is typically implied in discussions of inductive grammar work—see, for example, Gollin, 1998). Overall, the teachers' general beliefs about students' well-being, motivation, and sense of achievement seemed to be at least as powerful an influence on how grammar was taught as specific beliefs about L2 grammar learning itself. Writing about the teaching of English in Japan, Aspinall (2006) also noted that the teachers were often concerned about the general well-being of their students rather than focusing solely on their progress as language learners. These findings remind us that in seeking to understand how teachers teach grammar, it is important to explore not just the views they hold about L2 learning but also their broader underlying, and sometimes more powerful (see Phipps & Borg, 2009), beliefs about learning in general.

Studies of teachers' practices and cognitions must also take into account the effect that context has on mediating the relationship between what

teachers believe and what they do, and this was the focus of our third research question. There is evidence in the literature that teachers' beliefs about L2 learning may sometimes be at odds with how they teach (for a recent review of this literature see Basturkmen, 2012) and the explanation often given for such mismatches is the constraining influence of the context, especially assessment (see, for example, Pahissa & Tragant, 2009, who illustrate the impact of a high-stakes examination on grammar teaching in Spain).

In this study, the teachers considered test preparation important, but it did not emerge as a major constraint on what they did. The teachers felt, though, that the learners' low proficiency and motivation were powerful influences on how they approached grammar. Two other contextual factors that were influential were a lack of time and collective targets. Thus, the teachers felt they had limited time to cover the syllabus and also felt under pressure to cover the textbook at the pace needed to meet the targets agreed by all colleagues teaching any particular level. This combination of time and collective targets was cited by the teachers as one reason why so little classroom time was allocated to grammar practice or to any form of pair work or group work, even though the teachers acknowledged the value of communicative work. Rather, they prioritized the presentation and analysis of grammar as this was seen as the most direct and time-efficient way of promoting understanding. This finding reflects that of Sato and Kleinsasser (2004), in which teachers avoided using communicative activities even in oral communication lessons in order to "keep pace" with other classes. Our analysis of contextual factors here was limited to those within the teachers' school; the larger historical and socio-cultural context for education in Japan (e.g., decisions made at Ministry level) will also inevitably have had a bearing on what teachers do in the classroom.

Overall, then, our results add to existing evidence of the complex and multi-faceted nature of L2 grammar teaching. What emerges from this study is a portrait of how three high school teachers of English in Japan approached grammar, together with insights into the beliefs (about grammar, L2 learning, and education more broadly) and the contextual factors which, interactively, shaped the teachers' pedagogical choices.

## **Conclusion**

In concluding this study, we acknowledge that the insights presented here have arisen from the analysis of a total of six lessons and six interviews involving three very experienced male Japanese teachers of English working in

the same high school. We are not seeking to make more general claims about the teaching of grammar in Japanese high schools (although research noted earlier about the implementation of communicative language teaching in Japan does point to pedagogical practices which reflect those described here). Similarly, we are not arguing that the cognitions articulated by these teachers are in any way typical of the wider population of teachers of English in Japan. The value of this research lies in the qualitative insights it provides into what teachers actually do in their classrooms and into the manner in which teachers rationalize their practices. As noted earlier, in the last 15 years there has been increasing interest in exploratory case studies of this kind in the context of L2 grammar teaching, yet (in contrast with larger scale survey studies of teachers' attitudes and practices) specific analyses of teachers' practices and cognitions in the Japanese context remain limited. Much further qualitative research of this kind is needed as it can contribute to an understanding of the challenges and dilemmas teachers face in teaching grammar and, in particular, inform a critical analysis of the forces which may counter policies which seek to promote communicative language teaching more widely in Japanese high schools.

Three interlinked issues to emerge here appear to merit, we believe, further study in ways that extend beyond the specific research questions we have addressed. The first stems from the contrast between contemporary views of good practice in L2 grammar teaching and the pedagogies the teachers adopted here. Where do the beliefs underlying such pedagogies (such as the view that grammatical forms are best initially encountered out of context) come from? To what extent do teachers believe these pedagogies are effective and how do they make such judgements? Why are such instructional practices resilient despite efforts to make the teaching of English in Japanese high schools less grammar focused? The second issue (which may underlie the first) is the limited role that knowledge about L2 teaching played in teachers' rationales for their work. We recognize that experience will always be a major influence on what teachers do but argue that experience alone, particularly that which is not subject to critical reflection informed by propositional knowledge, may not necessarily promote productive pedagogy. Thus, what propositional knowledge about L2 grammar teaching do teachers hold and to what extent does this inform the pedagogical decisions teachers make? Finally, there was evidence here that teachers' general educational beliefs about keeping students interested and motivated were at least as powerful as their subject-specific beliefs about how to teach grammar. Thus, how do these different belief systems

interact in influencing how teachers teach grammar? Insights into all these issues can enhance understandings of grammar teaching and contribute to teacher development not only in Japan but also in the field of L2 learning more generally.

## Acknowledgements

We thank the teachers who participated in this study for their cooperation.

*Masateru Nishimuro* holds an MA TESOL degree from the University of Leeds and currently teaches English at Kwansai Gakuin Senior High School in Japan.

*Simon Borg* is a professor of TESOL at the University of Leeds. His interests are language teacher cognition, teacher education, grammar teaching, and teacher research.

## References

- Andrews, S. (2007). *Teacher language awareness*. Cambridge: Cambridge University Press.
- Aspinall, R. W. (2006). Using the paradigm of 'small cultures' to explain policy failure in the case of foreign language education in Japan. *Japan Forum*, 18, 255-274.
- Basturkmen, H. (2012). Review of research into the correspondence between language teachers' stated beliefs and practices. *System*, 40, 282-295.
- Basturkmen, H., Loewen, S., & Ellis, R. (2004). Teachers' stated beliefs about incidental focus on form and their classroom practices. *Applied Linguistics*, 25, 243-272.
- Borg, S. (1998). Teachers' pedagogical systems and grammar teaching: A qualitative study. *TESOL Quarterly*, 32, 9-38.
- Borg, S. (2001). Self-perception and practice in teaching grammar. *ELT Journal*, 55, 21-29.
- 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.
- Borg, S. (2006). *Teacher cognition and language education: Research and practice*. London: Continuum.
- Borg, S., & Burns, A. (2008). Integrating grammar in adult TESOL classrooms. *Applied Linguistics*, 29, 456-482.



- Denscombe, M. (2002). *Ground rules for good research*. Buckingham, UK: Open University Press.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.
- Duff, P. (2008). *Case study research in applied linguistics*. New York: Erlbaum.
- Eisenstein-Ebsworth, M., & Schweers, C. W. (1997). What researchers say and practitioners do: Perspectives on conscious grammar instruction in the ESL classroom. *Applied Language Learning*, 8, 237-260.
- Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly*, 40, 83-107.
- Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford: Oxford University Press.
- Farrell, T. S. C., & Lim, P. C. P. (2005). Conceptions of grammar teaching: A case study of teachers' beliefs and classroom practices. *TESL-EJ*, 9(2), 1-13.
- Gollin, J. (1998). Deductive vs. inductive language learning. *ELT Journal*, 52, 88-89.
- Gorsuch, G. (2000). EFL educational policies and educational cultures: Influences on teachers' approval of communicative activities. *TESOL Quarterly*, 34, 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 from <http://epaa.asu.edu/ojs/article/view/339/465>
- Jean, G., & Simard, D. (2011). Grammar teaching and learning in L2: Necessary, but boring? *Foreign Language Annals*, 44, 467-494.
- Kanno, Y., & Stuart, C. (2011). Learning to become a second language teacher: Identities-in-practice. *The Modern Language Journal*, 95, 236-252.
- Kikuchi, K., & Browne, C. (2009). English educational policy for high schools in Japan: Ideals vs. reality. *RELC Journal*, 40, 172-191.
- Kim, E. J. (2011). An activity theory analysis of a teachers' experience. In K. E. Johnson & P. R. Golombek (Eds.), *Research on second language teacher education* (pp. 225-238). London: Routledge.
- Kuzborska, I. (2011). Links between teachers' beliefs and practices and research on reading. *Reading in a Foreign Language*, 23, 102-128. Retrieved from <http://nflrc.hawaii.edu/rfl/April2011/articles/kuzborska.pdf>
- Nassaji, H., & Fotos, S. S. (2011). *Teaching grammar in second language classrooms*. London: Routledge.
- Newby, P. (2010). *Research methods for education*. Harlow, UK: Pearson Education.

- Nishino, T. (2008). Communicative language teaching: An exploratory survey. *JALT Journal*, 30, 27-50.
- Nishino, T. (2011). コミュニカティブ・アプローチに関する日本人高校英語教師の信条と実践 [Japanese high school teachers' beliefs and practices regarding communicative language teaching: An exploratory survey]. *JALT Journal*, 33, 131-155.
- Nitta, R., & Gardner, S. (2005). Consciousness-raising and practice in ELT course-books. *ELT Journal*, 59, 3-13.
- Pahissa, I., & Tragant, E. (2009). Grammar and the non-native secondary school teacher in Catalonia. *Language Awareness*, 18, 47-60.
- Phipps, S., & Borg, S. (2009). Exploring tensions between teachers' grammar teaching beliefs and practices. *System*, 37, 380-390.
- Sakui, K. (2004). Wearing two pairs of shoes: Language teaching in Japan. *ELT Journal*, 58, 155-163.
- Sakui, K. (2007). Classroom management in Japanese EFL classrooms. *JALT Journal*, 29, 41-58.
- Sato, K., & Kleinsasser, R. C. (2004). Beliefs, practices, and interactions of teachers in a Japanese high school English department. *Teaching and Teacher Education*, 20, 797-816.
- Schulz, R. A. (1996). Focus on form in the foreign language classroom: Students' and teachers' views on error correction and the role of grammar. *Foreign Language Annals*, 29, 343-364.
- Schulz, R. A. (2001). Cultural differences in student and teacher perceptions concerning the role of grammar instruction and corrective feedback. *The Modern Language Journal*, 85, 244-258.
- Storch, N. (2010). Researching grammar. In B. Paltridge & A. Phakiti (Eds.), *Continuum companion to research methods in applied linguistics* (pp. 205-221). London: Continuum.
- Taguchi, N. (2005). The communicative approach in Japanese secondary schools: Teachers' perceptions and practice. *The Language Teacher*, 29(3), 3-12.
- Tomlinson, B. (2007). Using form-focused discovery approaches. In S. Fotos & H. Nassaji (Eds.), *Form-focused instruction and teacher education* (pp. 177-192). Oxford: Oxford University Press.
- Underwood, P. R. (2012). Teacher beliefs and intentions regarding the instruction of English grammar under national curriculum reforms: A theory of planned behaviour perspective. *Teaching and Teacher Education*, 28, 911-925.

## Appendix A

### *Pre-Observation Interview Questions*

#### Section 1: Background

- Age
- Qualifications
- Years of ELT experience

#### Section 2: Experience of learning/teaching grammar

1. How were you taught English/grammar when you were a student?
2. Was a particular approach to grammar promoted in your teacher training courses?
3. The school culture
  - Does the school you work for promote any particular style of teaching (grammar)?
  - Are there any restrictions on the kind of materials you use or on the content and organization of your lessons?

#### Section 3: General questions about grammar teaching and ELT

1. The role of grammar/formal instruction/ELT in general
  - Do you think formal grammar instruction is essential to acquire English?
  - Do you think there should be more formal study of grammar in your English course?
  - Do you feel students receive enough grammar teaching?
  - Do you think grammar teaching should be integrated with other skills teaching?
2. Decisions about what grammar to teach
  - How do you decide what grammar items to teach?
  - The guidelines? Exams? The textbooks? Students' expectation or others?

## Appendix B

### *Post-Observation Interview Questions*

The numbered headings were covered with all three teachers. However, the precise questions varied across the teachers depending on what was observed during the lessons. The questions below are from Endo's interview.

1. The use of L1
  - You used Japanese almost 100% during your lessons. Do you have any reasons for doing so? Or for not using English?
2. Presentation of grammar
  - You presented grammar as it was presented in the coursebook without giving any context. Do you think grammar should be presented with/without context?
  - What was your purpose of using handouts?
  - You made your students use a dictionary. What was your purpose in relation to grammar teaching?
  - You always presented two different sentences to get students to notice the different structures and the usages. Do you always do that?
3. The use of grammar terminology
  - You often used grammar terminology when you were explaining grammar. Do you think it is necessary or useful?
  - Although you told students that they don't need to memorize the terms, you wrote them on the blackboard. What was your intention?
4. Grammar practices
  - You simply used the exercises from the coursebook. Do you sometimes add your original exercises?
  - There were no communicative activities or pair/group activities. Do you think they are necessary for grammar teaching?
  - Do you think students need more exercises?
5. The roles of teachers/students
  - You spent 35 minutes giving a lecture, and 5 minutes for students to work on the exercises. What do you think of this allocation of time? What do you think the teacher's role is?
6. The use of materials
  - You simply followed the coursebook. Do you feel obliged to use exactly what is written in the coursebook?
7. Integration of grammar and skills teaching
  - You did not integrate any skills teaching. Do you think you should teach grammar separately or together with skills?

# EFL Teachers' Learning: Transitional Experiences from an Overseas Teacher Education Program to Japanese School Settings

Yuka Kurihara  
*Tokai University*

Drawing on Vygotsky's sociocultural theory of mind as a conceptual framework, I discuss the nature of L2 teachers' learning to teach. In particular, I examine how EFL teachers who participated in overseas teacher education programs for professional development and returned to their native countries appropriated the pedagogical resources (hereinafter *tools*) presented in the programs into their own classroom instruction. EFL teachers' experiences are complex because their learning involves not only transferring the tools but also negotiating cultural boundaries between the two key contexts. Three secondary school English teachers from Japan, four U.S. program instructors, and two school administrators in Japan participated in this qualitative case study. The cases illustrated that an individual teacher's process of learning and the social world were intricately interwoven and influenced one another. The teachers also attempted to (re)construct new knowledge about ELT by negotiating cross-cultural challenges. Implications for L2 teacher education programs are discussed.

本研究はヴィゴツキーの社会文化理論を枠組みに、英語教員の学びの本質を探る。英語を外国語として教えているEFL教員の中でも、特に海外で教員研修プログラムを経験し現在自国で教える日本人教員が、研修で学んだツール(知識やスキル)を日本の現場でどのように自分のものとして使用しているかについて調査した。教員は異なる環境へのツールの移行や、互いの文化的側面を乗り越え折り合いをつけようとする、複雑な学びを経験する。この質的研究には、中等教育の日本人英語教員3名、米国の教員研修プログラム関係者4名、そして日本の学校管理職2名が参加した。結果、教員の学びとは、個々の学びの過程とその社会環境が密接に関

係し互いに影響をもたらしながら起こっていることが明らかになった。また、教員は文化的差異から生じる問題を解決する際、新たな英語教授アプローチを見出そうしていることも分かった。最後に教員研修への応用を述べる。

**I**n the last 20 years, there has been a growing trend among EFL teachers toward participation in L2 teacher education programs in North American universities for professional development, after which they generally return to their native countries to teach. Given this circumstance, it is critical to understand the nature of EFL teachers' learning to teach, in particular, how they adapt the pedagogical resources (hereinafter *tools*) presented in the programs into their teaching settings. By using Vygotsky's sociocultural theory as a conceptual framework (Grossman, Smagorinsky, & Valencia, 1999; Vygotsky, 1930s/1978), I investigated the transitional experiences of three Japanese EFL teachers newly trained in the U.S. as they reacclimated to their teaching situations in Japan. Because of its focus on the relationship of teachers with their contexts, Vygotsky's theory helps researchers better understand not only what sources contribute to change, but also how teachers deal with possibly competing goals and practices emphasized in various settings, and how they influence the contexts in which their learning to teach occurs. This qualitative case study, conducted for approximately 10 months, provides insights into how the teachers appropriated pedagogical tools in conjunction with their internal goals and histories, and how these are closely related to the social contexts where their learning to teach occurred. The results of this study may further allow us to better understand the influence of L2 teacher education.

## Review of the Literature

### *The Knowledge Base of L2 Teacher Education*

A central topic in recent L2 teacher education literature is what constitutes a professional knowledge base in the development of effective programs for English language teachers (Freeman & Johnson, 1998; Grabe, Stoller, & Tardy, 2000; Johnson, 2006; Johnston & Goettsch, 2000; Yates & Muchisky, 2003). Some consider that the core of the field should be based on a range of disciplines such as applied linguistics and second language acquisition (e.g., Yates & Muchisky, 2003). They further emphasize that exposure to the subject matter should be central so as to enable L2 teachers to conduct informed teaching practice. However, since the late 1980s, this view has been challenged by constructivists who seek to understand the complex nature

of L2 teachers' learning to teach by exploring their cognition (e.g., Woods, 1996). Drawing on sociocultural theory, scholars have further discussed the influential effects of the social world, such as schools, on teachers' learning to teach (e.g., Freeman & Johnson, 1998).

Although studies from sociocultural perspectives have portrayed the dynamic nature of teachers' learning to teach, studies which investigate EFL teachers, in particular those who participate in overseas L2 teacher education programs, are scarce. Their learning experiences as they move from overseas programs to their native teaching settings need to be examined so that we can add new insights to the field of L2 teacher education.

### ***Vygotsky's Sociocultural Theory of Mind***

A strand of research in social science currently draws on Vygotsky's sociocultural theory of mind as its conceptual framework. This theory posits that people learn through their social engagements, which are embedded in historical, cultural, and social conditions (Lantolf & Thorne, 2006; Morita, 2002; Vygotsky, 1930s/1978; Wertsch, 1991). The theory also emphasizes a critical role for the mediation of human activity by using physical and psychological tools (e.g., Ellis, Edwards, & Smagorinsky, 2010). Tools mediate "social and individual functioning and connect the external and the internal, the social and the individual" (John-Steiner & Mahn, 1996, p. 192).

Recent literature on teacher education in the L1 field explores the implications of Vygotsky's sociocultural theory (e.g., Ellis et al., 2010; Johnson, Smagorinsky, Thompson, & Fry, 2003). This research reveals that teachers' learning to teach encompasses multiple activity settings, including university-level teacher education course work, field practicums, school contexts, and prior experiences, so their learning experiences may turn out to be distinctly different based on the relationships among these settings (e.g., Grossman et al., 1999). Each setting has its own cultural history throughout which members of the community have established specific goals that guide their actions within the setting (e.g., Lave & Wenger, 1991). At the same time, because of their internal goals and histories, individual teachers are not just passive subjects who only reproduce "culturally valued concepts" (Edwards, 2010, p. 64).

Vygotsky's sociocultural theory also helps us recognize the roles of teachers' use of pedagogical tools which mediates their learning to teach (Grossman et al., 1999). Newell, Gingrich, and Johnson (2001) argued that rather than focusing on static internalization of knowledge transmitted during the

course work, Vygotsky's theory provides a framework for examining teachers' "appropriation" of pedagogical tools, which, as Wertsch (1998) explains, is "the process . . . of taking something that belongs to others and making it one's own" (p. 53). The scholarly discussion in these studies implies the need to explore L2 teachers' appropriation of pedagogical tools so that new insights can be added to existing teacher education research.

## The Study

### *Research Questions*

The main purpose of this study is to better understand the nature of EFL teachers' learning to teach by highlighting their appropriation processes within and across a variety of activity settings in which their learning to teach occurred. These settings included teacher education course work in the U.S., school settings in Japan, cross-cultural ELT issues, and teacher beliefs and backgrounds. A unique aspect of this present study is the examination of EFL teachers' transitional experiences as they cross from one setting to another. The study explores how teachers deal with possibly overlapping or competing values (or both), goals, and practices stressed in various activity settings and how they are being shaped by and shape the social settings. The following research questions were identified to achieve this aim:

1. What appropriation takes place when Japanese EFL teachers adopt pedagogical tools presented in the U.S. teacher education programs into their classroom instruction in Japan?
2. What sources and settings have influenced their appropriation?

## *Method*

### *Participants and School Settings*

The primary participants in this study were three Japanese senior high school teachers of English who had completed a 6-month or 1-year in-service teacher education program in the U.S. sponsored by the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT). These teachers were selected from questionnaire respondents who expressed interest in the qualitative case study. I employed an "intensity sampling" strategy to learn as much as possible from the "information-rich cases that manifest the phenomenon of interest intensely" (Patton, 2001, p. 234). The teachers—Mr. Fujii, Mr. Suzuki, and Mr. Kato (pseudonyms)—were all born in Japan and received their formal education before or right after the implementation of



EFL curriculum reform, which stresses the communicative-based pedagogy promoted by the Japanese government.

Fujii participated in a 6-month MEXT program in 1998 while at his first school of employment in northwestern Japan. Before, during, and after the program, he had various professional development opportunities inside and outside school. His current high school was located in a residential area. The students were “academically high” and most of them entered either universities or colleges upon graduation (Fujii interview, 19 Oct 2005).

Suzuki, who participated in a 12-month MEXT program in 1998-1999, transferred to his high school after the completion of the program. The school, located in a small rural area in the westernmost region of the main island of Japan, was designated as a Super English Language High School (SELHi). SELHi is a 3-year project promoted by the Japanese government in which each designated school has a responsibility to create an English curriculum and share it with the public (MEXT, 2004). According to his school’s brochure, the main goal of the program was to develop highly practical English communication skills, especially self-expression, through English education and international understanding.

Kato attended a 6-month program in 2003. His high school was located in a large commercial city in western Japan and was characterized as a “historical school.” According to the school principal, the school was implementing educational reform to prepare students for university examinations by letting them focus on specialized subjects.

Table 1 contains a summary of the participants and their school settings.

### *MEXT Programs*

Another key setting of the participants’ learning to teach was MEXT, which had shifted the goals of English education at the secondary school level from fact-oriented teaching to a communicative-oriented approach (MEXT, 1999, 2009). Accordingly, Japanese English teachers’ professional development opportunities inside and outside Japan were greatly increased. For example, from 1998 through 2003, MEXT sent secondary school Japanese EFL teachers to a total of 12 state or private universities in the U.S. Individual universities created programs for the participants that satisfied the requirements proposed by MEXT (CIEE, 2003). The main goals of the programs were to improve Japanese teachers’ English language skills, to better understand English language teaching methodology and its application, to broaden understandings of the various cultural and social aspects of the U.S. (and other countries), and to conduct research within the relevant EFL area (CLED, 1999, 2004).

**Table 1. Summary of Case Study Participants and School Settings**

Teacher	Fujii	Suzuki	Kato
Years of teaching	12	18	20
Degree	BA in Linguistics	BA in English Literature	BA in English Language
School characteristics & goals	Previous school	English focused course	English specialized course
	Current school	“Zest for living”; University entrance examinations	English focused school; University entrance examinations
Class size - course title	24 - 11th grade comprehensive English;	39 - 12th grade reading;	27 - 12th grade writing;
	41 - 11th grade writing	9 - 12th grade communication	27 - 12th grade reading
Responsibilities	English text developer	Head of the SELHi program & the English department; Graduate student in a distance TESOL program in England	Manager of the general affairs at school

I chose two U.S. university MEXT-sponsored programs for this study: one 6-month and one 12-month; among the U.S. programs, they had the largest number of participants. These programs were held in university-affiliated English language schools within large research-based institutions. The course work offered in each program was institutionally planned. For example, one university stressed “language learning as a social construct” (Program coordinator interview, 8 Aug 2005). In the methodology class, the participants were required to read a textbook, *The Practice of English Language Teaching* by Jeremy Harmer, and other selected materials containing both practical (e.g., communicative activities) and theoretical knowledge, including an overview of learning theories and ELT approaches as well as concepts such as multiple intelligences. This course also had the participants observing foreign language classrooms in local middle and high schools, reflecting on their own learning, conducting a teaching project, and considering the

potential application of communicative methods in their current teaching situations.

The other program emphasized “communicative methodology and task-based learning” (Program coordinator interview, 13 Sep 2005). The reading texts required in the methodology course included *How Languages are Learned* by Patsy Lightbown and Nina Spada and *Techniques and Principles in Language Teaching* by Diane Larsen-Freeman. These textbooks contained both theoretical knowledge and practical ideas about ELT (e.g., an overview of various teaching methodologies, theories of L2 learning, and effective teaching techniques), and were used to develop lesson plans that adapted the required English textbooks in Japan to make them more communicative. Teachers also spent about one week in a local secondary school to observe ESL classes.

### *Data Collection*

For the study, I employed a qualitative case study approach by triangulating multiple methods such as classroom observations, interviews, teacher reports, and questionnaires. Data were collected from August 2005 through May 2006. At the beginning of the study, questionnaires were administered to ascertain teachers’ professional and academic backgrounds and to trace the influence of their past experiences on their ELT conceptions and practices. Another important set of data came from classroom observations of the primary participants’ English teaching practices in each school setting in Japan and debriefings about my observations. These data were collected by visiting schools in October 2005 (two to three observations per teacher), and then through reports by the teachers on an approximately bimonthly basis (two to five reports per teacher). During the classroom observations, I took field notes of the participants’ lessons and described the sequence of teaching activities in class, the content and approaches they used, and the interactions between students and teachers. To understand teachers’ appropriation of pedagogical tools in more depth, audiotaped semi-structured interviews were also conducted at the beginning and the end of the study. In addition, four coordinators or instructors of the two MEXT programs and the school administrators in two of the schools where the teachers were employed each gave me permission to conduct an interview. All the data, except for the interactions with program hosts, were collected in Japanese. I also reviewed written documents obtained from the host programs such as course syllabi and the participants’ reflective journals written in English, which were required by the host programs.

### Data Analysis

The processes of analyzing the qualitative case study involved seeking out salient patterns, categories, themes, and theories (Patton, 2001). Vygotsky's sociocultural perspectives on teacher education based on Grossman et al. (1999) also informed the data collection and analysis of this study. To build an empirically based case portrait of each teacher, I collaboratively coded the transcripts of the data with two peer debriefers.

The coding system was developed by following Johnson et al.'s (2003) study. The categories used in the data analysis were *pedagogical tools*, *teaching areas*, *source of tools*, and *problems*, which are derived from Vygotsky's sociocultural theory. Regarding pedagogical tools, two types, *conceptual* and *practical*, were used to understand the tool-mediated nature of instruction given by the teachers. Conceptual tools included *problem-solving learning* and *collaborative learning*; and practical tools included *group work activities* and *Ministry-approved textbooks*. Teaching areas in which the teachers employed particular tools included *reading*, *writing*, *speaking/listening*, *language*, and *management*. The source of tools was where the teachers had learned a particular tool, including *prior experiences*, *school contexts*, and the *MEXT program*. Finally, to understand the teachers' goal-oriented actions, I also identified what problems they had attempted to solve through their use of particular tools. The problems included *students' learning*, *the application of pedagogical tools* presented in the MEXT programs, and *school goals* at their teaching sites. Based on these results, tentative assertions about each teacher were tested. Finally, cross-case analysis was conducted to explain the similarities and differences across the teachers' cases.

## Findings

### **An Alignment Among Learning Settings: Fujii's Progress**

To the rank-ordering question on the questionnaire (see Appendix), Fujii responded that professional development opportunities inside and outside Japan and his teaching sites are the most influential factors that have affected his ideas about teaching. The interview and classroom observation data also suggest that, when he adopted the pedagogical tools presented in the MEXT program, these settings interactively mediated his appropriation of the tools.

Prior to the MEXT program, he was somewhat familiar with those concepts of sociocultural theory which value social interaction of learning. Referring to a pedagogical tool, language use in real-life situations, he explained from

whom he had learned this concept of language teaching and how he used it in his practice:

I came to attempt to make learning relevant to students' real-life situations by discussing this with a professor in study meetings . . . I started to include the idea before participating in the [MEXT] program . . . it is perfectly natural to follow what Vygotsky suggests, so I often provide senior students with a theme of problem solving. (Fujii interview, 12 May 2006)

*Other professional development opportunities* which were found as the source of his pedagogical tools suggest Fujii had actively sought knowledge about ELT outside of the MEXT program. However, in problem-solving activities he had used before participating in the MEXT program, he had struggled with students' frequent avoidance of responses, their using "I don't know" (Fujii interview, 12 May 2006). He described his classroom instruction around that time as follows: "I taught English in my own way before the [MEXT] program and was unsatisfied every day" (Fujii interview, 19 Oct 2005).

The data suggest that he further developed an understanding of sociocultural theory, especially tool-mediated learning and social interaction, as a result of participating in the MEXT program. While observing one of the local classes, for example, he realized the significance of drawing for students' learning.

During the activity [drawing part of the story with peers], they were actively discussing it with each other by saying, "This mountain should be higher and that house should be bigger." I learned that participatory learning enabled quiet students to be actively involved in class. (Fujii interview, 12 May 2006)

As can be seen in this excerpt, Fujii appropriated the tool, *visual and auditory aids*, by considering a principle, *peer learning*, which he had learned in the MEXT program.

A challenge that Fujii and other teacher participants faced during the MEXT program, however, was English learning in EFL contexts where students have few opportunities to expose themselves to the target language beyond the classroom. Their reservations about whether the ideas about ESL teaching presented in the program would work in Japan allowed them to reconsider the issue and eventually they had new insights:

There [in the US], English is being used outside the classroom . . . . The other participants and I always discussed what we could do to marry the two [ESL and EFL]. Then, an idea came up, or the only choice we thought we could have is that, due to limited English vocabulary and learning environments [in Japan], we should not think of English learning opportunities as a separate activity such as reading as just a reading class . . . . The expressions and ideas they learn in one setting are transferable to other contexts. (Fujii interview, 12 May 2006)

This excerpt suggests that the participants collaboratively developed the idea that the principle of *integrating skills and courses across settings* supports EFL students' learning because they can utilize and apply their knowledge in a range of new contexts.

Reflecting on the time when he returned from the program to his teaching site in Japan, he commented, “[Back at my] previous school, I tried out almost everything of what I had wanted to do and ideas that I had had in my mind during the program” (Fujii interview, 28 Aug 2005). My classroom observations also suggest that he intentionally had his students use the four English skills of reading, writing, listening, and speaking within each lesson and linked their work inside the classroom to outside (Field note, 19 Oct 2005). For example, to make grade 11 students' learning more effective in his comprehensive English class, he used a variety of reading materials about World Heritage sites as *tools to think*, and asked them to discuss with peers in English the good and bad points about a recently designated site in Japan. The main conceptual framework for using this group work included peer learning and the use of real-life situations. The primary goal of this activity was to fashion learning that helped students “discover what they share and what they come to newly realize” by exposing them to ideas that are different from their own (Fujii interview, 12 May 2006). Referring to *eishaku* [language training] in the same lesson, Fujii further explained how this practical technique also helped his students connect their language skills gained in the practice with the problem-solving activity in which they engaged:

By having my students repeatedly read [texts] aloud, I want to provide them with training opportunities within 50-minute lessons . . . . Through such activities, I want my students to learn and remember as much useful English as possible. Then, students can borrow the language from there and connect the knowledge to their actual utterances. I learned this idea in

study group meetings in Japan [after the MEXT program] and keep adapting it myself. (Fujii interview, 12 May 2006)

Fujii's understanding of these conceptual and practical tools evolved as he realized, reflected upon, and tested the problems about his teaching practices (e.g., few opportunities for students to use English beyond the classroom in Japan) by making use of the activity settings available to him such as the MEXT program, his school sites, and study meetings. Through this problem-solving process, he eventually constructed new knowledge about ELT.

### ***From Self-Serving to Goal-Oriented Teaching: Suzuki's Negotiation***

In the rank-ordering portion of the questionnaire (see Appendix), Suzuki responded that two activity settings, professional development opportunities (in particular the MEXT program) and his teaching sites at schools, had influenced his ideas about teaching the most. However, the findings of the case study suggest that multiple activity settings interactively mediated his appropriation of the pedagogical tools. For example, his educational background as an EFL learner also seemed to be influential in his instructional decision making. He explained:

During my 3 years at high school, I studied English so hard. Right after entering university, however, I stopped. In summer I planned to take the STEP Eiken [English proficiency exam] and prepared for that. But I found I did not remember any words in the vocabulary book I had studied just 3 months before. The reason why is because I did not use what I studied. (Suzuki interview, 23 Mar 2006)

This excerpt suggests that all the efforts he had made to pass university entrance examinations (e.g., memorizing words) when he was a high school student did not contribute to his later English learning, but shaped his English teaching principles. That is, the actual use of English knowledge is indispensable for EFL learning.

This conceptual principle seemed to be further reinforced in the MEXT program, in which he was exposed to instructional techniques in reading and listening, including pre- and postreading activities. The following conversation between Suzuki and me illustrates how his educational background and the pedagogical tools newly acquired in the program were eventually connected:

Kurihara: You noted in your report that the idea about learning English through expressing what you learned came from not only your own experiences in the past but also an influence from Dr. Lidgley (pseudonym) in the MEXT program. Could you elaborate on that?

Suzuki: I learned about receptive skills, reading and listening [in the methodology class]. For example, reading includes pre-, while-, and postreading. I learned that, instead of completing [reading] just with [while-] reading, it is good to include a follow-up activity . . . . The idea [about actual use of English] came from this experience. (Suzuki, Report 4, 2 Feb 2006)

These comments suggest that Lidgley's methodology course had enabled Suzuki to realize the significance of using these practical tools (e.g., post-reading activities) grounded in the main conceptual framework, actual use of language.

His retrospective interviews, however, suggest that when Suzuki made the transition from the MEXT program to his teaching site in Japan, he experienced moments of disappointment and challenges:

Around that time [right after returning to my teaching site], I mainly taught reading classes. I had my students work on very conventional university-driven workbooks [assigned to the class] and included what I called Culture Time in the interval. During the Culture Time, I let them listen to real news and music, and showed TV commercials . . . . It was like making use of whatever I had learned in the MEXT program . . . . After my students entered the 3rd year, I realized that their reaction toward my teaching became negative . . . . Last year, I struggled with it. I'm afraid my rambling instruction caused this consequence. (Suzuki interview, 27 Oct 2005)

This excerpt suggests that for the first few years after the completion of the MEXT program, Suzuki's inclination toward EFL teaching was to try out "whatever [he] had learned in the MEXT program" and he eventually experienced dismay about his "rambling instruction." The 4th year after the program, however, he became a coordinator of the SELHi project and started to create the English curriculum with his colleagues. This collaborative work eventually allowed him to see the benefits of using pedagogical tools that he



had learned in the MEXT program to develop more sequential teaching approaches. Reflecting on the framework of reading and listening instruction learned from Lidgley, he commented thus:

I am now able to teach by taking the future of my students into consideration. This includes what would it be after a 1-hour lesson of a three-credit course as a short-term consideration, and also after 1 year and 3 years as a long-term consideration . . . . It was after I started to make the syllabi of English courses with other English teachers [that I had this vision]. (Suzuki interview, 29 Aug 2005)

This excerpt suggests that Suzuki transformed his professional self from a teacher who was self-serving with his instruction to a teacher who takes students' learning into consideration by framing his instruction with teaching goals in mind.

Throughout the study, Suzuki consistently distinguished *teaching to examinations* and *teaching beyond examinations*, which he and his colleagues attempted to conduct at his school. For example, in his 12th grade reading class, after he had explained English grammar points (idioms) to students, which according to him often appear on university entrance examinations, he asked the students to write English sentences of up to 20 words using each idiom as a postreading activity. Students were required to write sentences relevant to their own lives (Suzuki, Report 5, 9 Feb 2006). Suzuki explained the reasons behind this activity:

[Even] if students have finished the lesson by learning the grammar points in the workbook, they will not have situations where they can actually use what they learned. When they express English with the knowledge, it finally becomes their own. I offer my lessons based on this idea. (Suzuki, Report 5, 9 Feb 2006)

Suzuki appropriated a pedagogical tool learned in the MEXT program, a postreading activity, to facilitate students' language learning. When using the tool, he took into consideration the goal of the SELHi curriculum (to develop students' self expression), the responsibility to successfully send students to universities, and his desire to have them actually use English in the lessons.

### **“Going Back to My Original Teaching”: Kato’s Challenges**

To the rank-ordering question in the questionnaire (see Appendix), Kato responded that his learning experience in the MEXT program is what has most influenced his teaching ideas. However, his appropriation of the pedagogical tools presented in the program involved great challenges because of seemingly incompatible goals across and within his main activity settings. These settings included his pre-training experiences as an EFL teacher, the MEXT program, and his school culture.

In the 12th grade Writing and Reading lessons I observed, for example, Kato prepared a worksheet and had his students fill in the blanks with answers to workbook questions, key grammar points he had explained, and the Japanese translations of English words or expressions (Field note, 17 Oct 2005). He called this activity *sagyō* [performance] and explained the possible sources of his instructional decision to use this tool:

There are many students who do not think of English as being necessary and who are not very good at it, so I could not expect effective learning results from lecture-styled teaching by giving the students lots of knowledge. As a result, I decided to use an effective learning method that I had used when I worked at a night school before . . . . To get the information they wanted to know about the story, they did translating *sagyō* by actively using a dictionary. (Kato, Report 1, 28 Oct 2005)

As can be seen in this excerpt, Kato’s professional background prior to the MEXT program, especially working at a night school, as well as his concerns about students’ learning seemed to play an important role in guiding his instruction.

During the MEXT program, this previously shaped teaching perspective appears to have had a significant effect on his learning and understanding of what was presented in the program. The following retrospective interview suggests that his experiences in the ESL coursework in the program further reinforced his belief about his EFL approach to *sagyō*:

In the [ESL] class, we did not learn grammar only from a grammar book, but learned through actual writing, actual reading, and actual presentations. By doing so, [we found that these practices] gradually became ingredients . . . . It is not just teachers’ lecturing but students’ doing actual *sagyō* . . . . Through

doing *sagyo* over and over, students will be able to come to use it [English]. (Kato interview, 27 May 2006)

This excerpt suggests that Kato carried over his prior knowledge about *sagyo* to the MEXT program and integrated it with another new tool, which he called “actual practice of language over and over” (Kato interview, 27 May 2006).

When he made the transition from the MEXT program to his teaching site in Japan, Kato experienced feelings of being torn between the seemingly incompatible goals of the two settings, his school and the MEXT program:

Right after I came back from the U.S., . . . I did a group work activity in my class once, but stopped using it . . . In the end, group work did not work well in Japan due to the reality of university entrance examinations. I discussed it [group work] in my research project during the program, but what I wrote in the project and the teaching instruction to the exam do not match well. (Kato interview, 17 Oct 2005)

This excerpt suggests that once Kato resumed EFL teaching in Japan, his school became a key activity setting that affected his appropriation of the pedagogical tools. According to Kato, “the highest expectation of [my] school is probably to successfully prepare students for university entrance examinations” (Kato interview, 27 May 2006). This ultimate goal eventually compelled him to focus on “what would be useful in [his] school and what would be appropriate to meet [his] school expectations” (Kato interview, 17 Oct 2005). As a result, he selectively used pedagogical tools such as “*sagyo* over and over,” because these activities allowed him to fulfill his professional responsibilities at the school and his desire to try out the pedagogical tools he had learned in the MEXT program.

Kato eventually developed a sense of doubt about his own teaching approach. For example, reflecting on the last 2 years since he returned from the MEXT program to his teaching site, he commented on the reshaping of his professional self:

I returned to Japan with a feeling that I wanted to conduct lessons using this kind of practice and that kind of practice. I also thought what teaching I could offer to my students with the English skills I had developed. But the environment surrounding me has not changed, so [my instruction] is gradually going back to my original teaching. (Kato interview, 17 Oct 2005)

Throughout the study, I found that Kato frequently expressed concerns about the impact that his instruction may have on his students' learning. Reflecting on his 12th grade reading lessons, he noted the following:

As long as I offer a lesson and give them an exam, my students can get points, but to be able to get points has been a priority rather than feeling fulfilled through lessons . . . I am worried that I am creating students who patiently study English even though they suffer and hate to learn it. (Kato, Report 2, 17 Jan 2006)

Although Kato expressed a positive point about participating in the MEXT program in that he had gained confidence in his English language skills, these excerpts suggest that he felt disappointed when he realized that the established school goals did not allow him to try out the tools he had learned in the program. Unlike the other two participants, because Kato could not trial what he had learned in the MEXT program, he could not thereby reflect on his pedagogical application inside or outside school. Perhaps the lack of these opportunities was partly due to his commitment as a manager in the general affairs department which "took most of his professional time" and "made it difficult for [him] to design the curriculum of English education as a main English teacher for a certain grade" (Kato interview, 17 Oct 2005).

## Discussion

This study illustrated how three Japanese EFL teachers took various paths in appropriating pedagogical tools as they made the transition from a U.S. teacher education program to their teaching sites in Japan. The results suggested that the teachers' appropriation of tools depended on their histories and the professional experiences they brought to the given contexts, as well as their relationship with social settings, including the MEXT program (e.g., what pedagogical tools the program made available to them, what they took from the program), school culture (e.g., goals, students' needs), and other professional development opportunities (e.g., study group meetings). This finding follows Vygotsky's sociocultural theory that humans' cognitive development is simultaneously unique to individuals and socially constructed (e.g., John-Steiner & Mahn, 1996). Thus, as Morita (2002) contends, it is unproductive to treat individuals and social contexts as separate entities.

The goals, values, and practices emphasized in the settings overlapped and competed across and within the settings in dynamic ways, so they affected

the teachers' appropriation processes. Although all the teachers, to greater or lesser extents, experienced moments of both making progress and having challenges, when there was an alignment of goals and practices stressed in the settings, the teachers tended to be able to draw on what they had learned from the MEXT program (see Grossman et al., 1999 for a similar discussion). However, in cases of competing goals and practices among the settings, they appeared to encounter challenges in appropriating the tools into their teaching sites (see Grossman et al., 1999 for a similar discussion).

Because the teachers' learning involved crossing cultural boundaries, they attempted not only to transfer the instructional tools but also to negotiate various demands emphasized in the key settings in order to meet the needs of their teaching sites. The challenges they faced included integrating institutionally and nationally established cultural norms of what is good learning and their newly constructed assumptions about English teaching and learning. However, the case study data suggest that the teachers did not simply adopt one cultural aspect of English teaching (e.g., the group work promoted in the U.S. MEXT program) and ignore another one (e.g., university entrance examinations considered to be important in schools in Japan), but they attempted to negotiate the problems derived from the different sets of values in the two settings (see Morita, 2002 & 2004, for similar discussions about L2 learners' identity construction). It was through this problem-solving process that the knowledge they gained in the MEXT program, the responsibilities they had in their teaching sites in Japan, and their own goals for classroom instruction became intertwined. Two participants eventually constructed pedagogically stronger and more context-sensitive ELT approaches. Fujii and Suzuki attempted to negotiate the differences and eventually adapted some of the tools presented in the MEXT program in ways that worked in their own teaching contexts. Constraints they faced in light of cross-cultural boundaries proved "a positive set of limitations that provide[d] the structure for productive activity" (Grossman et al., 1999, p. 7). Teachers are not only passive recipients of culturally valued goals and practices in given contexts but also actors having an impact on them (Edwards, 2010).

The findings also suggested that the teachers' appropriation processes involved constructing their professional selves within the settings in which their learning to teach occurred. A theme that frequently emerged in the three teachers' cases was "the reciprocal relationship between their appropriating the tools and their sense of professional selves" that developed as they applied the tools presented in the MEXT program to their classroom instruction in Ja-

pan (G. E. Newell, personal communication, 14 Mar 2007; and for L2 learners' negotiating identities, also see Morita, 2002, 2004). For example, Kato, who struggled with the incongruity between the goals of his school and his desire to use the tools learned in the program, stopped trying out some of the tools and eventually developed a sense of doubt about his own teaching approach. This relatively negative professional self that he constructed through doubting his own teaching seemed to make it more difficult for him to appropriate the tools in the school culture. In contrast, Suzuki, who had also found himself torn between the two settings for the first few years, collaboratively developed the English curriculum with his colleagues and began to enthusiastically appropriate the pedagogical tools presented in the program into his classroom lessons. Around that time, his professional self was transformed from a single player with "self-serving instruction" to a team player with a more coherent sequence of teaching practices. That broader social relationships and identities constantly interact and mutually shape one another is also something Lave and Wenger (1991) discuss.

## **Implications**

This study offers implications for L2 teacher education regarding how to foster teachers' appropriation processes and eventually develop effective teacher education programs for EFL teachers. First, rather than assuming that providing teachers with training readily promotes new ways of thinking and behavior, teacher educators need to consider the teaching conceptions and practices teachers have constructed and how these elements interact with other elements in their learning contexts (Ellis et al., 2010; Grossman et al., 1999; Newell et al., 2001; and for a similar discussion, see Morita, 2002, 2004) such as school culture, teacher education, cross-cultural negotiation, and professional self. Viewing teachers' learning as a complex process involving various sources and settings allows us to provide additional insights into the discussion about what constitutes a professional knowledge base when developing effective programs. That is, the effects of teacher education need to be viewed with other relevant activity settings in which teachers' learning to teach occurs because of the interconnected relationship between teacher education programs and other elements in their learning contexts (see Grossman et al., 1999 for a similar discussion).

Second, however, teacher education programs can play a critical role in (re)shaping teachers' teaching conceptions and practices if they provide teachers with activity settings to examine their beliefs about EFL teaching and learning, understand pedagogical tools, try out the tools in their teaching

sites, and reflect on their pedagogical application. In this study, the teachers' learning seemed to be promoted when they had these opportunities in the processes of appropriating the tools.

The findings of this study further suggest that ongoing access to the appropriation of tools after teacher education programs is necessary to help teachers deal with competing goals and practices stressed in various learning settings. Such continuing access would allow them to better understand the instructional challenges they are faced with and critically examine their classroom practices. This would eventually result in better ELT instruction for their students.

## Acknowledgements

I would like to thank the Japanese secondary school EFL teachers who participated in this study. I am also thankful to Keiko Samimy, George Newell, and Kathleen Romstedt for their support throughout this study. My gratitude also goes to Naoko Morita who inspired me to conduct research.

*Yuka Kurihara* is an associate professor in Center for Liberal Arts Education at Tokai University. Her research interests include L2 teacher education, educational psychology, and L2 literacy.

## References

- CIEE (Council on International Educational Exchange). (2003). *Ministry of Education-Japanese Teacher Program: Campus Consultation 2003*. Boston, MA: CIEE.
- CLED (Georgetown University Center for Language Education and Development). (2004, 1999). *Japanese Ministry of Education study in the USA, one year teacher training program at Georgetown University, Center for Language Education and Development*. Washington, DC: Georgetown University.
- Edwards, A. (2010). How can Vygotsky and his legacy help us to understand and develop teacher education? In V. Ellis, A. Edwards, & P. Smagorinsky (Eds.), *Cultural-historical perspectives on teacher education and development* (pp. 64-77). New York: Routledge.
- Ellis, V., Edwards, A., & Smagorinsky, P. (2010). Introduction. In V. Ellis, A. Edwards, & P. Smagorinsky (Eds.), *Cultural-historical perspectives on teacher education and development* (pp. 1-10). New York: Routledge.
- Freeman, D., & Johnson, K. E. (1998). Reconceptualizing the knowledge-base of language teacher education. *TESOL Quarterly*, 32, 397-417.

- Grabe, W., Stoller, F. L., & Tardy, C. (2000). Disciplinary knowledge as a foundation for teacher preparation. In K. J. Hall & G. W. Eggington (Eds.), *The sociopolitics of English language teaching* (pp. 178-194). Clevedon, UK: Multilingual Matters.
- Grossman, P. L., Smagorinsky, P., & Valencia, S. (1999). Appropriating tools for teaching English: A theoretical framework for research on learning to teach. *American Journal of Education, 108*, 1-29.
- Johnson, K. E. (2006). The sociocultural turn and its challenges for second language teacher education. *TESOL Quarterly, 40*, 235-257.
- Johnson, S. T., Smagorinsky, P., Thompson, L., & Fry, G. P. (2003). Learning to teach the five-paragraph theme. *Research in the Teaching of English, 38*, 136-176.
- John-Steiner, V., & Mahn, H. (1996). Sociocultural approaches to learning and development: A Vygotskian framework. *Educational Psychologist, 31*, 191-206.
- Johnston, B., & Goettsch, K. (2000). In search of the knowledge base of language teaching: Explanations by experienced teachers. *The Canadian Modern Language Review, 56*, 436-468.
- Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford: Oxford University Press.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.
- MEXT (Ministry of Education, Culture, Sports, Science and Technology). (2004). *Heisei 18-nendo sūpā ingurishshu rangēji haisukūru (SELHi) no kettei ni tsuite* [Decision about Super English Language High Schools (SELHi) in 2006]. Retrieved from [http://www.mext.go.jp/b\\_menu/houdou/18/04/06041201.htm](http://www.mext.go.jp/b_menu/houdou/18/04/06041201.htm)
- MEXT (Ministry of Education, Culture, Sports, Science, and Technology). (1999, 2009). *Kōtōgakkō gakushū yōryō kaisetsu* [The explanation of the course of study for upper secondary school]. Tokyo: MEXT.
- Morita, N. (2002). *Negotiating participation in second language academic communities: A study of identity, agency, and transformation* (Unpublished doctoral dissertation). The University of British Columbia, Vancouver, Canada.
- Morita, N. (2004). Negotiating participation and identity in second language academic communities. *TESOL Quarterly, 38*, 573-603.
- Newell, G. E., Gingrich, S. R., & Johnson, B. A. (2001). Considering the contexts for appropriating theoretical and practical tools for teaching middle and secondary English. *Research in the Teaching of English, 35*, 302-343.
- Patton, Q. M. (2001). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage.



- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. In M. Cole, V. John-Steiner, S. Scribner, & E. Soubberman (Eds.), Cambridge, MA: Harvard University Press. (Originals works published 1930s)
- Wertsch, J. V. (1991). *Voices of the mind: A sociocultural approach to mediated action*. Cambridge, MA: Harvard University Press.
- Wertsch, J. V. (1998). *Mind as action*. New York: Oxford University Press.
- Woods, D. (1996). *Teacher cognition in language teaching*. New York: Cambridge University Press.
- Yates, R., & D. Muchisky. (2003). On reconceptualizing teacher education. *TESOL Quarterly*, 37, 135-147.

## Appendix

### Part III Teaching Principle

*Questionnaire – Teachers’ Post-Overseas Training Experiences (Relevant Section of Longer Questionnaire)*

I would like to know your daily English teaching practices and beliefs or principles in teaching. Please provide your answers in the space below each question in this section.

- 1a) First, I would like to know your typical English lesson. Please describe how you recently go about teaching English in class. Please try to be specific.
- 1b) Please explain why you teach that way. Please be sure to include your own beliefs (e.g., principles, values) as a teacher.
2. Please recall your formal educational experiences in learning and teaching English. To what extent do you think each of the following experiences has affected your beliefs as English teacher? Please circle the number that best describes the degree of affect with each statement.

*Degree of affect*

- |                              |                               |
|------------------------------|-------------------------------|
| 4. Affected me a great deal  | 3. Affected me to some extent |
| 2. Affected me only a little | 1. Did not affect me at all   |

- |   |   |   |   |   |
|---|---|---|---|---|
| a. Secondary school courses when you were a student | 4 | 3 | 2 | 1 |
|---|---|---|---|---|

b. Undergraduate (graduate, if any) level (teacher education) courses	4	3	2	1
c. Teaching practicum conducted during the undergraduate teacher education program	4	3	2	1
d. Teacher development opportunities other than No. 2 (e.g., domestic short-term training, overseas MEXT programs, conferences)	4	3	2	1
e. Teaching experiences at school environments (e.g., school, students' parents', and/or administrators' expectations, colleagues)	4	3	2	1
f. National policies described in "The Course of Study"	4	3	2	1

3. Please describe the most influential event(s) (including any other life experiences) that have/has affected your beliefs as an English teacher.

# 学内開発プレイスメントテスト得点解釈と使用 の妥当性の評価について

## Evaluating Validity for In-House Placement Test Score Interpretations and Uses

熊澤孝昭(くまざわ たかあき)

Takaaki Kumazawa

関東学院大学

*Kanto Gakuin University*

Validity theory has evolved dramatically in the past few decades. The most prominent theory in recent years is an argument-based validity framework, proposed by Kane (1992, 2004, 2006). To evaluate test score interpretations and uses based on Kane's framework, test developers first need to provide interpretive arguments and then validity arguments by proving sound warrants for the following four inferences: (a) scoring from observation to an observed score, (b) generalization from the observed score to the universe score, (c) extrapolation from the universe score to a target score, and (d) decision from the target score to use.

In the field of language testing, a number of studies have been conducted to investigate the validity of test score interpretations and uses, especially for the ones considered to be high-stakes such as the TOEFL (Chapelle, 2008; Chapelle, Enright, & Jamieson, 2010). However, not many studies have been conducted to validate in-house placement test score interpretations and uses, and no study has evaluated the validity of such low-stakes tests using Kane's validity framework. Regardless of

whether the tests are high or low stakes, test developers need to be responsible for validating their test score interpretations and uses in order to attest to the validity.

This study uses Kane's (2006) argument-based validity framework to evaluate the validity of in-house placement test score interpretations and uses. The research questions are as follows: (a) to what extent do examinees get placement items correct and high-scoring examinees get more placement items correct; (b) to what extent are placement items consistently sampled from a domain sufficient in number so as to reduce measurement error; (c) to what extent do the difficulty of placement items match the objectives of a reading course; and (d) to what extent do placement decisions made to place examinees in their proper level of the course have an impact on washback in the course?

An in-house placement test made up of 40-item grammar, 40-item vocabulary, and 10-item reading sections was developed and administered to 428 first-year private-university students in April 2010. The item format adopted was all multiple-choice so the answer sheets could be easily scored with a reader. Based on their test scores, about 60 high-scoring students and 50 low-scoring students were placed into one of two advanced or one of two basic reading classes. The remaining students were placed into one of several intermediate classes. A 55-item grammar achievement test was administered twice (once as a pretest and then again as a posttest) to the two basic and two intermediate classes. In addition, a 51-item class evaluation survey was administered to investigate students' participation in the reading classes and to gauge students' satisfaction with the classes and study support.

Warrants for a validity argument of score inference were based on the results of the item analysis. A warrant for a validity argument of generalization inference was based on the composite generalizability coefficient of .92. A warrant for a validity argument of extrapolation inference was based on FACETS analysis, showing that difficulty estimates of learning levels were in an expected difficulty order. A warrant for validity arguments of decision inference was based on the basic-level students' score gain on an achievement test and their positive reactions to a class evaluation survey. All the validity arguments presented in this study support the validity of the placement test score interpretations and uses. However, to further improve the validity of the test score interpretations and uses, it is necessary to investigate washback effects of the placement test in the reading classes and to revise the test to make grammar, vocabulary, and reading sections with 30 items each.

**本** 研究では論証型妥当性枠組み(Kane, 2006)を用いて大学一年生必修リーディング科目用のプレースメントテスト得点の解釈と使用についての妥当性を評価した。その結果、項目分析の結果は良好であった(得点化の妥当性論証の論拠)。測定領域より項目は抽出されており、多変量一般化可能性理論を用いて合成一般化可能性係数を求めた結果、.92と高い結果が得られた(一般化の妥当性論証の論拠)。FACETS分析を用いて学習レベル困難度推定値を求めた結果、難易度推定値は予想通りであった(外挿の妥当性論証の論拠)。基礎クラスと中級クラ

ス履修者を対象に、到達度テストを事前・事後テストとして実施し、その平均値差を検証した結果、基礎クラス履修者は得点を上昇させることができた。また、授業評価アンケート結果によると、学習支援など指導が効果的だったということがわかった(決定の妥当性論証の論拠)。

Validation is simple in principle, but difficult in practice. The argument-based framework provides a relatively pragmatic approach to validation. (Kane, 2012, p. 15)

日本の高等教育では集団基準準拠テスト(norm-referenced tests)の一種である一般入試やAO入試を含む入学試験制度の多様化などの要因から、学生の英語力の差が学部内でも顕著になり、学習者の英語力により適した指導ができるよう習熟度別カリキュラムが重要視されるようになった。それに伴い、クラス分けに用いられるプレースメントテスト(placement tests)の関心が高まり、その研究(Culligan & Gorsuch, 1999; Westrick, 2005等)が行われてきた。また、よりカリキュラムの内容を反映した目標規準準拠テスト(criterion-referenced tests)の特性を持ち合わせたテスト開発についての研究も行われてきた(Brown, 1989; 熊澤, 2010)。

妥当性の評価方法は様々であり、妥当性理論も進化してきた。市販のテストや学内開発プレースメントテストが多用される中、実際にそれらのテスト得点の解釈(interpretations)と使用(uses)がどの程度妥当であったかを評価する研究はさほど行われてこなかった。しかし、実施されたテスト得点の解釈と使用が妥当であったということを検証することは開発者の義務ではないかと考える。また、妥当性を論証することにより改善点が明白になり、さらに妥当性が高いものに改訂することができる。よって、本研究では妥当性理論の変遷について述べ、Kane(1992, 2004, 2006)が考案した論証型妥当性枠組み(argument-based validity framework)をもとに、ある大学で一般教養一年生必修リーディング科目のクラス分けを目的に開発されたプレースメントテスト得点の解釈と使用の妥当性を評価する。

## 妥当性について

妥当性は「テスト得点の解釈と使用がいかに論拠と理論によって支持されるか」と定義されている(American Educational Research Association/American Psychological Association/National Council on Measurement in Education, 1999, p. 9)。心理測定分野では妥当性について書かれた文献は多々ある(e.g., Lissitz, 2009; Wainer & Braun, 1988)。その中でも妥当性という概念がいかに進化してきたかを詳述しているのはKane(2001)である。それによると妥当性理論の変遷は(a)基準的妥当性(criterion-based validity)と内容的妥当性(content-based validity)、(b)構成概念妥当性(construct validity)、(c)妥当性の現代の見解(current view of validity)と三つの時期を経ている。第一の時期では、基準的妥当性を検証するにはあるテストと外的基準となるテストとの相関関係を検証する。例えば、TOEFLとあるテストとの相関係数が高いことにより妥当性が高いとする。しかし、外的基準となるテストが無く基準的妥当性を検証することが困難な場合、テスト項目と測りたい領域との一致度を専門家が判断することで内容的妥当性が検証されていた。

第二の時期では、Cronbach and Meehl(1955)があるテストが何の構成概念をどの程度測っているかを検証する構成概念妥当性という概念を提唱した。その後、複数の構成概念とテスト形式の相関関係を算出することで妥当性を検証する多特性多方法(multitrait-multimethod; Campbell & Fiske, 1959)など構成概念妥当性を検証する方法が考案された。その後、Cronbach(1971)はテスト自体が妥当であるかではなくテスト得点の解釈と使用が妥当であること、テスト得点の解釈と使用を妥当化するには複数の証拠(evidence)が必要であることを説いた。また、Cronbach(1980)とMessick(1981)は妥当性の評価方法と概念を確立すべきだと主張した。

第三の時期では、Messick(1989)が単一的構成概念妥当性(unitary construct validity)を発表し、証拠基準(evidential basis)と影響基準(consequential basis)を含むテストの正当化(test justification)、および解釈と使用を含むテストの機能(test function)の二つの関連した相(facet)を妥当性の枠組みとした。また、構成概念妥当性には内容的側面(content aspect)、本質的側面(substantive aspect)、構造的側面(structural aspect)、一般化的側面(generalizability aspect)、外的側面(external aspect)、影響的側面(consequential aspect)の六つの側面があるとした(Messick, 1995)。単一的構成概念妥当性は特に信頼性の概念とテストの影響(consequence)について妥当性理論に大きな影響を与えた。以前は、信頼性は妥当性の必要条件として独立したかたちで捉えられていたが、信頼性は一般化という概念に代わり妥当性の一部として捉えられるようになった。テスト得点は受験者に影響を与え、利害が大きい(high-stakes)テストほど影響も大きくなる。Messick(1989)はこのような影響も妥当性の枠組みに入れることを主張した。

Kane(1992, 2004, 2006)は妥当化の実践的でなおかつ系統的な方法を示すため論証型妥当性枠組み(argument-based validity framework)を考案した。Kane(2006)によると妥当性を評価するためには解釈的論証(interpretive argument)と妥当性論証(validity argument)を提示する必要がある。前者はテスト結果の解釈と使用に内在する連鎖的な推論(inference)に対する前提条件を明示することであり、後者は複数の分析結果を用いてその解釈的論証を評価するための論拠(warrant)を提示することである。図1にあるように、観測(observation)から観測得点(observed score)、観測得点から測定領域得点(universe score)、測定領域得点から目標得点(target score)、目標得点から用途(use)をそれぞれ結ぶ推論は得点化(scoring)、一般化(generalization)、外挿(extrapolation)、決定(decision)と四つがある。ブレイスメントテストの解釈的論証の場合、得点化についての推論は、ある項目がいかに採点されるかを述べ、それが適切であることを前提とする。一般化についての推論は、測定領域を代表する項目が抽出されており(representative of the universe)、また誤差を最小に抑えられるよう項目数は十分であることを前提とする。外挿についての推論は、項目は授業目標の到達度を示すのに適切であり、誤差はテスト方法などの要因の影響をさほど受けていないことを前提とする。決定についての推論は、低得点者に適切な難易度の授業目標を学習させることで、より適切な成績評価ができるなど利益があることを前提とする。

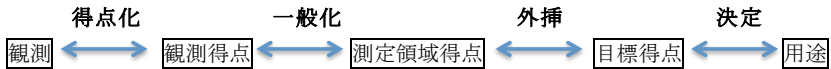


図1. 解釈的論証における推論の連鎖について  
(Kane, Crooks, & Cohen, 1999, p. 9参考)

プレイズメントテストをもとに、いかに解釈的論証を評価するための妥当性論証の論拠を提示するかを一例として挙げる(Kane, 2006)。得点化の妥当性論証の論拠としては古典的テスト理論であれば項目容易度(item facility)と項目弁別力(item discrimination)、項目応答理論であれば項目困難度(item difficulty)と適合度(fit)を含む項目分析をもとにすべての項目は正確に得点されていることが論拠となる。一般化の妥当性論証の論拠としては内部一貫性信頼性クロンバック $\alpha$ 係数を用いて論拠とすることができる。または、一般化可能性理論(Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Brennan, 2001a; Shavelson & Webb, 1991)を用い、一般化可能性研究(generalizability study [G研究])で各相の分散成分を求め、それをもとに決定研究(decision study [D研究])で集団基準準拠テストであれば一般化可能性係数(generalizability coefficient)を、目標標準準拠テストであれば信頼度指数(dependability index)を論拠とする。外挿の妥当性論証論拠を提示するには分析型と実証型の方法がある。分析型の方法では、測定領域と目標領域(target domain)との関係を調べるため、どのように問題を解いたかを発話させ、それを分析すること(think-aloud protocol analysis)で論拠を得る。実証型の方法では従来の基準的妥当性や因子分析の結果を論拠に用いることができる。決定の妥当性論証にはテスト得点をもとに下した決定の影響や波及効果(Cheng & Watanabe, 2004)について述べ、下した決定が適切であったことを示すことを論拠とする。

言語テストの分野では妥当性について記述した文献は多々ある(Cumming & Berwick, 1996; Kunnan, 1998; Weir, 2005等)。Bachman and Palmer(1996)は信頼性(reliability)、構成概念妥当性(construct validity)、真正性(authenticity)、相互性(interactiveness)、影響(impact)、実用性(practicality)というテストの質からなるテストの有用性(test usefulness)という概念を発表した。Bachman(2005)はKaneの論証型妥当性枠組みとBachman and Palmer(1996)のテストの有用性を踏襲し、評価妥当性論証(assessment validity argument)と評価用途論証(assessment utilization argument)からなる評価使用論証(assessment use argument)という概念を発表した。評価妥当性論証とはKane(2006)でいう解釈に関する論証であり、得点化から外挿までの推論と同意である。評価用途論証はKane(2006)では決定の推論にあたり、Bachmanはさらに決定の推論を細分化し、関連性(relevance)、用途(utility)、意図した影響(intended consequences)、充足(sufficiency)とした。また、論拠に対する反論(rebuttal)を挙げることも重要視した。関連性とは得点解釈が得点をもとに下す決定と関連性があるかをいう。用途とは得点解釈が意図した決定を下すのに役に立つかをいう。意図した影響とは得点解釈をもとに下した決定が受験者、教育機関、会社など社会全体に有益な影響を及ぼすかをいう。充足とは行われた評価が決定を下すのに十分な情報を提供するかをいう。よって、Bachman(2005)の枠組みは得点化から充足まで七つの推論(得点化、一般化、外挿、関連性、用途、意図した影響、充足)にもとづいている。



Bachman and Palmer (2010) はさらに新たな論証型妥当性枠組みを考案した。まず、タスクをもとにパフォーマンスが観測される。そして、評価記録 (assessment records) は一貫性があるかの論拠と反論を検証する。次に、解釈は意義があるか (meaningful)、公平であるか (impartial)、一般化できるか (generalizable)、関連性があるか (relevant)、充足するものか (sufficient) の論拠と反論を検証する。次に、決定は価値があるか (values sensitive)、公平か (equitable) の論拠と反論を検証する。最後に影響は有益なものか (beneficial) の論拠と反論を検証する。よって、この枠組みは一貫性から有益性まで九つの点 (評価記録に一貫性があるか、解釈に意義・公平性・一般化・関連性・充足があるか、決定に価値・公平性があるか、影響が有益か) についての論拠と反論を検証することになる。

Chapelle (2008) は主に Kane (2006) の枠組みを踏襲し、測定領域記述 (domain description)、評価、一般化、説明 (explanation)、外挿、用途 (utilization) の推論について論証すべきとしている。測定領域記述とは測っている概念を記述することである。評価と用途は Kane (2006) では得点化と決定と同意である。説明とは分析結果をもとに測っている概念はなにかを明示することである。よって、Kane (2006) の枠組みに測定領域記述と説明を加え六つの推論 (測定領域記述、評価、一般化、説明、外挿、用途) にもとづくのが Chapelle (2008) の枠組みである。

表1. 妥当性枠組みの比較について

	得点解釈に関する推論	得点使用に関する推論
Kane (2006)	得点化、一般化、外挿	決定
Bachman (2005)	得点化、一般化、外挿	関連性、用途、意図した影響、充足
Bachman & Palmer (2010)	評価記録は一貫性があるか、解釈には (意義があるか、公平であるか、一般化できるか、関連性があるか、充足するものか)	決定には (価値があるか、公平であるか)、影響は (有益か)
Chapelle (2008)	測定領域記述、評価、一般化、説明、外挿	用途

言語テストの分野ではテスト得点解釈と使用の妥当性を評価した実証研究がある。まず、Pardo-Ballester (2010) は Bachman (2005) の枠組みを用いてウェブ上で実施するスペイン語リスニングテストの有用性 (usefulness) の特性である信頼性、構成概念妥当性、真正性の論拠を挙げている。信頼性にはラッシュ信頼性係数 (Rasch reliability coefficient)、項目適合度、受験者適合度、主成分分析 (principal component analysis) の結果を提示している。構成概念妥当性にはタスクの難易度を三レベルに区分けし、項目応答理論にもとづくそれぞれのレベルの項目困難度平均値差の結果を挙げている。真正性はテスト項目が授業で行った問題と類比しているかなどを調査するアンケートの結果を提示した。



Chapelle, Chung, Hegelheimer, Pendar, and Xu (2010)ではコンピュータ上で受験できる文法テストを開発し、Chapelle (2008)の妥当性評価の枠組みである六つの解釈的論証の推論のうち測定領域記述、評価、一般化、説明、外挿の五つの妥当性検証の論拠を挙げている。領域記述に対しての妥当性検証の論拠にはその文法テストの内容、実施目的、開発過程などの記述を用いている。Chapelle (2008)がいう評価というのはKane (2006)がいう得点化にあたるが、この論拠として採点方法の記述を用いている。一般化の論拠ではクロンバック  $\alpha$  係数が、説明の論拠には第二言語習得理論にもとづく文法習得過程と項目困難度との一致が、そして外挿の論拠には他の文法テストとの高い相関関係および習熟度別の三グループの平均値に有意な差があったことが挙げられている。

Beglar (2010)はMessick (1995)の単一的構成概念妥当性の枠組みを用いてラッシュモデル(Rasch model)から得られた結果をもとに語彙サイズテスト(Vocabulary Size Test [VST])得点解釈の妥当性を評価している。内容的側面には項目困難度(item difficulty)と項目適合度を用いている。本質的側面としては語彙の頻度レベル(frequency level)とそれぞれのレベルの項目困難度が一致しているかと能力別に語彙数を調査した結果を提示している。構造的側面としてはVST項目の一次元性(unidimensionality)についての結果を示した。一般的側面としてはラッシュ信頼性係数を提示した。

Koizumi et al. (2011)は日本人学習者対象に文法診断テストを開発し、項目弁別力、錯乱肢分析(distracter analysis)、信頼性、名詞節の難易度の結果を妥当性論証の論拠として挙げた。項目弁別力は比較的高い、選択肢は機能している、信頼性は高い、名詞節の難易度はほぼ予想通りであったことからそのテスト得点解釈の妥当性を評価している。

## 研究課題

妥当性理論は進化を遂げ、また言語テストの分野でもTOEFLなどのテスト得点解釈と使用の妥当性の評価がされてきた(Chapelle, 2008; Chapelle, Enright, et al. 2010)。しかし、教員がカリキュラムの内容を反映するプレースメントテストを開発し、その妥当性を評価した研究はさほど行われてこなかった。ましてや、現代の妥当性理論であるKane (2006)の妥当性枠組みを用いてプレースメントテスト得点解釈と使用の妥当性を評価した論文はほぼ皆無であろう。

本研究ではKane (2006)の枠組みを用いるが、その理由は明確さと簡潔さである。まず、得点化から決定までの推論である一連の関連性が非常に明確であることが挙げられる。つまり、解答が得点化され観測得点となり、その観測得点が一般化され測定領域得点となり、その測定領域得点が外挿され目標得点となり、その目標得点をもとに決定が下され使用されるという一連の推論の関係性が明確であるということである。次に、論証型妥当性枠組みは簡潔であるべきであり、Kaneのものがもっとも簡潔であるといえる。枠組みが複雑になることで、論拠をもとに妥当性検証をすることが教員にとってより困難になるのではないか。妥当化を行うことは容易いことではないので、枠組みが明確で簡潔なほうが教員にとって取り組みやすいということである。

本研究の目的は大学一年生対象一般教養必修リーディング科目のクラス分けを目的に開発されたプレースメントテスト得点解釈と使用の妥当性をKane (2006)の論証

型妥当性枠組みを用いて評価することである。つまり、得点化、一般化、外挿、決定の推論に付随する前提条件 (assumption) を挙げることにより解釈的論証を行い、それに対する論拠をもとに妥当性論証をすることである。以下にある四つの推論に対する解釈的論証の前提条件を研究課題とする。

1. プレイメントテスト項目はどの程度意図した正解を受験者が解答でき、高得点者がより正解できるか。
2. プレイメントテスト項目はどの程度一貫して測定領域から抽出されていて、また項目数は誤差が最小になる程度あるか。
3. プレイメントテストにおける学習レベルはどの程度リーディング科目の授業目標と難易度が一致しているか。基礎レベルの学習目標は文法であれば中等教育で学習した事項、語彙であれば2000語程度、読解であれば基礎的な英文理解ができる程度とする。
4. プレイメントテスト得点をクラス分けの判断材料として使用し、習熟度別クラス編成し、受講者のレベルにより適した授業内容を提供することにより、どの程度正の波及効果が生じるか。

## 方法

### 対象者

本研究の対象者は関東地方にある私立大学法学部に在籍するプレイメントテストを受験した428名の2010年度入学一年生である。外国人学習者一名を除いてはすべて日本人学習者であった。法学部開講必修英語は一年次には四科目あり、そのうちの二科目はリーディング、それ以外はコミュニケーション中心の授業であった。前者の科目のみ習熟度別に編成され、プレイメントテスト高得点者の60名程度が上級クラスである二クラス、低得点者の50名程度が基礎クラスである二クラスへ得点順に振り分けられた。それ以外の受験者は中級クラスである十クラスのいずれかに自動的に振り分けられた。上級クラスでは速読と語彙学習が中心で、基礎クラスではリメディアル教育の一環として中等教育で学習した文法事項を復習するのが主な授業内容であった。基礎クラスの学習者はさらなる教育支援が必要と思われるので授業外でも教員から指導を受けられる体制が整っていた。

### プレイメントテスト

プレイメントテストを開発した理由は、一年生必修リーディング科目を習熟度別にクラス編成するためである。学部で独自にテスト開発をすることにより、よりカリキュラムを反映する内容になると考えた。また、履修生に迅速に指定クラスを掲示するためマークシートリーダーで採点できるテストが必要であった。リーディング科目のプレイメントテストということで、最終的には文法 ( $k = 40$ )、語彙 ( $k = 40$ )、読解 ( $k = 10$ ) の三つのセクションから成るテストとした。

文法項目を作成するにあたり、リーディング科目では中等教育で学習したことをもとにさらに読解力を伸ばすということが目標であることを考慮し、まず中学校・高等学校の教科書(霜崎他, 2006, 高橋他, 2005)にあるすべての重要文法項目を調べ、リス

トを作成した。中学校および高等学校の学習文法事項を参照するためにその教科書を用いた理由は一般的に広く用いられていると思われるため、また教科書によって扱う学習事項が若干異なるので統一したほうがいいと判断したためである。その後、重要文法事項を用いて英文を作成し、次に文法事項の個所を空欄にし、そして正解を含む選択肢を四つ作成した。よって、項目形式は多肢選択式項目である。2006年に実施した試行テストでは70項目を20名の受講者が受験し、項目容易度の値が極端に高いまたは低い項目で、項目弁別力の値が低い30項目を除外した。特に、中学校一年生で学習する文法事項は受験者にとって易しすぎたので削除した。残った40項目は143名の受講者対象に試行テストとして再度実施され、その結果をもとに改良したものが本試験で用いられた。本試験の前に再度四人の教員が項目の文言などをみて修正を加えた。2007年以降毎年ほぼ同じ項目を用いこのプレイズメントテストを実施しているが、2010年度の実施では文言と選択肢を何点か変更した。例題は以下の通りである。

Hi, I ( ) Ken.

1. am    2. are    3. is    4. be

受験者は英文を読んで空欄個所にあてはまるもっとも文法的に正しいものを選ぶ形式であり、正解は1である。配点は各設問につき一点である。

語彙セクションの開発過程では、まず大学英語教育学会基本語リストJACET List of 8000 Basic Words(大学英語教育学会基本語改訂委員会, 2003)より1,000語から2,000語の頻度にある48語を選んだ。この語彙リストを選んだ理由としては日本人学習者に即しているという点である。まず、受験者の語彙レベルを把握するため20名の受講者にテストを実施した。その結果、得点結果にはばらつきがさほどみられなかったため、本試験では1,000から3,000語の頻度範囲から40語を選んで40項目を作成した。そして、その40項目は試行試験として実施され、その結果をもとに改良を加えたものが本試験で用いられた。また、文法セクション同様2007年以降、複数の文言と選択肢に変更を加えた。例題は以下の通りである。

Bring

1. 送る    2. 持ってくる    3. 鳴る    4. 購入する

受験者が英単語を見てもっとも意味が近い和訳を選ぶという形式であり、正解は2である。配点は各設問につき一点である。

読解セクションは2008年より追加された。その主な理由はやはりリーディング科目のプレイズメントテストであるにも関わらず、読解項目がなく語彙と文法のみだったためである。本文は基礎クラスで使用されている教科書にある123語から成るものと、上級クラスで用いられている教科書にある341語から成るものを選んだ。低得点者は基礎クラスで用いている教科書にある本文の理解度が低く、高得点者はほぼ理解できると想定した。それぞれの本文に対して五項目ずつ英文で書かれた多肢選択式項目を作成した。項目数を増やしたかったのだが文法と語彙項目数だけですでに80項目あり、試験時間も45分であるので断念した。文法と語彙項目を減らし読解項目を増やすことも検討したが、項目を減らすことによって、前年度の結果と比較できなくなるとの理由で行わなかった。四名の教員と一名の英語母語話者教員が英語を確認し、不自然な文言などを修正した。試行テストは行われなかった。例題は以下の通りである。

How do they travel?

1. by plane 2. by bus 3. by car 4. by train

受験者は本文と設問を読み、その解答を選択肢から選ぶという項目形式である。配点は各設問につき二点である。その理由は満点を100点にするため、またリーディング科目のプレイズメントテストとして用いるのでリーディングの配点を大きくしたいためである。

### 実施の手順

教員と職員とでテスト問題やマークシートを用意し、教員が試験を監督した。試験監督者は監督の手引きを試験前に受験者に読み上げ、このプレイズメントテストの結果はクラス分けと研究目的で使用されるとの説明がなされた。試験時間が45分に保たれるように監督者は留意した。マークシートはリーダーによりデータ化され、エクセルで採点がされた。未解答は不正解とみなした。テスト結果は受験者へは返却されなかった。

上級クラス、中級クラス、基礎クラスともに通年でおよそ30回の授業が行われた。上級クラスでは速読と語彙学習を中心に授業が行われた。複数ある中級クラスの内、二クラスのみ、プレイズメントテストとは異なる55項目から成る文法到達度テストを2010年四月事前テストとして、2011年一月事後テストとして実施した。このテストは基礎クラスでの学習内容の到達度を測る目的とプレイズメント目的で開発された(熊澤, 2010)。また、最終授業内で51項目から成る授業評価アンケートを実施した。六件法((6)つよくそう思う、(5)そう思う、(4)まあそう思う、(3)あまりそうは思わない、(2)そうは思わない、(1)まったくそうは思わない)を採用し、授業の満足度、履修者の参加度合い、学習支援内容などを評価した。例えば、項目1は「この授業でよくできたのは学習支援室でのサポートがあったからだ」である。中級クラス授業内では文法学習や訳読などが行われた。

基礎クラスでは中級レベルの二クラスと同様に文法到達度テストが実施された。しかし、基礎クラスのみ文法到達度テストの結果が成績に考慮された。また、同クラスでは授業評価アンケートも最終授業内で実施された。基礎クラス授業内では基礎的な文法を学習すること、英文を訳すこと、学習内容の習熟をみるため小テストなどが行われた。授業外では文法練習帳を課題として出題し、わからない点や答え合わせなど教員から指導を受ける支援があった。

### 分析の手順

得点化の妥当性論証の論拠を提示するため、古典的テスト理論および項目応答理論(item response theory)を用いた。古典的テスト理論にもとづく分析では項目容易度および項目弁別力の値を求めた。項目応答理論を用いた分析は以下で述べる。一般化の妥当性論証の論拠を提示するため、多変量一般化可能性理論(multivariate generalizability theory)を用いて文法、語彙、読解セクションそれぞれの受験者( $p$ )、項目( $i$ )、受験者 $\times$ 項目( $p \times i$ )の分散成分推定値を一般化可能性研究(generalizability study [G研究])で求め、その結果をもとに各セクションの一般化可能性係数(generalizability coefficient)と合成一般化可能性係数(composite

generalizability coefficient)を決定研究(decision study [D研究])で求めた。よって、このG研究のデザインは $p \cdot X^p$ でD研究は $p \cdot X^D$ である。 $\bullet$ は全セクションともに受験者は同じであることを意味し、 $\circ$ は各セクションに設けられた項目は異なるということの意味する。多変量一般化可能性理論を用いた理由はこのテストは三セクションから成るので、各セクションの項目数を変更することにより、いかに一般化可能性係数が変動するかを検証したいがためである。mGENOVA (Brennan, 2001b)をG研究とD研究に用いた。多変量一般化可能性理論についてはBrennan(2001a)が解説書を書いている。また、言語テストの分野でも多変量一般化可能性理論を用いた論文がある( Lee, 2006; Sawaki, 2007; Xi, 2007等)。

得点化および外挿の妥当性論証の論拠を提示するため、項目応答理論の一種である多相ラッシュモデル(multifaceted Rasch model; Linacre, 1989)をもとにした統計ソフトであるFACETS (Linacre, 2002)を用いた。インフィット平均二乗に用いる基準は0.80~1.20 (Bond & Fox, 2001)とし、この項目の値が基準を満たしている場合、一次元性が保たれたと仮定することができる(Bond & Fox, 2001)。分析対象とした相は受験者能力、項目困難度、学習内容困難度で、それぞれの推定値を求めた。得点化の論拠には項目困難度推定値とその標準誤差を用いた。外挿の論拠としては一次元性と学習レベル困難度を用いた。学習レベル困難度推定値は高校文法レベルより中学文法レベル、JACET3000語彙レベルよりJACET1000語彙レベル、上級読解レベルより基礎読解レベルのほうがより低くなると仮定された。ラッシュモデルについてはBond and Fox (2001)が解説書を書いており、言語テストの分野でも多相ラッシュモデルを用いた研究は多々ある(e.g., Coniam, 2008; 熊澤, 2010; Schaefer, 2008)。

そして、決定の論拠を求めるため、プレイスメントテスト、文法到達度テスト、および授業評価アンケートの記述統計を求めた。まず、習熟度別にクラスが編成されているかを確認するため全体、上級クラス1、上級クラス2、中級クラス、基礎クラス1、基礎クラス2ごとのプレイスメント得点平均値と標準偏差値を求めた。この記述統計では読解項目のみ配点を各項目二点とした。次に、中級群と基礎群ではどの程度教育効果があったかを検証するため事前テスト時と事後テスト時の文法到達度テスト得点平均値と標準偏差値を求めた。最後に、中級群と基礎群がどのように授業を評価したかを調査するために実施されたアンケートのそれぞれの平均値と標準偏差値を求めた。

## 結果

表2は古典的テスト理論とFACETS分析による項目分析の結果についてである。項目容易度(IF)の値には大きなばらつきがみられる。例えば、項目41の値は.99で項目28の値は.15となっている。プレイスメントテストでは、値が.30から.70の範囲にある項目がよいとされる(Brown, 2005)。その基準をもとに項目の良否を判断すると表中にある太字で表記されている28項目は値が適切ではなく、受験者にとって易しい、または難し過ぎるということになる。しかし、高得点者が低得点者より正解できるかを示す項目弁別力をみると、不良項目と判断される基準である.19以下の項目は表中にある太字で表記されている四項目のみである。

項目困難度推定値のばらつきは大幅にあり、分離指数(separation index)は6.87でラッシュ信頼性係数は.98であった。これは項目困難度推定値の最大値と最小値がそ



れぞれ2.33と-3.79で、項目困難度推定値のばらつきが広範囲になっていることを意味する。FACETS分析の利点は図2にあるように項目困難度と受験者能力値などの推定値が同一の間隔尺度上に示され、視覚的にその推定値のバランスを検証できることである。大半の受験者能力推定値は-1.00から1.00の範囲にあり(74%)、またその範囲を測る項目も多くある(82%)。しかし、能力推定値が1.00以上となった受験者は若干いる(20%)のものにもかかわらず1.00以上の能力推定値を測る項目がさほどない(7%)。また、能力推定値が-2.00以上になった受験者はいないが、項目41と項目44は大幅にその推定値を下回っているのを削除してもいいかもしれない。標準誤差は最大値が項目41の.50で、項目が易しく99%の受験者が正解したため、この項目の推定値を算出するのにデータが足りず、誤差が大きくなったことを示す。他の値は問題なく推定値の誤差はさほどない。インフィット平均二乗は基準である0.80~1.20(Bond & Fox, 2001)に項目67以外のすべての項目の値が収まりモデルに適合しているといえる。項目67の出題単語はacquireで選択肢は1から4の順に、「合唱する、必要とする、取得する、支払う」で正解は3であるが、高得点者が不注意でrequireと間違い2を選択してしまったなどの理由が挙げられる。しかし、アウトフィット平均二乗ははずれ値に敏感なため0.80~1.20の基準の範囲以外の項目が14項目もある。特に問題視すべきなのは項目50で、項目弁別力が-.05、アウトフィット平均二乗の値が2.20になった。これは、高得点者が誤答してしまう確率があり、選択肢などを改訂すべきであることを示唆している。この項目内容をみると英単語がruleで、選択肢が「役割、規定、命令、事実」となっており、正解は規定である。高得点者も不注意でroleと間違えてしまい、選択肢の役割を選んでしまったことが推測される。一方、受験者については、インフィット平均二乗値が0.80~1.20の基準からはずれ、不適合(misfit)と判定された者は16名(4%)であった。その内14名は1.20以上になり適合不足(underfit)と判定され、集中力が持続しなかったなどの理由が挙げられる。

表2. 古典的テスト理論とFACETS分析による項目分析の結果について (N = 428)

項目	項目内容	古典的テスト理論		FACETS分析			
		IF	ID	Diff	SE	Infit MS	Outfit MS
1	過去形不規則変化(中)	.81	.37	-0.74	0.13	1.00	0.90
2	過去進行形(中)	.52	.25	0.79	0.10	1.20	<b>1.30</b>
3	接続詞when(中)	.81	.45	-0.74	0.13	0.90	0.80
4	～あるthere(中)	.78	.45	-0.54	0.12	0.90	0.90
5	接続詞and(中)	.80	.42	-0.71	0.13	0.90	<b>0.70</b>
6	命令形Will(中)	.58	.47	0.51	0.11	0.90	0.90
7	比較(中)	.65	.43	0.20	0.11	1.00	1.00
8	受身(中)	.76	.39	-0.43	0.12	1.00	0.90
9	have to=must(中)	.51	.25	0.87	0.10	1.10	1.20
10	仮定法(中)	.67	.49	0.06	0.11	0.90	0.90

項目	項目内容	古典的テスト理論		FACETS分析			
		IF	ID	Diff	SE	Infit MS	Outfit MS
11	疑問詞How long (中)	.82	.45	-0.83	0.13	0.90	0.80
12	現在完了形・経験 (中)	.73	.37	-0.24	0.12	1.00	0.90
13	形式主語 (中)	.67	.50	0.09	0.11	0.90	0.80
14	so that構文 (中)	.59	.29	0.51	0.11	1.10	1.10
15	動名詞形容詞の用法 (中)	.60	.40	0.43	0.11	1.00	1.00
16	過去分詞形容詞の用法 (中)	.23	.21	2.33	0.12	1.10	<b>1.40</b>
17	関係代名詞Who (中)	.67	.53	0.12	0.11	0.90	0.80
18	関係代名詞Which (中)	.69	.38	-0.01	0.11	1.00	1.00
19	how to (中)	.81	.47	-0.69	0.13	0.90	0.80
20	because of (中)	.52	.31	0.83	0.10	1.10	1.00
21	不定詞の名詞的用法 (高)	.64	.52	-0.68	0.11	0.90	0.90
22	現在完了形 (高)	.39	.36	0.52	0.11	1.00	1.10
23	関係代名詞what (高)	.56	.47	-0.29	0.11	1.00	0.90
24	分詞の形容詞的用法 (高)	.62	.49	-0.59	0.11	0.90	0.90
25	seemの使い方 (高)	.54	.54	-0.20	0.11	0.80	0.80
26	分詞 (高)	.73	.47	-1.15	0.11	0.90	0.80
27	受動態の完了形 (高)	.39	.30	0.53	0.11	1.10	1.10
28	分詞構文 (高)	.15	.34	2.03	0.15	0.90	1.00
29	時制の一致 (高)	.71	.36	-1.05	0.11	1.00	1.10
30	with+名詞+形容詞の使い方 (高)	.32	.38	0.88	0.11	1.00	1.00
31	関係副詞(非制限用法) (高)	.50	.42	-0.04	0.10	1.00	1.00
32	形式目的語 it (高)	.43	.21	0.29	0.11	1.20	1.20
33	強調構文 (高)	.41	.32	0.39	0.11	1.10	1.10
34	過去完了進行形 (高)	.51	.43	-0.05	0.10	1.00	1.00
35	未来完了形 (高)	.66	.41	-0.80	0.11	1.00	0.90
36	未来進行形 (高)	.19	.18	1.64	0.13	1.10	<b>1.30</b>
37	倒置 (高)	.29	.23	1.01	0.12	1.20	<b>1.30</b>
38	助動詞 would の用法 (高)	.75	.40	-1.29	0.12	1.00	1.00
39	不定詞の完了形 (高)	.55	.39	-0.28	0.11	1.00	1.00
40	動名詞の意味上の主語 (高)	.34	.26	0.78	0.11	1.10	1.20
41	Country (139)	.99	.14	-3.79	<b>0.50</b>	1.00	<b>0.50</b>

項目	項目内容	古典的テスト理論		FACETS分析			
		IF	ID	Diff	SE	Infit MS	Outfit MS
42	Problem (179)	.95	.22	-1.92	0.21	1.00	0.90
43	Bring (281)	.79	.42	-0.32	0.12	0.90	0.80
44	Company (294)	.97	.24	-2.75	0.31	1.00	<b>0.50</b>
45	Century (335)	.95	.31	-1.97	0.22	0.90	<b>0.60</b>
46	National (368)	.61	.35	0.66	0.11	1.10	1.10
47	Certain (428)	.54	.55	1.03	0.11	0.90	0.80
48	Consider (453)	.67	.40	0.38	0.11	1.00	0.90
49	Law (489)	.64	.39	0.55	0.11	1.00	1.00
50	Rule (495)	<b>.89</b>	<b>-.05</b>	-1.15	0.16	1.10	<b>2.20</b>
51	Population (558)	.59	.47	0.79	0.11	0.90	0.90
52	Various (647)	.74	.43	0.01	0.12	0.90	0.80
53	Development (786)	.78	.45	-0.24	0.12	0.90	0.80
54	Standard (825)	.93	.24	-1.67	0.19	1.00	1.00
55	Industry (873)	.69	.53	0.31	0.11	0.90	0.80
56	Reduce (878)	.66	.44	0.44	0.11	1.00	0.90
57	Represent (956)	.55	.36	0.99	0.11	1.10	1.10
58	Particular (966)	.66	.36	0.43	0.11	1.10	1.10
59	Tend (986)	.70	.48	0.25	0.11	0.90	0.90
60	Worth (994)	.52	.58	1.10	0.10	0.80	0.80
61	Solution (1186)	.50	.54	0.13	0.10	0.90	0.90
62	Available (1299)	.48	.59	0.24	0.11	0.80	0.80
63	Prison (1336)	.73	.28	-1.00	0.11	1.10	1.10
64	Environmental (1356)	.65	.50	-0.59	0.11	0.90	0.80
65	Victim (1486)	.47	.50	0.25	0.11	0.90	0.90
66	Status (1571)	.64	.20	-0.55	0.11	1.20	<b>1.40</b>
67	Acquire (1646)	.46	<b>.01</b>	0.34	0.11	<b>1.30</b>	<b>1.40</b>
68	Typical (1648)	.66	.42	-0.64	0.11	1.00	0.90
69	Poverty (1851)	.46	.36	0.33	0.11	1.10	1.10
70	Legal (1874)	.51	.39	0.09	0.10	1.00	1.00
71	Promote (1898)	.41	.48	0.54	0.11	0.90	0.90
72	Wage (1949)	.43	.51	0.46	0.11	0.90	0.90



項目	項目内容	古典的テスト理論		FACETS分析			
		IF	ID	Diff	SE	Infit MS	Outfit MS
73	Convince (1964)	.38	.31	0.73	0.11	1.10	1.10
74	Actual (1989)	.54	.48	-0.08	0.11	0.90	0.90
75	Interpret (2000)	.33	.28	0.96	0.11	1.10	1.20
76	Criminal (2003)	.50	.43	0.17	0.10	1.00	1.00
77	Discrimination (2411)	.50	.30	0.19	0.11	1.10	1.10
78	Superior (2622)	.47	.41	0.30	0.11	1.00	1.00
79	Punish (2859)	.51	.56	0.09	0.10	0.90	0.80
80	Equality (2974)	.62	.55	-0.40	0.11	0.90	0.80
81	基礎	.58	.54	-0.38	0.11	0.90	0.90
82	基礎	.55	.50	-0.24	0.11	0.90	0.90
83	基礎	.50	.56	-0.02	0.10	0.90	0.90
84	基礎	.52	.50	-0.08	0.10	1.00	1.10
85	基礎	.30	.45	0.99	0.11	1.00	1.00
86	上級	.49	.45	-0.36	0.11	1.10	1.10
87	上級	.52	.40	-0.50	0.10	1.10	1.10
88	上級	<b>.26</b>	.27	0.80	0.12	1.20	<b>1.40</b>
89	上級	<b>.25</b>	.34	0.86	0.12	1.10	<b>1.30</b>
90	上級	<b>.26</b>	.38	0.81	0.12	1.10	<b>1.30</b>

注.(中)=中学文法学習レベル、(高)=高校文法学習レベル、(139)=JACET8000による出現頻度、基礎=基礎読解クラスレベル、上級=上級読解クラスレベル、IF=項目容易度、ID=項目弁別力、Diff=項目困難度、SE=標準誤差、Infit MS=インフィット平均二乗、Outfit MS=アウトフィット平均二乗。

表3は $p \cdot X \cdot p$ デザインによるG研究の結果を示す。受験者の分散成分%はテスト得点の分散成分を100%とした場合、文法、語彙、読解、それぞれ11%、12%、11%であった。つまり、受験者の能力には10%程度のばらつきがあり、受験者の能力はテスト得点の分散の10%を占めることがわかった。受験者の文法と語彙、文法と読解、語彙と読解の共分散はそれぞれ.02、.02、.02で相関係数(disattenuated correlation)は.83、.79、.79となった。つまり、三つのセクションの得点は高い相関関係にある。項目の分散成分%は文法、語彙、読解、それぞれ13%、13%、7%で、項目の難易度には10%程度のばらつきがあったことになる。受験者X項目交互作用の分散成分%は76%、75%、82%で、受験者と項目以外にも様々な要因が得点のばらつきを生じさせていることを示す。

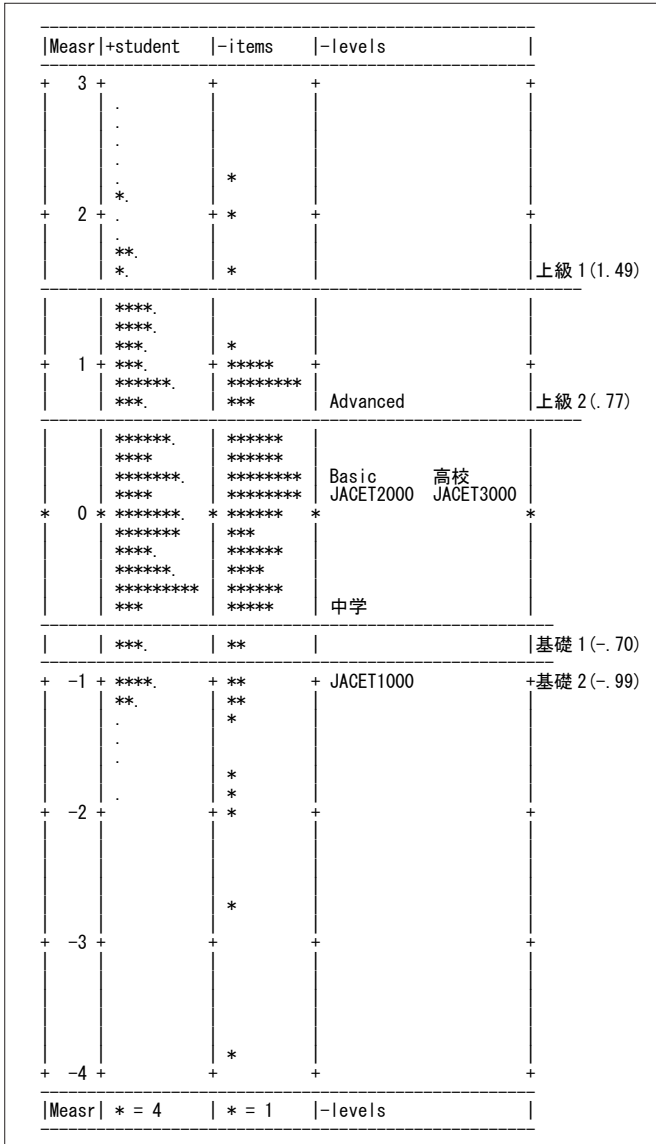


図2. グラフ化したFACETS分析の結果. 中学=中学文法レベル、高校=高校文法レベル、JACET1000=JACET1000語彙レベル、JACET2000=JACET2000語彙レベル、Basic=基礎読解クラスレベル、Advanced=上級読解クラスレベル。

表3. p・X i°デザインによるG研究の結果について (N=428)

相	文法		語彙		読解	
	分散成分	分散成分%	分散成分	分散成分%	分散成分	分散成分%
受験者 (p)	.02709	11%	.83205		.78810	
	.02288		.02791	12%	.79338	
	.02144		.02191		.02732	11%
項目 (i)	.03281	13%				
			.03030	13%		
受験者X項目 (p X i)					.01781	7%
	.18603	76%				
			.17753	75%		
分散成分合計	.24593		.23574		.19925	82%
				.24438		

注. 太字=分散成分、対角線上=相関係数、対角線下=共分散。

表4はp・X i°デザインによるD研究の結果を表している。文法、語彙、読解の項目数がそれぞれ40、40、10の場合、母得点(universe score [ $\sigma^2(\tau)$ ])は受験者分散成分推定値と同じで、.02709、.02791、.02732となり、これは古典的テスト理論では真値(true score)に対応する。プレイスメントテストは集団基準準拠テストの一種であるので、目標規準準拠テストに用いる信頼度指数ではなく、一般化可能性係数を提示すべきである。そのためにはまず相対誤差(relative error [ $\sigma^2(\delta)$ ])を求める必要があり、これは文法の場合、受験者X項目分散成分推定値である.18603を項目数である40で割ることで求めることができ、.00465(.18603/40)となる。語彙と読解の相対誤差はそれぞれ.00444、.01992である。そして、母得点を相対誤差に母得点を足したもので割ると一般化可能性係数を求めることができ、文法、語彙、読解それぞれの係数は.85(.02709/[.00465 + .02709])、.86、.58となる(表4中にあるD研究7参照)。読解のみ係数が十分ではなく、内部一貫性が欠如していることになる。合成一般化可能性係数(composite generalizability coefficient [ $p^2$ ])を求めるにはまず合成母得点(composite universe score [ $\sigma^2(\tau)$ ])と合成相対誤差(composite relative error [ $\sigma^2(\delta)$ ])を求める。その値はすべてのセクションの母得点の値と重み付け(weight)および相対誤差の値と重み付けで求められる。ここでは特定の重み付けを指定し算出する名義重み付け(nominal weight)ではなく統計的に求めた効果的重み付け(effective weight)を用いた(Brennan, 2001a)。項目数がセクションごと40、40、10の場合、効果的重み付けは.44、.44、.11であった。合成母得点は.02452([(44)(.44)(.0271)] + [(44)(.44)(.0280)] + [(44)(.11)(.0273)] + [2(.44)(.44)(.0229)] + [2(.44)(.44)(.0219)] + [2(.44)(.11)(.0214)])で、合成相対誤差は.00204であった。そして、合成母得点を合成母得点と合成相対誤差を足した値で割ると合成一般化可能性係数が求められ、.92(.02452/

[.02452+.00204]であった。合成一般化可能性係数は75項目の場合は.91(表4中にあるD研究3参照)で120項目の場合でも.94で(表4中にあるD研究19参照)さほど変化はなく、プレイスメントテストには十分高い値である。各セクションの一貫性を保つためには一般化可能性係数で.80は必要である。各セクション25項目だと.80には達せず、30項目であれば達することがわかった。やはり、各セクションの一般化可能性係数が十分に高く、そのうえ合成一般化可能性係数も高いほうが望ましい。

表4.  $p \cdot X I^{\circ}$ デザインによるD研究の結果について ( $N=428$ )

D研究	文法		語彙		読解		合計	
	$k$	$\rho^2$	$k$	$\rho^2$	$k$	$\rho^2$	$k$	$p^2$
1	35	.84	30	.83	10	.58	75	.91
2	30	.81	35	.85	10	.58	75	.91
3	25	.78	25	.80	25	.77	75	.91
4	35	.84	35	.85	10	.58	80	.91
5	40	.85	35	.85	10	.58	85	.92
6	35	.84	40	.86	10	.58	85	.92
7	40	.85	40	.86	10	.58	90	.92
8	30	.81	30	.83	30	.80	90	.92
9	45	.87	40	.86	10	.58	95	.93
10	40	.85	45	.88	10	.58	95	.93
11	40	.85	40	.86	15	.67	95	.93
12	45	.84	45	.83	10	.58	100	.91
13	40	.85	40	.86	20	.73	100	.93
14	40	.85	40	.86	25	.77	105	.93
15	50	.88	50	.89	10	.58	110	.94
16	40	.85	40	.86	30	.80	110	.93
17	40	.84	35	.85	35	.83	110	.93
18	40	.85	40	.86	35	.83	115	.94
19	40	.85	40	.86	40	.85	120	.94

注.  $k$ =項目数、 $\rho^2$ =一般化可能性係数、 $p^2$ =合成一般化可能性係数。

表5はFACETS分析による学習レベル困難度推定値の結果についてである。図2でも視覚化された結果がある。学習レベル困難度の分離指数は14.37でラッシュ信頼性係数は1.00なので推定値にはばらつきがある。受験者にとってもっとも容易だったのはJACET1000語彙レベルであり、順に中学校文法レベル、JACET3000語彙レベ

ル、JACET2000語彙レベル、高校文法レベル、基礎読解レベル、上級読解レベルであった。本文を読んで設問に解答するには、文法や語彙などより多くの知識を必要とするため読解項目の難易度が高かったと思われる。標準誤差の値は低く、推定値の誤差は少なかった。インフィット平均二乗は基準の範囲以内であるが、アウトフィット平均二乗は上級レベル読解の値が基準範囲以外となり不適合となった。これは、アウトフィット平均二乗は外れ値に敏感であるためだと思われる。

表5. FACETS分析による学習レベル困難度推定値の結果について (N=428)

学習レベル	Diff	SE	Infit MS	Outfit MS
中学文法レベル	-0.65	0.03	1.00	1.00
高校文法レベル	0.29	0.02	1.00	1.00
JACET1000語彙レベル	-0.94	0.03	1.00	0.90
JACET2000語彙レベル	0.15	0.03	1.00	1.00
JACET3000語彙レベル	0.12	0.05	1.00	1.00
基礎読解レベル	0.30	0.05	1.00	1.00
上級読解レベル	0.73	0.05	1.10	1.30

注. Diff=学習レベル困難度、SE=標準誤差、Infit MS=インフィット平均二乗、Outfit MS=アウトフィット平均二乗。

図2にあるグラフ化されたFACETS分析の結果をみると学習レベルと級ごとの習熟度合いが把握できる。上級1の分割点は受験者能力推定値の1.49であった。上級1の履修者には高校までの学習レベルは容易に正解できることがわかる。例えば、もっとも推定値が高かった上級読解レベル(.73)でも50%以上正解できる確率があり、また基礎読解レベルには75%以上正解できる確率がある。その逆に基礎2の履修者にとってJACET1000語彙レベルは50%程度の確率で正解を導きだすことはできるが、それ以外のレベルは正解できる確率が低く、上級読解レベルに正解する確率は0%に等しい。この結果から、基礎クラス履修者は中学文法学習レベルにも到達していないことがわかる。

表6はクラス別のプレイズメントテストの得点結果についてである。全体的に上級クラスと判定された受講者は平均値が高く、文法であれば中等教育の学習内容、語彙であればJACET3000語彙レベル、読解であれば基礎レベルの本文はほぼ理解できることがわかる。他方、基礎レベルと判定された受講者は、平均値が低く習熟していないということがわかる。この結果から、上級クラス履修者が基礎レベルの授業内容を学習することはすでに学習済みの内容なので不利益になるといえ、基礎クラス履修者には上級クラスの授業内容は難しすぎるので不利益になるといえる。また、上級1と上級2の平均値差は顕著であるが、基礎1と基礎2の平均値差はさほどない。今後、

この基礎二クラスをより分別する項目を入れたほうがいいかもしれない。しかし、級ごとに得点順になっており、また得点差も顕著であることから習熟度別に編成されているといえる。

表6. クラス別ブレイスメントテスト得点結果について

クラス	n	文法		語彙		読解		合計	
		M	SD	M	SD	M	SD	M	SD
上級1	36	33.53	3.38	35.94	2.44	15.22	2.72	84.69	4.17
上級2	33	30.33	2.39	34.09	1.84	12.42	2.59	76.85	1.37
中級	310	22.48	5.41	24.13	5.58	7.72	3.55	54.31	10.99
基礎1	25	13.04	2.68	16.76	3.35	4.32	2.55	34.12	0.76
基礎2	24	11.17	2.91	14.42	2.89	3.33	2.55	28.92	3.46
全体	428	22.82	7.13	24.91	7.19	8.27	4.35	56.00	16.25

表7は中級クラスの二クラスを中級群とし、学習支援およびリメディアル教育を受けた基礎クラスの二クラスを基礎群とした場合の文法事前テストと文法事後テストの結果についてである。まず、事前テストの結果をみると中級群と基礎群の平均値差は十点ほどある。事後テストの結果は中級群の場合、事前テストの平均値と比較すると、十点ほど値が下がっている。これはテスト結果が成績に考慮されないのでテスト受験に集中しなかったということが主な原因として挙げられる。標準偏差の値が大きく、真剣に受験した履修生とそうでない者がいたため得点のばらつきが生じた可能性がある。基礎群の事後テスト平均点は、事前テスト平均点よりも六点ほど上昇している。いずれにしても、中級群と基礎群の得点低下または得点上昇の理由にはさまざまな要因があるため学習効果のためと断定することはできない。最終授業内で六件法の授業評価アンケートを実施した結果、中級群と基礎群の平均値が1.00ほど差があった項目は学習支援についてのもので、基礎クラス履修者にとって学習支援を受けることは重要であると「まあそう思った」(M [基礎群]=4.11; M [中級群]=2.16)、役に立ったと「まあそう思った」(M [基礎群]=4.32; M [中級群]=2.63)という結果になった。しかし、中級群と基礎群の授業満足度を調査する「全体的にこの授業は満足できた」という項目における平均値差(M [基礎群]=4.63; M [中級群]=4.51)はさほど顕著ではなく、全体的にやや満足したようである。

表7. クラス別文法到達度テスト得点結果について

クラス	文法事前テスト ( $\alpha=.85$ )			文法事後テスト ( $\alpha=.92$ )		
	n	M	SD	n	M	SD
中級1	26	30.38	6.34	21	12.14	2.50
中級2	25	32.36	8.47	24	28.63	7.93
中級群	51	31.35	7.45	45	20.93	10.24

クラス	文法事前テスト ( $\alpha=.85$ )			文法事後テスト ( $\alpha=.92$ )		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
基礎1	25	20.80	5.09	22	26.82	5.21
基礎2	24	19.88	4.29	23	26.78	5.95
基礎群	49	20.35	4.69	45	26.80	5.53

## 考察

考察では解釈的論証に対しての妥当性論証の論拠を挙げることを主な目的とする。まず、得点化から決定までの推論に対する解釈的論証の前提条件と妥当性論証の論拠を表8に示す。

表8. プレイスメントテストの解釈的論証の前提条件と妥当性論証の論拠について

推論	解釈的論証(前提条件)	妥当性論証(論拠)
得点化	<ul style="list-style-type: none"> <li>プレイスメント項目は意図した正解を受験者が解答でき、高得点者がより正解できる</li> </ul>	<ul style="list-style-type: none"> <li>ほぼすべての項目が機能していたことから受験者の解答は適切に観測得点へ得点化された</li> </ul>
一般化	<ul style="list-style-type: none"> <li>プレイスメント項目は一貫して測定領域から抽出されていて、また項目数は誤差が最小になる程度ある</li> </ul>	<ul style="list-style-type: none"> <li>全項目は測定領域から抽出されており、合成一般化可能性係数が高かったことから誤差が少ないため観測得点は一貫性をもって測定領域得点へ一般化された</li> </ul>
外挿	<ul style="list-style-type: none"> <li>プレイスメントテストレベルはリーディング科目の授業目標と難易度が一致している</li> </ul>	<ul style="list-style-type: none"> <li>学習レベルの難易度は予想通りであったため測定領域得点から適切に目標得点へ外挿された</li> </ul>
決定	<ul style="list-style-type: none"> <li>プレイスメントテスト得点をクラス分けの判断材料として使用し、習熟度別クラス編成したため、受講者のレベルにより適した授業内容を提供できることにより、正の波及効果が生じる</li> </ul>	<ul style="list-style-type: none"> <li>得点順に習熟度別にクラスが編成されており、基礎群の授業内容に対する評価が高く、到達度テストで得点上昇が生じたことから目標得点から用途へ決定が下された</li> </ul>

研究課題1はプレイスメントテスト項目はどの程度意図した正解を受験者が解答でき、高得点者がより正解できるかであった。古典的テスト理論の観点からみると、受験者にとって項目容易度の値が高いまたは低いということで、易しすぎるまたは難しすぎるという項目が90項目中28項目あった。プレイスメントテストは得点のばらつきを



生じさせることでよりの確なレベル判別ができるので、受験者の能力をいかに弁別しているかの項目弁別力が重要であり、その値が低かったのはわずか四項目のみであった。よって、項目容易度の値が不適切な項目があったが、集団基準準拠テストの一種であるブレイスメントテストの主たる目的である得点のばらつきを生じさせるという点では項目弁別力が高い項目が大半であったことから、目的は達成できたといえる。また各項目には正解が明確にあり、高得点者がより確実に正解できるものであったといえる。

項目応答理論の観点からみると、本テストは項目分離指数が高く項目困難度推定値にはばらつきがあり広範囲の能力推定値を測定できるので、適切である。しかし、項目困難度推定値が低く受験者の能力値を測るにはさほど必要ない項目が二項目あった。また、現時点では合計得点が高いまたは低い順に上級、または基礎クラスに振り分けているが、今後項目応答理論を用い分割点 (cut-off point) を設けることで振り分ける場合、分割点の推定値あたりに項目が多くあるほうがより適切に受験者の能力を判別できるのでよいかもしれない (Hudson, 1991)。例えば、 $-0.50$  から  $-1.00$  くらいの推定値に収まる項目がより多くあれば、基礎クラス1か基礎クラス2を履修したほうがいいのかがより正確に判断できる。インフィット平均二乗は項目67以外すべての項目は基準範囲以内であり、これはほぼすべての項目は作成時に意図していた正解に受験者が解答でき、高得点者がより正解できたということがいえる。したがって、研究課題1は達成でき、受験者の解答は適切に観測得点へ得点化された。

研究課題2はブレイスメントテスト項目はどの程度一貫して測定領域から抽出されていて、また項目数は誤差が最小になる程度あるかであった。まず、このブレイスメントテスト開発手順に記述したが、測定領域である文法であれば中学・高校の教科書にある文法事項、語彙であればJACET1000~3000レベル、読解であれば基礎クラスと上級クラスで実際に用いられた本文より項目を抽出しているので代表性は保たれていると判断する。次に、多変量一般化可能性理論を用い、どの程度の項目数があれば誤差が最小になるかを検証した。合成一般化可能性理論は文法、語彙、読解の項目数が本テストと同じでそれぞれ40、40、10項目の場合、.92になり十分な値ではあるが、読解の項目数が十項目と少なく、一般化可能性係数が.58と低い結果となった。よって、全体的には項目は一貫して測定領域から抽出されており、誤差が少ないが、読解の項目数のみ少なく誤差がより大きい。誤差を最小にし、全セクションともに一般化可能性係数を確実に.80以上にするにはテストを改訂し、各セクションの項目数をそれぞれ30項目にする必要がある。しかしながら、研究課題2は十分に達成でき、観測得点は一貫性をもって測定領域得点へ一般化された。

ブレイスメントテストの項目分析を行い信頼性を検証した既存研究と本研究結果を比較する。Culligan and Gorsuch (1999) では読解と聴解150項目を含むSLEPテスト (Secondary Level English Proficiency Test) を日本人大学生に実施したところ、項目弁別力が.20以上であった項目は66項目のみであった。また、信頼性係数 (KR-20) は.81であった。Westrick (2005) は日本人大学生161名に読解、文法知識、語彙知識を測る目的に開発された60項目から成るQPT-PPT (the Quick Placement Test - Pen and Paper Test) を実施したところ、項目弁別力が.20以上であった項目はわずか23項目で、信頼性係数 (KR-20) は.55であった。既存研究結果と比較しても不良項目がほとんどなく、一般化可能性係数が高い結果を示す本研究結果のほうが数段とよい。また、Beglar (2010) でのラッシュ項目信頼性 (.95以上) や Koizumi et al. (2011) で提示さ



れた項目弁別力平均値( $M = .45$ )と合成一般化可能性係数( $p^2 = .93$ )と比べ、本研究結果は項目弁別力平均値が.40で合成一般化可能性係数が.92であることからほぼ同等だといえる。Chapelle, Chung et al. (2010)で17項目からなる発信型文法テストをもとに算出されたクロンバック係数( $\alpha = .82$ )より本研究の合成一般化可能性係数のほうが高い。これはやはり市販のテストをそのまま使用するのではなく、カリキュラムを考慮し本学部独自のテストを作成し、試行テストの結果をもとに改訂したためであろう。

研究課題3はプレイスメントテストレベルはどの程度リーディング科目の授業目標と難易度が一致しているであった。文法学習レベルは中学校より高等学校、語彙学習レベルはJACET1000レベルよりJACET3000レベル、読解学習レベルは基礎レベルより上級レベルのほうが高いことが予想された。FACETS分析により学習レベル困難度推定値を算出した結果、JACET2000レベルとJACET3000レベルにおいて学習レベル困難度推定値がほぼ同等になった以外は、他の推定値は予想通りであった。基礎クラスではbe動詞から分詞構文までを学習することになっていた。図2をみると、基礎クラスの履修生はJACET1000の語彙レベルにはほぼ到達しているがそれ以外の学習レベルにはさほど到達していないことから、適切な難易度の分割点で、基礎クラスのクラス分けがなされていたと考えられる。また、上級クラスの履修生は中等教育での学習事項は習熟していることを前提とし、それにもとづき読解指導が行われる。このテスト結果から上級クラス履修生は上級レベルの読解設問以外は到達していることがわかった。よって、適切な難易度の分割点で、上級クラスのクラス分けがなされていたと考える。したがって、研究課題3は達成でき、測定領域得点から適切に目標得点へ外挿された。

Beglar (2010)は語彙出現頻度順に項目困難度推定値の平均値を算出した。高頻度の語彙レベルの値は低く、順に高くなっており、Kane (2006)の枠組みでは外挿にあたる本質的側面を評価している。本研究でもJACET1000レベルよりJACET2000レベルとJACET3000レベルのほうが学習レベル困難度推定値は高く、Beglarと同様の結果が得られたため、高頻度の語彙ほど習熟しやすいということがいえる。Chapelle, Chung et al. (2010)では第二言語習得理論の分野で発表された文法習得順序をもとに、文法項目を初級レベル、中級レベル、上級レベルと三段階に区分けをした。項目容易度平均値は予想通り初級レベル、中級レベル、上級レベルの順に高くなり、初級レベルの文法項目がもっとも習得しやすいという結果になり、Chapelle, Chung et al. はKaneの枠組みでは外挿にあたる説明の妥当性論証の論拠として用いている。本研究でも文法学習レベルを区分けした結果、中学校学習レベルのほうが推定値は低くなったことから、中学校学習レベルのほうが習熟しやすいということがいえる。

研究課題4はプレイスメントテスト得点をクラス分けの判断材料として使用し、習熟度別にクラス編成したため、受講者のレベルにより適した授業内容を提供できることにより、どの程度正の波及効果が生じるかであった。基礎クラスの履修者は学習支援などを含むリメディアル教育を受け、また彼らの習熟度に合った学習目標が設定されたことにより、基礎クラスの履修生にはより高い波及効果があったといえるだろう。その論拠として30回の授業を受けた結果、事前テストと比べ事後テストの得点平均値は上昇していた。また、アンケート結果でも学習支援は役に立ったとの意見があり、波及効果の論拠として挙げられるであろう。しかし、基礎群の得点上昇は顕著ではな

く、また教育効果以外にも多様な要因があると思われるので決定の論拠としては若干弱いと考える。また、到達度テスト得点とアンケート結果では直接プレイメントテストの波及効果を検証しているわけではなく、間接的な影響なので若干弱い論拠ではある。したがって、研究課題4は論拠としては若干弱いと達成されたとみなし、目標得点から用途へ決定が下された。

## 結論

教育的示唆としては第一にKane(2006)の論証型妥当性枠組みは明確であり簡潔であるということが挙げられる。つまり、本研究のように教員が項目分析、信頼性の算出などを行って期末テストなどを含むテスト得点の解釈と使用の妥当性を評価することが可能である。また、妥当性を評価することで問題点なども明確になることから、Kane(2006)の枠組みを用いて妥当性を評価することは重要である。本研究では決定の推論に対する論拠が弱かったため、習熟度別に編成することは各レベルを履修している履修者に利益または不利益があるか、また正または負の波及効果があるかを今後さらに直接的に検証する必要がある。第二に多変量一般化可能性理論の有用性である。一般化可能性理論ではセクションごとにG研究を行わないといけないうことに対し、多変量一般化可能性理論を用いることで複数のセクションから成るテストでも一回のG研究だけで済み、またセクション間の相関関係や合成一般化可能性係数も求めることができるため、非常に応用の幅が大きい理論であるといえる。第三は多相ラッシュモデルの有用性である。このモデルを用いることでいくつもの相の推定値を同一上の間隔尺度に変換でき、また、ある受験者がある相に属す項目に何パーセントの確率で正解できるかを推定することができる。これは目標設定が受験者に一致しているかなどを判断する上では最適な判断材料になるのではないかと。

本研究ではKane(2006)が提唱する論証型妥当性枠組みを用い大学一年生対象必修リーディング科目のクラス分け目的で開発したプレイメントテストの解釈的論証を挙げ、それに対する妥当性論証を挙げ、テスト得点の解釈と決定についての妥当性を評価した。ほぼすべての項目に対し受験者は意図した正解に解答し、高得点者が正解できる項目がほとんどであったことが得点の妥当性論証の論拠である。また、合成一般化可能性係数が.90以上になったことから、項目は測定領域から一貫して抽出されており、項目数は誤差を最小にするために十分であったことが一般化の妥当性検証の論拠である。FACETS分析を用いて学習レベル困難度推定値を求めた結果、推定値は予想通りの難易度を示し、また授業目標は履修生に適切に設定されていることを外挿の論拠とした。到達度テストを基礎クラスと中級クラス履修者を対象に事前・事後テストとして実施し平均値差を検証した結果、基礎クラス履修者は得点を上昇させることができた。また、授業評価アンケートにより学習支援など指導が効果的だったという意見が示された。したがって、プレイメントテストの波及効果があったことが示され、それを決定の妥当性論証の論拠とした。このようにプレイメントテスト得点の解釈と決定についての妥当性を評価することができた。今後、D研究の結果にある通り、各セクションの項目数を30項目になるようテスト改良をすることで各セクションの誤差を最小にする、実証型のみで外挿の妥当性論証論拠を提示したので分析型で求めた論拠も提示する、および直接授業を観察するなどして波及効果を検証することでよりテスト得点の解釈と使用の妥当性を向上させることができるであろう。

## 謝辞

この研究ができたのも多くの教員と事務員のおかげです。ここに感謝を述べたいと思います。編集長と査読者にこの研究について貴重なご指摘をいただいたことに感謝をいたします。また、佐藤敬典氏(メルボルン大学)にも丁寧にみていただきましたことに感謝を述べます。本研究に残る誤りはもちろんわたくしの責任です。

熊澤孝昭(*Takaaki Kumazawa*)は現在関東学院大学法学部所属准教授である。Temple Universityより教育学(博士)を取得した。専門分野は英語教育学で、特にテスト理論、第二言語習得理論、カリキュラム開発に興味がある。

## 引用文献

- American Educational Research Association/American Psychological Association/  
National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2, 1-34. doi:10.1207/s15434311laq0201\_1
- Bachman, L. F., & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University.
- Bachman, L. F., & Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University.
- Beglar, D. (2010). A Rasch-based validation of the vocabulary size test. *Language Testing*, 27, 101-118. doi:10.1177/0265532209340194
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model*. Mahwah, NJ: Erlbaum.
- Brennan, R. L. (2001a). *Generalizability theory*. New York: Springer.
- Brennan, R. L. (2001b). mGENOVA (version 2.1) [Computer software]. Iowa City, IA: The American College Testing Program.
- Brown, J. D. (1989). Improving ESL placement tests using two perspectives. *TESOL Quarterly*, 23, 65-83. doi:10.2307/3587508
- Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment*. New York: McGraw-Hill College Press.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validity by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105. doi:10.1037/h0046016

- Chapelle, C. A. (2008). The TOEFL validity argument. In C. A. Chapelle, M. K. Enright, & J. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 319-352). London: Routledge.
- Chapelle, C. A., Chung, Y., Hegelheimer, V., Pendar, N., & Xu, J. (2010). Towards a computer-delivered test of productive grammatical ability. *Language Testing*, 27, 443-469. doi: 10.1177/0265532210367633
- Chapelle, C. A., Enright, M., & Jamieson, J. (2010). Does an argument-based approach to validity make a difference? *Educational Measurement: Issues and Practice*, 29, 3-13. doi: 10.1111/j.1745-3992.2009.00165.x
- Cheng, L., & Watanabe, Y. (Eds.). (2004). *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Erlbaum.
- Coniam, D. (2008). An investigation into the effect of raw scores in determining grades in a public examination of writing. *JALT Journal*, 30, 69-84.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.) (pp. 443-507). Washington, DC: American Council on Education.
- Cronbach, L. J. (1980). Selection theory for a political world. *Public Personnel Management*, 9, 37-50.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements*. New York: Wiley.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302. doi:10.1037/h0040957
- Culligan, B., & Gorsuch, G. (1999). Using a commercially produced proficiency test in a one-year core EFL curriculum in Japan for placement purposes. *JALT Journal*, 21, 7-25.
- Cumming, A., & Berwick, R. (Eds.). (1996). *Validation in language testing*. Clevedon, UK: Multilingual Matters.
- 大学英語教育学会基本語改訂委員会. (2003). *JACET list of 8000 basic words*. Tokyo: JACET.
- Hudson, T. (1991). Relationships among IRT item discrimination and item fit indices in criterion-referenced language testing. *Language Testing*, 8, 160-181. doi:10.1177/026553229100800205
- Kane, M. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112, 527-535. doi:10.1037/0033-2909.112.3.527
- Kane, M. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38, 319-342. doi:10.1111/j.1745-3984.2001.tb01130.x

- Kane, M. (2004). Certification testing as an illustration of argument-based validation. *Measurement, 2*, 135-170. doi:10.1207/s15366359mea0203\_1
- Kane, M. (2006). Validation. In R. Brennan (Ed.), *Educational measurement* (4th ed.) (pp. 17-64). Westport, CT: Greenwood Publishing.
- Kane, M. (2012). Validating score interpretations and uses. *Language Testing, 29*, 3-17. doi:10.1177/0265532211417210
- Kane, M., Crooks, T., & Cohen, A. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice, 18*, 5-17. doi:10.1111/j.1745-3992.1999.tb00010.x
- Koizumi, R., Sakai, H., Ido, T., Ota, H., Hayama, M., Sato, M., & Nemoto, A. (2011). Development and validation of a diagnostic grammar test for Japanese learners of English. *Language Assessment Quarterly, 8*, 53-72. doi:10.1080/15434303.2010.536868
- 熊澤孝昭. (2010). 多肢選択式項目の項目形式が文法テストパフォーマンスに与える影響について. *JALT Journal, 32*, 169-188.
- Kunnan, A. J. (Ed.). (1998). *Validation in language assessment*. Mahwah, NJ: Erlbaum.
- Lee, Y. (2006). Dependability of scores for a new ESL speaking assessment consisting of integrated and independent tasks. *Language Testing, 23*, 131-166. doi:10.1191/0265532206lt325oa
- Linacre, J. M. (1989). *Many-facet Rasch measurement*. Chicago: MESA.
- Linacre, J. M. (2002). FACETS (Version 3.41) [Computer software]. Chicago: MESA.
- Lissitz, R. W. (Ed.). (2009). *The concept of validity*. Charlotte, NC: Information Age.
- Messick, S. (1981). Evidence and ethics in the evaluation of tests. *Educational Researcher, 10*, 9-20. doi:10.3102/0013189X010009009
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 13-103). New York: Macmillan.
- Messick, S. (1995). Validity of psychological assessment. *American Psychologist, 50*, 741-749. doi:10.1037/0003-066X.50.9.741
- Pardo-Ballester, C. (2010). The validity argument of a web-based Spanish listening exam: Test usefulness evaluation. *Language Assessment Quarterly, 7*, 137-159. doi:10.1080/15434301003664188
- Sawaki, Y. (2007). Construct validation of analytic rating scales in a speaking assessment: Reporting a score profile and a composite. *Language Testing, 24*, 355-390. doi:10.1177/0265532207077205
- Schaefer, E. (2008). Rater bias patterns in an EFL writing assessment. *Language Testing, 25*, 465-493. doi:10.1177/0265532208094273

- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.
- 霜崎實他. (2006). *New Crown English Series I & II*. 東京:三省堂出版.
- 高橋貞雄他 (2005). *New Crown English Series 1, 2, & 3*. 東京:三省堂出版.
- Wainer, H., & Braun, H. I. (Eds.). (1988). *Test validity*. Hillsdale, NJ: Erlbaum.
- Weir, C. (2005). *Language testing and validation: An evidence-based approach*. New York: Palgrave Macmillan.
- Westrick, P. (2005). Score reliability and placement testing. *JALT Journal*, 27, 71-94.
- Xi, X. (2007). Evaluating analytic scoring for the TOEFL Academic Speaking Test (TAST) for operational use. *Language Testing*, 24, 251-286. doi:10.1177/0265532207076365

# Research Forum

## The Role of Think-Aloud and Metacognitive Strategies in L2 Meaning-Inference During Reading

Megumi Hamada  
*Ball State University, U.S.A*

Chaehee Park  
*Sunmoon University, Republic of Korea*

Lexical inference is an important word learning method, yet it is still unknown what kind of instruction will improve inference accuracy. In this study we investigated whether think-aloud can enhance metacognitive strategies while inferring the meaning of unknown words encountered during reading and lead to higher inference accuracy. Two groups of college-level Korean L1 ESL students inferred the meanings of pseudowords in a short passage in a pretest-posttest design. The think-aloud group ( $n = 19$ ) performed the task verbalizing their thoughts, while the control group ( $n = 20$ ) performed it silently. The results indicated a significant gain in inference accuracy for the think-aloud group. Based on further analyses of the think-aloud protocol, the study addresses the importance of the quality of think-aloud in L2 performance.

字句推論は重要な語彙学習方法であるが、推論の正確さを向上する方法はまだ分かっていない。本研究は、読書中に出てくる未習の語彙の意味を推測する際に、声に出して考えるという方法が、メタ認知を促進することによって、高い正確率をもたらすかどうかという調査を行った。事前・事後テストのデザインで、2グループの韓国語を母語とするESL学習者が、短い文章に含まれている偽単語の意味を推測した。声に出して考えるグループ ( $n = 19$ ) はそのタスクの間考えを声に出しながら、一方コントロールグループ ( $n = 20$ ) はタスクを無言で行った。結果は声に出して考えるグループの推測正確度が大幅に向上したことを示した。プロトコロの更なる分析をふまえて、本研究は声に出して考える内容の質が重要であることを述べる。



Word knowledge is crucial in all aspects of second-language (L2) learning throughout the proficiency levels. In the past decade, there has been increasing interest in the nature of this knowledge and its acquisition. One growing area of research focuses on lexical inference while reading. Lexical inference, hereinafter called meaning-inference, involves “making informed guesses as to the meaning of a word in light of all available linguistic cues in combination with the learner’s general knowledge of the world and awareness of context” (Haastrup, 1991, p. 40). Meaning-inference is also known as “incidental” word learning because it can occur as a by-product of other activities, such as reading (Huckin & Coady, 1999). The indirect nature of meaning-inference is appealing, and this method is suggested to be “the most important of all sources of vocabulary learning” (Nation, 2001, p. 232).

Despite the fact that meaning-inference has been widely incorporated into teaching, teachers often encounter “wild guesses” – the fact that not every student is able to infer the correct meaning of unknown words (e.g., Kaivanpanah & Alavi, 2008; Kelly, 1990). Although a number of findings have been reported on meaning-inference, findings concerning instructional techniques or strategies that enhance inference accuracy are still limited. This study focused on the use of think-aloud during meaning-inference, based on research in reading comprehension. A think-aloud procedure, originally developed in the field of psychology, is a commonly used research technique both in first language (L1) and L2 (Bowles, 2010). This technique requires research participants to verbalize their thought processes during or after a given task, and the verbal protocol generated from the think-aloud is used to reveal the participants’ thought processes while performing the task (Ericsson & Simon, 1993; Pressley & Afflerbach, 1995). A number of L1 studies suggest that think-aloud can be used to promote the use of metacognitive strategies, which in turn facilitates reading comprehension (e.g., Baumann, Seifert-Kessell, & Jones, 1992; Loxterman, Beck, & McKeown, 1994).

With the assumption that both reading comprehension and meaning-inference during reading are outcomes of cognitive processes and facilitated by metacognitive strategies, this study focused on the relationship between think-aloud, metacognitive strategies, reading comprehension, and meaning-inference during reading in the L2 context with Korean L1 college students enrolled in an intensive English program in the United States. This college-level learner population, both in South Korea and in Japan, often has basic skills in English yet has difficulty in reading advanced or authentic materials due to a lack of vocabulary knowledge. This study aimed to provide further



findings on vocabulary learning among Northeast Asian college-level learners. To clarify the relationships between meaning-inference during reading, reading comprehension, and think-aloud, the following section presents a review of relevant research.

## Literature Review

### *Think-Aloud*

Think-aloud is categorized into two types, concurrent or retrospective, depending on the timing of the task administration. Concurrent think-aloud requires learners to verbalize their thought processes as they work on a given task, and retrospective think-aloud requires verbalization after task completion. Both types of think-aloud are intended to reveal what thought processes participants are actually going through while performing a learning activity (Ericsson & Simon, 1993; Pressley & Afflerbach, 1995). The verbal reports generated through think-aloud are further categorized as nonmetalinguistic or metalinguistic, depending on the quality of thought processes reflected in the report (Ericsson & Simon, 1993). Nonmetalinguistic (or nonmetacognitive) think-aloud is defined as a report including thoughts *per se*, while metalinguistic (or metacognitive) think-aloud is defined as including deliberate thoughts such as reasons, explanations, and justifications of specific decisions made during task performance (Bowles & Leow, 2005).

Although think-aloud has provided valuable contributions to language research, its use may be questionable due to the reactivity issue. The reactivity of think-aloud refers to the problem that concurrent think-aloud potentially alters outcome results by influencing participants' cognitive processing while they are completing a given task (Ellis, 2001; Jourdenais, 2001). The main issues investigated in this line of research are reactivity in the accuracy and latency of participants' task performance. Think-aloud may influence the accuracy on outcome measures and the latency of time on task, yet current findings are still inconclusive. For example, regarding accuracy, Sanz, Lin, Lado, Bowden, and Stafford (2009) and Rossomondo (2007) reported positive reactivity, suggesting that think-aloud led to higher gain in receptive and productive knowledge of L2 grammar introduced through written input, whereas Egi (2008) reported nonreactivity in the recall of new L2 grammar introduced through spoken input. Examining reactivity in L2 writing, Sachs and Polio (2007) reported that think-aloud is negatively reactive, suggesting that think-aloud led to a decrease in the number of errors corrected in the

revised essay in Experiment 1 (within-participant design), while reporting nonreactivity in Experiment 2 (between-participant design) on the same issue. Regarding time, some studies report negative reactivity, suggesting that think-aloud increased time on task (e.g., Bowles, 2008; Bowles & Leow, 2005; Yoshida, 2008, Experiment 1 in Sanz et al., 2009), whereas others report nonreactivity (e.g., Sachs & Suh, 2007; Experiment 2 in Sanz et al., 2009).

The reactivity of think-aloud has been investigated in reading comprehension tasks as well. A majority of findings report nonreactivity (Bowles & Leow, 2005; Leow & Morgan-Short, 2004; Rossomondo, 2007; Yoshida, 2008), although there exists research that reports otherwise (Goo, 2010). For example, in Leow and Morgan-Short (2004), 77 college-level learners of Spanish, divided into a think-aloud group and a non-think-aloud group, read Spanish texts and answered multiple-choice reading comprehension questions written in their L1, English. The think-aloud group reported to a laboratory and was instructed to voice their thoughts aloud throughout the text reading and comprehension questions, whereas the non-think-aloud group remained in the classroom and completed the tasks silently. The results indicated that there was no significant difference between the two groups' reading comprehension scores.

Likewise, Bowles and Leow (2005) examined reactivity in reading comprehension among college-level Spanish L2 students and found a nonsignificant difference between the think-aloud and the silent control groups. They also took the important step of examining the effect of type of think-aloud in reactivity. The think-aloud group was divided into a nonmetalinguistic (or nonmetacognitive) think-aloud group, in which the participants were required to say aloud whatever passed through their minds while performing the task, and a metalinguistic (or metacognitive) think-aloud group, in which they were required to say aloud more specific, additional information, such as justification and reasoning for their answers. The results indicated that the nonmetacognitive think-aloud group scored significantly better than their counterpart, implying that metacognitive think-aloud was cognitively more demanding and interfered with reading comprehension. This finding suggests that the complexity of information verbalized during think-aloud clearly plays a role in determining the existence of reactivity.

### **Meaning-Inference During Reading**

Meaning-inference is a complex construct because it entails multiple processes such as analyzing, extracting, and integrating textual information with the reader's background knowledge. Sternberg (1987) identified the major sequential operations involved in meaning-inference during reading: (a) separating relevant from irrelevant text information for the purpose of inferring the meaning of an unknown word, (b) combining relevant textual cues to formulate a workable definition, and (c) evaluating the hypothesized meaning against information from the subsequent context. Because the main source of information or cues used for meaning-inference is the text, accurate comprehension of the text is crucial for accurate meaning-inference.

A number of L2 meaning-inference studies have investigated what cues learners tend to use in meaning-inference during reading. The cues are generally categorized as either local or global cues. Local cues include morphological structure, word analogy, and grammatical (syntactic) structure, wherein learners do not necessarily have to apply their understanding of the text in meaning-inference. For example, Huckin and Bloch (1993) investigated college-level ESL learners' meaning-inference behaviors and reported that the cue that was most important for accurate meaning-inference was the morphological structure of an unknown word.

In contrast, global cues involve more contextually based analysis that requires a deeper level of understanding of the text as well as world knowledge, such as background knowledge and the existence of schemata related to the text, rather than an analysis solely of the given text. It is important to note that learners may use multiple cues, including both local and global cues, in inferring the meaning of a single word. There are a handful of findings that suggest that using local cues alone tends to yield inaccurate inference (e.g., Chern, 1993; Haynes, 1993; Nassaji, 2003). For example, Haynes (1993) reported that an adult ESL student inferred "the end of spring" for *offspring* based on an incorrect morphological analysis (*off + spring*), despite the fact that the inferred meaning did not match the context. Similarly, Nassaji (2003) found that adult ESL learners incorrectly inferred meanings of unknown words based on graphic similarity of words (word analogy), such as the similarity of *permeated* to *meat* and *affluence* to *influence* (p. 653). Although local cues may not always lead to accurate inference, they are more popular than global cues among learners (e.g., Bensoussan & Laufer, 1984; Haynes, 1993). These findings suggest that local cues are more easily used by learners, presumably because local cues do not require as much comprehension of the text as global cues do.

### ***Metacognition in Reading and Meaning-Inference***

Metacognition is defined as “one’s knowledge concerning one’s own cognitive processes and products or anything related to them” (Flavell, 1976, p. 232). Metacognitive strategies promote conscious awareness of one’s cognitive processes, such as planning for learning, thinking about the learning process as it is taking place, monitoring of one’s production or comprehension, and evaluating learning after a task is completed (Purpura, 1997). In reading comprehension, different levels of cognitive processes are involved, such as visual, phonological, and semantic processes, word recognition, syntactic parsing, and discourse processes, and all of the processes contribute to the accurate comprehension of a text (e.g., Koda, 2005). Metacognition plays an important role in controlling these processes, and there has been a large volume of research, particularly in L1 reading instruction, that investigates metacognitive strategies that enhance reading comprehension. One strategy with empirical support for its effectiveness is think-aloud. Because think-aloud can reveal certain cognitive processes, using think-aloud while comprehending a text makes it possible to teach what cognitive processes need to be involved in reading comprehension. This point is discussed more in the next section.

In meaning-inference during reading, separating and combining textual cues and evaluating an inferred meaning (Sternberg, 1987) are involved in addition to the processes required in reading comprehension. Thus, as in reading comprehension, metacognition is necessary to facilitate the meaning-inference processes. For instance, in Nassaji (2003), 21 intermediate-level adult students with various L1 backgrounds (Arabic, Chinese, Persian, Portuguese, and Spanish), who were enrolled in a 12-week intensive English program, engaged in a meaning-inference task. Each student individually met with the researcher and was asked to read aloud a short passage which contained target words unfamiliar to them. They were then asked to infer the meaning of the words, using the think-aloud technique, in English. They were also asked to underline any other unfamiliar words in the passage and infer the meaning of the words, using think-aloud. The analysis of the verbal protocols indicated that “verifying” (evaluating whether the inferred meaning is correct) and “self-inquiry” (asking questions about a word or the inferred meaning of a word) in the think-aloud protocol were related to higher inference accuracy (p. 662), suggesting the importance of metacognitive strategies in meaning-inference during reading.

### ***Think-Aloud in Reading Instruction***

Think-aloud has also been an important technique for instructional purposes. The most prominent implementation of think-aloud is seen in reading comprehension instruction, mainly targeting L1 readers. This instructional implementation of think-aloud is based on the view that reading comprehension is a complex cognitive activity that requires a strategic planning and problem solving process (Kucan & Beck, 1997). Because think-aloud is able to reveal learners' cognitive processes, using think-aloud while comprehending a text makes it possible to teach what cognitive processes need to be involved for reading comprehension. In other words, what is essentially taught using think-aloud are metacognitive strategies that facilitate comprehension processes.

In order to incorporate metacognitive strategies, typical think-aloud instruction includes teacher modeling, using concurrent think-aloud, to show the way comprehension should be carried out and discussion questions or activities that direct students to be engaged in their comprehension processes (e.g., Maria & Hathaway, 1993; Oster, 2001; Walker, 2005). The effectiveness of think-aloud instruction in L1 reading comprehension has been reported in many studies (e.g., Baumann et al., 1992; Bereiter & Bird, 1985; Berne, 2004; Laing & Kamhi, 2002). For example, in Laing and Kamhi (2002), 40 third-graders were divided into groups of average and below-average readers and instructed with two conditions: listen through and think-aloud. After reading a story, the children answered comprehension questions. As expected, the comprehension accuracy was significantly better for both groups in the think-aloud condition. Given the empirical support, think-aloud instruction has practically become a standard method in reading comprehension instruction with native speaking students.

As for the use of think-aloud in L2 reading comprehension instruction, current findings are far fewer than in L1 research, despite the fact that there have been a number of suggestions and recommendations that metacognitive strategies are crucial in L2 reading comprehension (e.g., Block, 1992; Carrell, 1989; Casanave, 1988). As for empirical studies, McKeown and Gentilucci (2007) investigated whether think-aloud instruction, using concurrent think-aloud, improves middle school ESL learners' content area reading comprehension. Due to a small number of participants ( $N = 27$ ), all of them did a pretest, treatment (think-aloud instruction), and a posttest. In contrast to findings from L1 studies, the results indicated that none of the student groups (early intermediate, intermediate, and early advanced) showed a reliable gain in reading comprehension scores in the posttest. However, two

issues need to be noted in interpreting their results. There was no control group, which makes the research design of this study questionable. Another issue is that the think-aloud instructions were administered in English and the learners were asked to think-aloud in English, rather than their L1. Given the fact that the students had relatively low English proficiency, it is possible that think-aloud in their L2 did not completely reflect their metacognition.

## **Research Question**

As summarized above, in reading instruction, think-aloud is widely used as an activity or technique to enhance the use of metacognitive strategies, that is, to make students aware of their own cognitive processes in reading comprehension. The underlying assumption is that think-aloud can reveal certain cognitive processes (Ericsson & Simon, 1993). In other words, being able to think-aloud their own cognitive processes indicates that the learners are aware of the actual processes. Cognitive processes are involved in not only reading comprehension, but in meaning-inference as well. That is, if metacognitive strategies play a role in reading comprehension, it seems reasonable to argue that metacognitive strategies also play a role in meaning-inference, because both activities involve cognitive processes. Although a number of studies in L2 meaning-inference during reading have used think-aloud as a research technique for analyzing learners' cognitive processes (e.g., Nassaji, 2003; Paribakht & Wesche, 1999), whether the use of think-aloud influences the outcome of meaning-inference has not been extensively examined. Thus, the research question investigated in this study was:

Does the use of think-aloud while inferring the meaning of unknown words encountered during reading improve inference accuracy?

The research question was tested using a pretest-posttest design with two groups of participants, a control group vs. an experimental group. The two groups varied in terms of treatment – whether or not think-aloud was included in the treatment. In this study, think-aloud referred to metacognitive think-aloud, which is “explicit” verbalization of learners' thoughts on the task, such as planning, monitoring, reasoning, and evaluation, because explicit verbalization of thoughts requires more explicit metacognitive processes and thus serves more effectively as a metacognitive strategy in meaning-inference.

## Method

### Participants

The participants were 39 Korean L1 ESL learners who were enrolled in reading/writing courses in an intensive summer English program in a mid-size university in the U.S. They were randomly assigned to one of the groups, the think-aloud group ( $n = 19$ ), consisting of five males and 14 females, and the control group ( $n = 20$ ), consisting of 11 males and nine females. The mean age was 22.74 ( $SD = 2.16$ ) for the think-aloud group and 22.95 ( $SD = 1.73$ ) for the control group. The mean reading placement test scores are from the *ACT Compass Test*. This test, administered online by *ACT* (originally named American College Testing), is a standardized test, which includes English as a foreign/second language (similar to TOEFL) as well as academic subjects for native-speaking students (similar to the SAT in the United States). The mean reading comprehension scores (maximum 100 points) were 83.16 ( $SD = 8.18$ ) for the think-aloud group and 81.45 ( $SD = 7.06$ ) for the control group. This difference was found to be nonsignificant,  $t(37) = -.699, p = .402$ . The majority of the students were in an English-speaking country for the first time and had 5 weeks of experience living in the U.S. at the beginning of data collection, except for one student from the think-aloud group who indicated 4 months and one student from the control group who indicated 14 months. All of the participants, whose L1 was Korean, were from the same university in South Korea and were majoring in various fields. In each participant group, more than 50% majored in either engineering or business. The authors were not the participants' teachers. Participation was voluntary.

### Materials

Passages for the pretest and posttest were first selected. The criteria for selecting the passages were as follows: (a) they were approximately the same length; (b) the grammar and vocabulary items used in the passages were known to the participants; and (c) the passage topics did not require highly specialized background knowledge. The pretest passage, "When a Young Bird Leaves the Nest" (236 words), was selected from Chern (1993) (see Appendix A) and the posttest passage, "Folk Objects" (253 words), was selected from Hamada and Koda (2011) (see Appendix B). Each passage contained 10 pseudowords, whose meanings the participants were asked to infer. The pseudowords served as unfamiliar L2 words to be learned by the participants. Although the use of pseudowords inherently creates

an unnatural setting, we decided the use of pseudowords was the optimal solution to the design of this study. The use of pseudowords in vocabulary research is a common practice, particularly in experimental studies, because it can minimize the effects of participants' previous vocabulary knowledge and the frequency of real words (Kirsner, 1974). The proportion of known to unknown word coverage for the pretest and posttest passages was 95.76% and 96.05%, respectively. These percentages are above the minimum 95% requirement for meaning-inference to occur, suggested by Liu and Nation (1985).

In order to ensure that the participants would not have difficulty with the grammar or vocabulary items used in the passages except for the 10 pseudowords, their instructors checked the appropriateness. They indicated that the grammar was basic enough and should be known to the participants, but some of the vocabulary items might not be known. Therefore, two weeks prior to the pretest, the participants looked at both the pretest and posttest passages and underlined words whose meanings were unknown to them. This was done as part of their regular class activity and took no longer than 10 minutes. Two Korean L1 graduate students majoring in linguistics/TESOL created a glossary based on the participants' responses, so that the participants would know all of the vocabulary items, except for the pseudowords. Regarding the criteria on the topic, our goal was to not choose a topic that would be intelligible only to those who have specialized knowledge. After discussion with the participants' teachers, we determined that the topics were general enough to be understood by laypeople.

In an effort to examine the comparability of the pretest and posttest passages, ten native speakers of English who were enrolled in degree programs in the same U.S. university read the two passages, inferred the meanings of the pseudowords, and judged the difficulty of the passages. The mean rating for Passage A (pretest) was 2.3 ( $SD = 1.16$ ), and the mean rating for Passage B (posttest) was 2.6 ( $SD = .97$ ), where the scale was 1 very easy, 2 easy, 3 neutral, 4 difficult, and 5 very difficult. The mean rating scores were tested using a two-tailed  $t$  test, and the difference was found to be nonsignificant,  $t(18) = -.629, p = .538$ .

## **Tasks and Procedures**

### *Pretest*

The pretest was administered by the instructors to all of the students in the classroom. The participants read the passage and wrote down the



inferred meanings in both English and Korean. The reason for having them answer both in L2 and L1 was to minimize the influence of L2 proficiency on identifying the correct definition for each pseudoword. The instructors gave a maximum of 30 minutes to complete the pretest, although they noted that most of the participants completed it in 15-20 minutes.

### *Treatment*

The treatment was administered either one day or two days after the pretest. Due to the absence of some of the participants on a field trip, we were unable to schedule the treatment on the same day for all of the participants. The instructors first explained the inference strategies from Nassaji (2003) in class as part of a class session to all of the participants. After the strategy explanation, the control group was asked to stay in the classroom and work on practice inference using the practice passage. They were told to use the strategies introduced, but worked silently. When they completed the practice passage (i.e., wrote down the inferred meanings), they turned it in and were dismissed. As for the think-aloud group, after the strategy explanation provided in class, they worked on practice inference in a separate room, where they received think-aloud instruction from one of the three teaching assistants, Korean L1 graduate students in TESOL who were trained in instructing the think-aloud technique. The participants were asked to use think-aloud as they worked on inference using the practice passage.

The teaching assistants gave the following metacognitive think-aloud instructions in Korean: (a) Say everything that comes across your mind while guessing, but focus on voicing in detail the thought process you used to reach your answer; and (b) Use either Korean or English, whichever is more comfortable for you. The teaching assistants reminded them to keep verbalizing when they stopped doing so for more than 10 seconds. The teaching assistants were not allowed to respond to the participants' questions or utterances during think-aloud. The assistants were told that their role was to ensure that the participants use think-aloud, rather than to help them to infer the meaning of unknown words correctly or to serve as an active listener. The assistants also modeled think-aloud, which included both syntactically complete and incomplete sentences. For both groups, we set the time-on-task as 15 minutes each for the strategy explanation and for the practice inference.

### *Posttest*

The posttest was administered one day after the treatment. Each participant in the think-aloud group completed it individually in a quiet room in the presence of one of the three teaching assistants. Each participant was tested one at a time, and a Sony digital voice recorder was placed on the desk near the participant. After the assistant determined that the participant was relaxed and comfortable using think-aloud, the participant worked on the posttest. The think-aloud group's verbal reports were audio-recorded. The control group worked silently in the classroom on the posttest, administered by their instructors. For the control group, the instructors made sure to create an individual testing atmosphere (not a class activity), and for the think-aloud group, the teaching assistant made sure to be noninterruptive (as if he or she was not present). We were unable to schedule individual rooms for the control group, but we made sure that any difference that might arise would be minimal. As in the pretest, both groups were allowed a maximum of 30 minutes for the posttest.

### **Analysis Procedures and Results**

The following scoring system was used in order to best assess meaning-inference accuracy. Accuracy of the inferred meanings (i.e., participants' definitions of the pseudowords) was judged based on two sources: the original words that were replaced in the passages and the definitions given by the ten native speakers who performed the same task. The participants' definitions were scored following the criteria used in Haynes and Carr (1990). One point was awarded (a) when the definition for a particular pseudoword matched that of the original word; (b) when the definition matched one of the native speakers' definitions; or (c) when the definition was semantically identical (or synonymous) to the original word or one of the native speakers' definitions. A half point was awarded when the inferred meaning was semantically close to the original word or one of the native speakers' definitions. The keys for both passages are presented in Appendix C and Appendix D. Scoring was done by two independent raters who were fluent in both English and Korean (interrater reliability .95), and items whose scoring was disagreed upon were resolved by discussion.

**Table 1. Means and Standard Deviations of Pretest and Posttest Scores**

Group	Pretest	Posttest
Think-aloud group	3.58 ( <i>SD</i> = 1.72)	4.66 ( <i>SD</i> = 2.24)
Control group	3.98 ( <i>SD</i> = 1.31)	3.83 ( <i>SD</i> = 1.38)

Table 1 is a summary of the mean scores for the pretest and posttest. The think-aloud group's mean scores were 3.58 (*SD* = 1.72) for the pretest and 4.66 (*SD* = 2.24) for the posttest. The control group's mean scores were 3.98 (*SD* = 1.31) for the pretest and 3.83 (*SD* = 1.38) for the posttest. The Mann-Whitney test was selected for the analysis.

**Table 2. Means and Sums of Ranks**

Group	Mean/Sum	Pretest	Posttest	Difference
Think-aloud group	Mean	18.84	22.84	23.68
	Sum	358.00	434.00	450.00
Control group	Mean	21.10	17.30	16.50
	Sum	422.00	346.00	330.00

Table 2 is a summary of the means and sums of ranks. The mean and sum of ranks were calculated for the pretest, posttest, and the difference between the two. For the think-aloud group, the mean ranks were 18.84 for the pretest, 22.84 for the posttest, and 23.68 for the difference. For the control group, the mean ranks were 21.10 for the pretest, 17.30 for the posttest, and 16.50 for the difference. The Mann-Whitney *U* indicated the groups did not differ significantly for the pretest ranks and the posttest ranks,  $p = .534$  and  $p = .126$ , respectively. However, the difference between the pretest and posttest differed significantly between the groups,  $U(37) = 120$ ,  $Z = -1.982$ ,  $p = .047$ .

To further examine the data, we conducted a qualitative analysis of the think-aloud protocol from the think-aloud group. The verbal reports of the participants who had the top five reading placement test scores ( $M = 91$ ,  $SD = 2.35$ ) and the lowest five reading placement test scores ( $M = 75$ ,  $SD = 8.34$ ) were translated and transcribed in English, then coded into the two types, nonmetacognitive and metacognitive think-aloud. When the verbal report for each pseudoword item included specific information about planning,

monitoring, reasoning, and evaluating, the utterance was coded as metacognitive. When the utterance did not include such information, it was coded as nonmetacognitive. Two raters independently coded the verbal reports (interrater reliability .96), and items whose coding was disagreed upon were resolved by discussion. The protocol analysis excluded one participant's data due to poor audio recording quality.

**Table 3. Comparison of the Participants with Higher and Lower Reading Scores**

Higher/ Lower	Pretest score	Posttest score	Metacognitive think-aloud	Nonmeta- cognitive think-aloud
Higher	3.4 ( <i>SD</i> = 1.52)	5.7 ( <i>SD</i> = 3.29)	5.2 ( <i>SD</i> = 3.35)	4.8 ( <i>SD</i> = 3.35)
Lower	4.2 ( <i>SD</i> = 1.35)	4.4 ( <i>SD</i> = 2.61)	2.6 ( <i>SD</i> = 3.21)	7.4 ( <i>SD</i> = 3.21)

Table 3 is a summary of the mean pretest and posttest scores (max. 10) and number of metacognitive think-aloud protocols (max. 10) categorized for the participants with higher (top five) reading placement test scores and lower (lowest five) reading placement test scores. The high scoring group had a larger increase in meaning-inference accuracy in the posttest scores (2.3 points) than the low scoring group did (0.2 points). The high scoring group's protocol contained both metacognitive and nonmetacognitive information in similar amounts (the difference was only 0.4 points), while the low scoring group's protocol had much more nonmetacognitive information (4.8 points more) than metacognitive information.

## Discussion

The study investigated whether the use of think-aloud while inferring the meaning of unknown words encountered during reading improves inference accuracy. The results demonstrated that the think-aloud group showed a larger increase in meaning-inference accuracy between the pretest and the posttest than the control group did. As predicted, the think-aloud technique seems to have enhanced the participants' metacognitive processes, which in turn facilitated processes involved in meaning-inference. This finding is in line with the consensus in L1 research (e.g., Loxterman et al., 1994), while inconsistent with the result from the L2 reading comprehension study

(McKeown & Gentilucci, 2007). Although the focus of this study was not to test the reactivity issue, the current results suggest that think-aloud was positively reactive in meaning-inference during reading. Similar results, an increase in accuracy, were reported in L2 grammar learning studies (e.g., Rossomondo, 2007; Sanz et al., 2009), yet most L2 reading comprehension studies showed either no reactivity (e.g., Leow & Morgan-Short, 2004) or negative reactivity (e.g., Bowles & Leow, 2005). These findings offer a possible interpretation that meaning-inference involves cognitive processes, perhaps processes involved in problem-solving, more similar to grammar learning than to reading comprehension. Needless to say, further research is warranted to confirm the current findings and interpretations, given the limited amount of research in think-aloud technique and L2 learning, including reading comprehension and meaning-inference. In particular, more findings are necessary to verify whether the role of think-aloud in L1 learning tasks differs from that in L2 learning tasks and whether think-aloud plays a different role in the different learning activities, reading comprehension and meaning-inference.

The follow-up qualitative analysis on the protocol from the think-aloud group also offers interesting insights. The group with higher reading scores seemed to show a larger positive effect from the use of think-aloud in meaning-inference than the group with lower reading scores. Of particular interest is that the low-scoring group's protocol contained much more non-metacognitive information than the high scoring group's protocol did. These results appear to suggest that the use of think-aloud relates to meaning-inference accuracy and the relationship also varies according to individual learners' English proficiency. Although an effort was made to ensure that the two groups, the think-aloud and control groups, were comparable in terms of age and English proficiency, a more strict homogeneity of the participants will be necessary in future studies.

For those who used nonmetacognitive think-aloud more frequently, the overwhelmingly most common characteristic of the verbal protocol was to simply read and reread a portion of the passage. Following is one such example. Because the participant read from the passage, the entire utterance was in English.

“Some individuals ricate with objects as though they were people. . . . Some individuals ricate with objects as though they were people.”

In contrast, those who used metacognitive think-aloud revealed reasoning based mainly on the three pieces of information, morpho-syntactic structure, world knowledge, and discourse knowledge, which Nassaji (2003) reported lead to meaning-inference success. In the examples below, the text portions indicated by single quotation marks were uttered in English, and the rest was uttered in Korean.

#### *Use of morpho-syntactic structure*

“It is a noun if seeing next sentence. . . . ‘They may serve as symbols for social class, kede, or ethnicity’. . . ‘social class . . . object display’ . . . because this is a plural, should be many. . .”

“‘They may serve as symbols for social class, kede, or ethnicity’. . . ‘kede’. . . it is parallel . . . so . . . could be the same meaning as ‘social class,’ and ‘ethnicity’ . . . then religion?”

#### *Use of world knowledge*

“It has been taded that in Utah, one can find driveways lined with wheels, and gates built from commercial objects’ . . . um . . . this is something we can see in Utah . . .”

#### *Use of discourse knowledge*

“Although mailboxes must follow official standards of measurement’ . . . so . . . they should follow the standards, but they do not . . . so it means change.”

Because metacognitive think-aloud requires a deeper level of processing (Ericsson & Simon, 1993), it makes sense that those who were unable to perform metacognitive think-aloud were only able to read aloud a portion of the passage, only one of the tasks involved in meaning-inference while reading. This observation, although qualitative in nature, seems to indicate the interconnection between English proficiency, including proficiency in reading, the quality of think-aloud, and metacognitive strategies.

## **Conclusion**

This study explored the effectiveness of think-aloud in L2 meaning-inference during reading with college-level ESL students. The findings from the study suggest that think-aloud helps facilitate accurate meaning-inference, as shown by the increase in inference accuracy in the student group who

used the think-aloud technique. A follow-up analysis on the protocol from the think-aloud group seems to imply that the individual participants' English proficiency relates to the use of metacognitive strategies, the type of information uttered during think-aloud, and meaning-inference success. Given that this observation is based on a qualitative analysis, further research is warranted to provide any definitive conclusions, addressing the limitations of participant grouping (to have a more homogeneous group in terms of English proficiency) and task procedure (to have the control group given the posttest individually in a separate room, as was the think-aloud group).

Finally, pedagogical implications are addressed. It is recommended that meaning-inference instruction incorporate think-aloud strategy instruction as well. However, teachers and students need to be aware of false causality—simply using think-aloud does not necessarily lead to greater success in meaning-inference. The quality of think-aloud needs to be focused on during instruction in order to ensure that students' think-aloud actually reflects their metacognition and serves as a metacognitive strategy. Metacognitive strategies typically include planning, reasoning, monitoring, and evaluation of one's learning activity (Purpura, 1997). Without the existence of these characteristic components in learners' verbal reports, the effectiveness of think-aloud will be reduced. While the researchers explored the instructional implementation of think-aloud, they advise teachers and students to incorporate think-aloud in L2 meaning-inference with some caution. It is also recommended that teachers incorporate meaning-inference instruction only after knowing whether their students' English proficiency level is high enough to be conducive to correct inferences.

## Acknowledgements

We thank the students who participated in this study. Thanks also to Mary Theresa Seig, Kathy Ramos, and Kathleen Ulrey for their assistance in recruiting participants and Chaeyoung Kim Lee and Boyoung Roh for their assistance in data collection and analysis. Permission has been obtained to reproduce the contents in Appendix A from Ablex Publishing and Appendix B from Elsevier.

*Megumi Hamada* teaches at Ball State University. Her specialty is second language acquisition. *Chaehee Park* teaches in South Korea. His specialty is second language spelling and reading.

## References

- Baumann, J. F., Seifert-Kessell, N., & Jones, L. A. (1992). Effect of think-aloud instruction on elementary students' comprehension monitoring abilities. *Journal of Reading Behavior, 24*, 143-172.
- Bensoussan, M., & Laufer, B. (1984). Lexical guessing in context in EFL reading comprehension. *Journal of Research in Reading, 7*, 15-32.
- Bereiter, C., & Bird, M. (1985). Use of thinking aloud in identification and teaching of reading comprehension strategies. *Cognition and Instruction, 2*, 131-156.
- Berne, J. (2004). Think-aloud protocol and adult learners. *Adult Basic Education, 14*, 153-173.
- Block, E. L. (1992). See how they read: Comprehension monitoring of L1 and L2 readers. *TESOL Quarterly, 26*, 319-343.
- Bowles, M. A. (2008). Task type and reactivity of verbal reports in SLA: A first look at a L2 task other than reading. *Studies in Second Language Acquisition, 30*, 359-387.
- Bowles, M. A. (2010). *The think-aloud controversy in second language research*. New York: Routledge.
- Bowles, M. A., & Leow, R. P. (2005). Reactivity and type of verbal report in SLA research methodology: Expanding the scope of investigation. *Studies in Second Language Acquisition, 27*, 415-440.
- Carrell, P. L. (1989). Metacognitive awareness and second language reading. *The Modern Language Journal, 73*, 121-134.
- Casanave, C. P. (1988). Comprehension monitoring in ESL reading: A neglected essential. *TESOL Quarterly, 22*, 283-302.
- Chern, C.-L. (1993). Chinese students' word-solving strategies in reading in English. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 67-85). Norwood, NJ: Ablex.
- Egi, T. (2008). Investigating stimulated recall as a cognitive measure: Reactivity and verbal reports in SLA research methodology. *Language Awareness, 17*, 212-228.
- Ellis, R. (2001). Investigating form-focused instruction. *Language Learning, 51*, 1-46.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: MIT Press.
- Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231-235). Hillsdale, NJ: Erlbaum.
- Goo, J. (2010). Working memory and reactivity. *Language Learning, 60*, 712-752.



- Haastrup, K. (1991). *Lexical inferencing procedures or talking about words: Receptive procedures in foreign language learning with special reference to English*. Tübingen, Germany: Gunter Narr.
- Hamada, M., & Koda, K. (2011). Similarity and difference in learning L2 word-form. *System*, 39, 500-509.
- Haynes, M. (1993). Patterns and perils of guessing in second language reading. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 46-64). Norwood, NJ: Ablex.
- Haynes, M., & Carr, T. H. (1990). Writing system background and second language reading: A component skills analysis of English reading by native speaker-readers of Chinese. In T. H. Carr & B. A. Levy (Eds.), *Reading and its development: Component skills approaches* (pp. 375-421). San Diego, CA: Academic Press.
- Huckin, T., & Bloch, J. (1993). Strategies for inferring word meaning in context. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 153-178). Norwood, NJ: Ablex.
- Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language: A review. *Studies in Second Language Acquisition*, 21, 181-193.
- Jourdenais, R. (2001). Cognition, instruction and protocol analysis. In P. Robinson (Ed.), *Attention and awareness in foreign language learning* (pp. 183-216). Honolulu: University of Hawai'i, Second Language Teaching and Curriculum Center.
- Kaivanpanah, S., & Alavi, S. M. (2008). The role of linguistic knowledge in word-meaning inferencing. *System*, 36, 172-195.
- Kelly, P. (1990). Guessing: No substitute for systematic learning of lexis. *System*, 18, 199-207.
- Kirsner, K. (1974). Modality differences in recognition memory for words and their attributes. *The Journal of Experimental Psychology*, 102, 579-584.
- Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. Cambridge: Cambridge University Press.
- Kucan, L., & Beck, I. L. (1997). Thinking aloud and reading comprehension research: Inquiry, instruction, and social interaction. *Review of Educational Research*, 67, 271-299.
- Laing, S. P., & Kamhi, A. G. (2002). The use of think-aloud protocols to compare inferencing abilities in average and below-average readers. *Journal of Learning Disabilities*, 35, 436-447.
- Leow, R. P., & Morgan-Short, K. (2004). To think aloud or not to think aloud: The issue of reactivity in SLA research methodology. *Studies in Second Language Acquisition*, 26, 35-57.

- Liu, N., & Nation, I. S. P. (1985). Factors affecting guessing vocabulary in context, *RELC Journal*, 16, 33-42.
- Loxterman, J. A., Beck, I. L., & McKeown, M. G. (1994). The effects of thinking aloud during reading on students' comprehension of more or less coherent text. *Reading Research Quarterly*, 29, 353-367.
- Maria, K., & Hathaway, K. (1993). Using think alouds with teachers to develop awareness of reading strategies. *Journal of Reading*, 37, 12-18.
- McKeown, R. G., & Gentilucci, J. L. (2007). Think-aloud strategy: Metacognitive development and monitoring comprehension in the middle school second-language classroom. *Journal of Adolescent & Adult Literacy*, 51, 136-147.
- Nassaji, H. (2003). L2 vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *TESOL Quarterly*, 37, 645-670.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Oster, L. (2001). Using the think-aloud for reading instruction. *The Reading Teacher*, 55, 64-69.
- Paribakht, S., & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition: An introspective study of lexical inferencing. *Studies in Second Language Acquisition*, 21, 195-224.
- Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of the constructively responsive reading*. Hillsdale, NJ: Erlbaum.
- Purpura, J. (1997). An analysis of the relationships between test takers' cognitive and metacognitive strategy use and second language test performance. *Language Learning*, 47, 289-325.
- Rossomondo, A. E. (2007). The role of lexical temporal indicators and text interaction format in the incidental acquisition of the Spanish future tense. *Studies in Second Language Acquisition*, 29, 39-66.
- Sachs, R., & Polio, C. (2007). Learners' uses of two types of written feedback on a L2 writing revision task. *Studies in Second Language Acquisition*, 29, 67-100.
- Sachs, R., & Suh, B. R. (2007). Textually enhanced recasts, leaner awareness, and L2 outcomes in synchronous computer-mediated interaction. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 197-228). Oxford: Oxford University Press.
- Sanz, C., Lin, H.-J., Lado, B., Bowden, H. W., & Stafford, C. A. (2009). Concurrent verbalizations, pedagogical conditions, and reactivity: Two call studies. *Language Learning*, 59, 33-71.

- Sternberg, R. J. (1987). Most vocabulary is learned from context. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 89-105). Hillsdale, NJ: Erlbaum.
- Walker, B. J. (2005). Thinking aloud: Struggling readers often require more than a model. *The Reading Teacher*, 58, 688-692.
- Yoshida, M. (2008). Think-aloud protocols and type of reading task: The issue of reactivity in L2 reading research. In M. Bowles, R. Foote, S. Perpiñán, & R. Bhatt (Eds.), *Selected proceedings of the 2007 Second Language Research Forum* (pp. 109-209). Somerville, MA: Cascadilla Proceedings Project.

## Appendix A

### *Passage A—When a Young Bird Leaves the Nest*

Like people, young birds go through a difficult transition when it's time to strike out on their own. The fledgling must be (1) glurked while learning to feed itself. It must be protected while learning to fly. In some species, fledglings must even be (2) moxed by their parents during their first autumn migration.

In most cases, a young bird (3) tidly returns once it leaves the nest. But there are some (4) padons. The youth of certain kinds of woodpeckers, wrens and swallows fly back to the nest to sleep. Similarly, some eagles and large hawks (5) firk home for weeks to feed until they learn how to catch their own (6) pum.

When it comes to (7) snerdling, however, few fledglings need any lessons. Fifty years ago, a German scientist names J. Grohmann raised some young pigeons in narrow tubes that prevented them from moving their (8) lurds. At the same time he allowed another group of pigeons of the same age to be raised by their (9) medlons in a nest in the normal way, exercising their wings vigorously. When the two groups of pigeons were mature enough, Grohamm took them out into the open and tossed them into the air. Surprisingly, the pigeons raised in the tubes flew away as strongly as the ones that had been unrestrained in the nest. Grohamm thus proved that the instinctive (10) grumility to fly develops in young birds with or without the opportunity to practice.

## Appendix B

### Passage B—Folk Objects

The relationships that objects have with their human creators and owners are recognizable. Object forms show human characteristics, for example, chairs are (1) noked as having legs, lamps as having necks, and clocks as having faces. Some individuals (2) ricate with objects as though they were people. They give them names, talk to them, and decorate or “dress” them. In American culture, for example, cars are regularly named or personalized with special license plates or paintings. They may be praised for good performance or cursed for bad. Some (3) beek consider the purchase of new mats, covers, or ornaments as buying “gifts” for their cars. So, humans express their own ideas and (4) hoakings through objects and see them as reflections of themselves.

Objects can be used for display (5) clibes of their human creators and owners. They may serve as symbols for social class, (6) kede, or ethnicity. A (7) deany example of object display can be found in front of houses. It has been (8) taded that in Utah, one can find driveways lined with wheels, and gates built from commercial objects. Although mailboxes must follow official standards of measurement, owners (9) soanalyze them. The mailboxes are converted into symbols of personal, occupational, or regional identity. Cowboys and horses (10) ficed from steel are put on the tops of mailboxes. The bottom is built from milkcans and wheels. Many mailboxes have iron chains built into supports, and bent to form initials or abstract shapes. By using objects, humans display their characteristics within what they believe to be a more uniform culture.

## Appendix C

### Passage A Answer Key

Pseudoword	Original word (1 point)	Native speaker answer other than the original word (1 point)	Synonym of the original word or a native speaker answer (1 point)	Meaning is partially correct (0.5 point)
(1) glurked	fed	aided/assisted, guided, given food, protected, helped	prevented 보살핌을 받다	raised (기르다)

(2) moxed	accompanied	forced, led, carried, helped along, supervised, watched over, guided, helped, coaxed, urged, assisted	instruct (지도하다, 가르치다) to follow (따라가다), to manage	depend on (의지하다)
(3) tidly	rarely	seldom, never, quickly, hardly ever, almost never		
(4) padons	exceptions	problems, issues, cases, situations, birds that return to the nest even though they aren't expected to	except pattern tendency (경향)	types of behavior (행동양식) habit (습성), rule (규칙)
(5) firik	return	come, stay, visit, leave		
(6) pum	prey	food	meal	
(7) snerdling	flying	instincts		independence (독립) moving (이동하다)
(8) lurds	wings	bodies		
(9) medlons	parents	family, mother(s)		
(10) grumlity	ability	instinct, urge, desire, need	skills nature nature (성질)	power, will (힘, 의지) behavior, action (행동) sensitivity (감각) characteristic (특성, 특징) way to survive (생존방법)

## Appendix D

## Passage B Answer Key

Pseudoword	Original word (1 point)	Native speaker answer other than the original word (1 point)	Synonym of the original word or a native speaker answer (1 point)	Meaning is partially correct (0.5 point)
(1) noked	described	said to have, deemed, thought as, known, defined, recognized, tables, talked about	symbolize recognize represent (나타내어진다)	similar, seems (같다, 비슷하다) call as~ (~라고 부르다) seen (~처럼 보인다)
(2) ricate	interact	spend time, talk, identify, bond, relate		consider (간주하다), think, recognize express (나타내다)
(3) beek	individuals	owners, enthusiasts, people, supporters, drivers	car lover (차애호가) 사람	consumer, buyer (구매자, 소비자) mania (매니아)
(4) hoakings	feelings	identifications, tastes/styles, opinions, personalities, desires, likes	신념, 성격, 모토, 기호 thinking (생각), 가치관	image (이미지)
(5) clibes	purposes	characteristics, places, a part of, a piece of a whole, facets, pieces, boards, symbols, trophies	조각, 부분	thinking, emotion, opinion, idea 생각, 의견, 문화

(6) kede	occupation	age, religion, race, gender, membership, identity, culture	개인적 특성, personal characteristic, identity	tradition, rich (부), 명성(fame), status (지위), economic status, honor (명예), social status (계급), authority (권위), money (돈)
(7) deany	contemporary	good, common, great, perfect, prime	looking easily (쉽게 찾을 수 있는)	representative (대표적인), most (대부분의) 분명한, 적절한, 확실한, 알맞은, 전형적인, similar (비슷한), specific (구체적인)
(8) taded	reported	noticed, seen, said, stated, found, recorded, shown	showed, indicate, find, 나타내다	발생한, exist, 적용되다
(9) soanalize	personalize	decorate, dress them up, specialize	hope to make it specific (~특별하게 하고 싶다) identify, distinguish (개성화하다)	change
(10) ficed	made	cut, created, shaped	built, carve, form	꾸미다, 문양되다

# Reviews

***The Developing Teacher.* Duncan Foord. Peaslake, Surrey, UK: Delta, 2009. 96 pp.**

*Reviewed by*  
John Bankier  
Soka University

*The Developing Teacher* is the second book in Delta Publishing's teacher development series. It consists of a large number of activities designed to help teachers reflect on and improve their professional practice. The activities come from a combination of the author's conversations with other teachers and his own lengthy experience, as well as having been informed by writing from inside and outside the field of TESOL. Foord attempts to make the activities universal and appropriate for all, regardless of teaching experience or long-term commitment to the profession. The book is divided into three main parts. The first defines development and explains why it is important, detailing the rationale behind the book as a whole. Teacher development is identified as consisting of both "becoming better at what we do in the classroom" and developing "our career in teaching in the broadest sense." In development, teachers "reflect on what has happened, . . . make new things happen and . . . react to changes around [them]" (p. 8). The second part consists of the activities for professional development themselves. Foord separates the activities into five "circles of development," proceeding from those which can be done alone to activities which require the teacher to develop in the context of the profession as a whole. Though the majority of activities in the five sections are short, the final part of the book contains suggestions for three long-term projects.

Each section on the five circles of development is organized in the same way. It begins with an introduction, the first part of which is a checklist on which readers can rate from 0 to 5 things they have or have not done. For in-



stance, the first checklist includes “worked on my time management skills” and “kept a teaching diary” (p. 20). Readers can then identify those things they would like to do. The introduction then links the particular section to RISE (Recognition, Imposition, Self-improvement, Enjoyment), which Foord claims are the main motivations teachers have for taking part in development. For example, the section of the book dedicated to “You and Your Students” focuses on S and E, as teachers try to improve themselves as well as gain more enjoyment from the job. The end of each section’s introduction consists of a summary of key activities and a rationale for each. After the introduction, the activities are laid out in an easy-to-understand format including a rationale, key steps in the activity itself, and a template where necessary. Each section contains between 11 and 20 activities. The activities are written in an accessible style, with frequent humor, such as Step One asking readers to make a cup of tea, get two biscuits, and sit down on their sofa before beginning Step Two.

The first circle is “You.” All of the activities require only reflection or self-observation. Some, such as “Tea and Two Biscuits” (p. 31) mentioned above, comprise self-reflection on particular classes, sources of stress, or bigger questions such as “Why did I become a teacher?” Other activities asking the reader to self-observe at work and reflect do not require much preparation or materials beyond a piece of paper. For instance, in “RISE and Shine” (p. 22), readers are asked to assign motivation to instances in their careers, such as having done a diploma course partly out of a desire for recognition (R in RISE). Readers are to return to their observations in 3 months’ time and assess if their motivation has changed. This is typical of many of the activities in the book; though the concept presented is simple, actually doing the activity can be surprisingly revealing as following the steps compels one to focus on particular areas of development.

The second circle is “You and Your Students,” where the activities require interaction between the reader and students. Many of these activities seek to bridge the gap between teacher and learners, and introduce ways in which a teacher can involve learners in his or her teacher development. An example is “Face Down, Face Up” (p. 41). To get over the problem of students being unwilling to offer criticism or provide feedback on the approaches taken by their teachers, they are provided with a collection of possible comments, such as “I wish the teacher would correct my mistakes more.” Students only need to agree or disagree; this is followed by a discussion and feedback stage. Activities in this section allow the teacher to elicit comments on something new they may have tried in the classroom, such as a new method of teaching,

a different attitude to classroom roles, or simply a different arrangement of seating.

The third circle, "You and Your Colleagues," asks the reader to interact with colleagues to either assist them or be assisted by them in development. Some activities are geared towards improving communication between teachers, while others require a colleague to team-teach a class or observe one's teaching. These activities are generally short-term, but two are about mentoring new teachers for a period of weeks or months.

"You and Your School," the fourth circle, consists of activities about the work environment. These range from the micro, such as small annoyances with classroom design, to the macro, such as the curriculum, materials, or exams set by the school administration. The activities frequently require the reader to reflect on new ways to approach interactions between school and colleagues. One example is "Six Ways of Talking" (p. 69). Teachers role-play typical inter-staff conversations, such as the director of studies meeting with a teacher to discuss student complaints, and reflect on how they communicated.

The fifth and final section is "You and Your Profession." Here the reader is encouraged to interact across the field as a whole in various ways. At the simplest level, this might consist of attending a conference or posting on a blog, but builds up to ways for teachers to get involved in publishing both in academic journals and as contributors to classroom texts.

In the course of reviewing the book, I tried out a number of the activities, both on my own and in my classes. In much the same way that different ways of learning appeal to different people, I found that some activities suited me well, whereas others did not, such as "What Colour of Teacher are you?" (p. 29). However, the wide variety of activities means it is possible to find some that look interesting and rewarding, such as "First Things First" (p. 24) which helped me a great deal with my prioritizing of tasks. As the majority of activities can be completed in 20 or 30 minutes, it is easy to make time for them. Indeed, there is even an activity to improve time management for this very reason. In short, *The Developing Teacher* is an accessible, clearly written, and entertaining text. I would recommend it to all teachers, regardless of their experience, motivations, or professional goals.

***Can Do Statements in Language Education in Japan and Beyond: Applications of the CEFR.* Maria Gabriela Schmidt, Naoyuki Naganuma, Fergus O'Dwyer, Alexander Imig, and Kazumi Sakai (Eds.). Tokyo: Asahi Press, 2010. x + 315 pp.**

Reviewed by

Carol Begg

Kanda University of International Studies

Those of us with experience in the European language industry as educators, course designers, examiners, or trainers will surely have encountered the Common European Framework of Reference for Languages (CEFR). Developed by the Council of Europe (COE) through extensive research and consultation in the 1990s, the CEFR represents an action-orientated model of progressive language competency and the COE's vision for a communicative-focused *plurilingual* society. Accompanying the six common reference levels (A1—C2; denoting Basic, Independent, and Proficient users), the CEFR is supported by *can do statements*, positive descriptors of learners' practical competences at each stage (see 1.1 Imig & O'Dwyer, p. 3 for details on how these statements were developed). The framework itself has been translated into 38 languages, with a further two in translation (COE, n.d.), demonstrating its continuing rise in popularity and increasing circle of influence beyond Europe. The concept that led to the creation of *Can Do Statements in Language Education in Japan and Beyond: Applications of the CEFR* was Schmidt's (p. iii) proposal at the JALT2009 conference that a collection of can do-related texts be compiled. For better or worse, an assemblage of similarly themed papers is what this book is.

The CEFR and the can do statements, originally devised for European languages and now available in multiple languages, will be better known to teachers of languages other than English. It is perhaps because of this, as well as to embody the concept of plurilingualism, that the editors decided to include papers in English, Japanese, and German. This choice, while understandable, nevertheless makes the volume, in part, inaccessible to most readers. While the preface appears written in all three languages, this equality is not maintained. Most papers in the collection, including those written in Japanese or German, have an English abstract provided. Only one of the papers written in Japanese (2.3 Majima) is accompanied by a full English version which, much like the rest, would have benefited from consistent

editing. Lack of consistency is the most noticeable issue. It would have been preferable to have all the abstracts provided in each language and for the papers to all follow the same outlining structure.

The introductory chapter historically situates the CEFR, European Language Portfolio (ELP), and can do statements, giving those readers unfamiliar with the CEFR family an overview of the associated influences and issues, as well as references for further reading. Both papers are concise and informative; the continued defining of the CEFR in Chapter 2 “The CEFR and Can Do Statements” is redundant. That said, the papers included in this and subsequent chapters do cover a broad range of can do applications and research. This certainly fulfils the collection’s aim of overcoming the lack of awareness perceived by the editors to exist among educators on how to utilise the CEFR and can do statements.

The CEFR has been described as possibly being “the most relevant and controversial document in the [language education] field in the twenty-first century . . .” (Figueras, 2012, p. 477). In their paper, “1.2 An Overview of the International Influences of the CEFR,” Parmenter and Byram give an introduction to the criticisms levelled at the CEFR, highlighting its *foreignness* outside of Europe. However, it falls to later papers to explain these shortcomings. Green (“2.2 Conflicting Purposes in the Use of Can Do Statements in Language Education”) explores the friction sometimes created by the behavioural and proficiency objectives at the heart of the CEFR scales. This conflict grounds some of the Japan-centric issues voiced in later chapters: those of assessing the communicative ability (the capacity to carry out a language task) or mastery of the language necessary for *proficiency* to be evident. Horiguchi, Harada, Imoto, and Atobe (“3.5 The Implementation of a Japanese Version of the ‘ELP-Junior Version’ in Keio”), echo this issue, documenting the resistance of some teachers to adopting what they discern as a purely communicative approach assessment. The study also highlights the pedagogical limitations of a can do or task-based approach to language learning in a context with little opportunity for interaction beyond the classroom. In the classroom, textbooks and functional approaches to instruction, such as those described in “3.2 *Kannbeschreibungen im Deutschunterricht an Japanischen Universitäten?*” (outlining the creation of a conversational German textbook) and “3.3 Using German Can Do Statements as a Model for Other Languages Such as Russian and Chinese: A Special Project” (looking at the use of chunking as an approach to language instruction), support the notion of the CEFR’s foreignness but also offer insight into its effectiveness once this initial unfamiliarity is overcome.

Although the miscellany of papers can be difficult to navigate, and the plethora of voices can be grating, the variety is also a great strength of the collection. By including a range of contexts, practitioners, and researchers, the editors have succeeded in offering models for research or application in almost every Japanese institution or context. Small-scale action research such as Sato's 1-year survey of three freshman classes' reaction to the use of can do statements, "4.3 Using Can Do Statements to Promote Reflective Learning," is comprehensible and described in enough detail to be replicated. Collett and Sullivan's project, "4.2 Considering the Use of Can Do Statements to Develop Learners' Self-Regulative and Metacognitive Strategies," is broader in scale, but still furnishes the reader with a model for similar studies. Collett and Sullivan give more insight into the theories of learner autonomy and learning strategies, linking these to the can do statements before synthesising them into the *Study Progress Sheets* used in the project. Papers like theirs not only give credibility to the use of can do statements from an academic standpoint, but demonstrate how, through practitioner-led action research, effective learner resources can be developed. Larger scale studies, such as Horiguchi et al. ("3.5 The Implementation of a Japanese Version of the 'ELP-Junior version' in Keio") and Sargent and Winward-Stuart ("4.7 Implementation of a Can Do Based Syllabus in an *eikaiwa*"), are two of the strongest papers in this collection. Horiguchi et al. describe an institution-wide attempt to implement the ELP and delineate the obstacles encountered, including the general foreignness of the instrument and learners' lack of experience with self-assessment. The 7-year project undertaken by Sargent and Winward-Stuart at an *eikaiwa* is by far the largest described in this collection. It explores the complexities of the Japanese education context in terms of stakeholder expectations, the examination or assessment-centric nature of language learning, language learning beliefs, and the business framework in the private conversation and language schools. Their project shows a commendable focus on learner needs and the necessity for change both for the precarious *eikaiwa* industry and language education as a whole.

The Sargent and Winward-Stuart paper picks up on another sub-theme of the collection, that of the CEFR's relationship with examinations, clearly a vital quality in the Japanese context. Those readers already familiar with the CEFR will undoubtedly know its parallels with Europe-based exams (e.g., Cambridge Suite, BULATS, IELTS). However, for the sample of *eikaiwa* students in Sargent and Winward-Stuart's study, the correlations found between the CEFR-based levels and those of the largely Japan-based *Eiken* test were very interesting.

As a published volume, *Can Do Statements in Language Education in Japan and Beyond: Applications of the CEFR* can be unwieldy. As an assembly of papers, the sub-themes and unexpected insights contained could have been better presented if the book had followed a more consistent format. As a platform for research, however, the book is an amazing showcase of the contexts of language teaching in Japan and of the many talented practitioners and researchers working in the field. Like the can do statements themselves, this book needs to be adapted by the reader to suit his or her own needs. The Framework's ideological derivations and commentary on the evolution of language teaching are increasingly relevant in the globalised environment, making these papers timely and insightful. The collection certainly contributes to the literature on and the debate surrounding the CEFR.

## References

Council of Europe (COE). (n.d.). *Cadre*. Retrieved from [www.coe.int/t/dg4/linguistic/cadre\\_en.asp](http://www.coe.int/t/dg4/linguistic/cadre_en.asp)

Figueras, N. (2012). The impact of the CEFR. *ELT Journal*, 66, 477-485.

***English in Japan in the Era of Globalization*. Philip Seargeant (Ed.). London: Palgrave Macmillan, 2011. x + 208 pp.**

*Reviewed by*

Theodore Bonnah

Ritsumeikan University

At first glance, the average EFL teacher in Japan might not think that *English in Japan in the Era of Globalization* (hereinafter *EJEG*) is the most useful or relevant book for them. Although the first half of *EJEG* does explore various facets of English in the Japanese educational context, the book is not explicitly about ELT like a coursebook or teacher's manual, but is instead a collection of critical anthropological studies of English as it exists here. However, for reflective English teachers in Japan who have wondered about the historical and political forces which allow us to automatically become professional educators in another country, and which prize our linguistic birthright while never seeming to fully take it to heart, then *EJEG* is a welcome discovery. Although the book is replete with scholarship on how glo-

balizing forces have shaped how English is taught, learnt, manifested, and manipulated in Japan, it is quite engaging and easy to read, and would be useful for the native English-speaking educator trying to situate his or her identity and role in the ambiguous spaces of Japanese schools, institutions, and society in general.

*EJEG* begins with an Introduction by editor Philip Seargeant, followed by nine articles organized into two parts, entitled “English in the Education System” and “English in Society and Culture.” The Introduction alone is worth half the price; it gives the broad historical outlines of how English has come to hold its puzzling position in Japan, with a depiction of the contested and contradictory sociopolitical discourse that surrounds it today. Seargeant’s stated intent with *EJEG* is to examine English in Japan in terms of its relation to Japanese identity, the image and reality of the language itself, and the globalizing processes that promote and permeate how English is taught and used in Japan. This organization echoes Appadurai’s (1990) depiction of cultural products in the contested landscape of globalization, a notion that rings true to English speakers or teachers in Japan.

The five articles in Part I give impressions of English in Japanese education from various angles, from policymakers to teachers and students, as well as the average Japanese person in the street. The first essay, by Yamagami and Tollefson, tackles the discourse of English in education, especially the tension between *English-as-opportunity* and *English-as-threat*, which is engendered by opposing globalizing educational and nationalizing political forces. Next, Matsuda’s examination of student and teacher perceptions of English as global lingua franca is closest to research that JALT members may have read in its publications or heard at conferences, but goes further by asking and answering important questions about what English really means to Japanese people studying it. Stewart and Miyahara follow this with a look at the opposition of administrative reality and the globalization goals for a student-oriented curriculum in a Japanese university, which will resonate with anyone frustrated at teaching in higher education here. Fourth, Breckenridge and Erling discuss the essentialist yet alienating “native teacher ideal” of the JET Programme through three ALT narratives. This article looks at the alienation of ALTs from the teaching profession and how ALTs might be drawn toward and engaged in professional teaching associations. Rounding out the section, Kubota offers personalized narratives and policy views of the privileged status of English education and how this denies the experiences of other-language immigrants here. With one like me, who started as an ALT on the JET Programme, moved through various EFL positions, and

climbed my way into teaching at a university while doing my doctoral studies here in Japan, all these articles resonate and add a deeper understanding of the larger forces shaping the experiences I have had.

Part II presents four articles that step outside of the classroom to look at how English manifests itself in Japanese society at large. Starting off, Yano examines how “native-speaker syndrome” prevents the emergence of an indigenous Japanese tradition of English such as the ones in Singapore or the Philippines, instead leaving a heavy English footprint on the Japanese language. Next, Kamada uses analysis of discourse from mixed-parentage girls to depict how they form their own identity. Following that, Moody and Matsumoto examine how both behavioral characteristics as well as linguistic skills are expected of the idealized Japanese speaker of English in NHK “language entertainment” TV programs. Finally, Seargeant wraps the section up with a semiotic analysis of English used in signage in Japan, especially its various and sometimes opposing functional roles for foreigners and status symbolism for Japanese. The diverse topics covered in this section especially testify to the dual role of English in Japan in what Fairclough (2006) terms “globalization from above” and “globalization from below” (p. 37).

This book will likely be of particular value to those pursuing graduate degrees. Truthfully, some of the articles in *EJEG* are more qualitative than some readers might prefer, relating personal experiences as data, although this is mostly made up for by the depth of background research the writers of each article have done. Additionally, in true academic form, *EJEG* asks more questions than it answers, which educators used to teachers’ guides and texts with concrete steps and solutions may find less than satisfying. However, the book is thought-provoking and enlightening reading for English speakers and teachers who live in Japan, and I would recommend it to any fellow language teacher in Japan or anywhere else in Asia for that matter.

## References

- Appadurai, A. (1990). Disjuncture and difference in the global cultural economy. *Theory Culture Society*, 7, 295-310.
- Fairclough, N. (2006). *Language and gobalization*. New York: Routledge.



***Thinking About Language Teaching: Selected Articles 1982-2011.* Michael Swan. Oxford: Oxford University Press, 2012. xi + 230 pp.**

*Reviewed by*

Robert James Lowe

Rikkyo University

Michael Swan is well known among language teachers for his reference works, which are compulsory reading on the majority of TESOL certificate and diploma courses and usually essential for a few years afterwards. However, in the academic TESOL community he is at least as well known for his influential and sometimes provocative and contrarian articles about ELT methodology and practice. This book collects some of his more famous writings of this sort in the first 18 chapters, complete with some perspective-adding forewords from the author, as well as including a number of satirical pieces about the world of ELT in the final chapters (19-25).

Students of TESOL and applied linguistics will more than likely be aware of Swan's 1982 critique of the communicative approach, which appeared in *Applied Linguistics*, and led to a heated exchange between Swan and Henry Widdowson. Swan opined against the prevailing belief of the ELT community that the communicative approach was the most effective ELT methodology, arguing that the claimed benefits of function over form were weakly supported by theory and risked turning classrooms into places where students merely "did things" with language, rather than actually learn it. His two articles on the subject, along with this exchange, are reproduced in Chapters 1 and 2 of this book, and are contextualized with a foreword by Swan and some concluding thoughts from Widdowson. These articles provide an entertaining opening to the book and establish not only Swan's character (the closest thing to an *angry everyman* that the world of ELT has to offer), but also one of the recurring themes in Swan's writing, the "new toy" effect, which he first describes on page 7.

Swan argues that in the same way children only want to play with their new toy, discarding all the old ones, so too do teachers become overly enamoured of new methods and often dismiss older approaches as "discredited," regardless of how useful they may actually be. He has remained critical of what he sees as evidence-free trends in methodology, a position which can best be summed up with a quote from the book in which he states that "like

eighteenth-century doctors, we work largely by hunch, concealing our ignorance under a screen of pseudo-science and jargon" (p. 26). He returns to this theme when he discusses task-based learning in Chapter 8, *chunks* in Chapter 9, and methods in general in Chapter 16 (with other gentle sprinklings of cynicism found throughout the collection). It is easy to sympathise with Swan's position, but towards the end of the collection, readers may start to feel that the point has been sufficiently, if eloquently, made.

Inherent to all these criticisms is Swan's suspicion of communicative approaches, in which students "notice" gaps in their L2 knowledge and "discover" language (rather than being told about or taught it by a teacher). In particular, he argues that these approaches focus too heavily on the L2, and not enough on the L1. This point receives the greatest attention in Chapter 7, during which he discusses the connection between vocabulary acquisition and the influence of the student's L1 and is brought up again in Chapter 13, where he defends contrastive analysis as a useful resource for teachers in identifying learner problems.

It may seem as if Swan's approach is more critical than constructive, but this is not in fact the case. Alongside his critiques, he offers insights into how TESOL courses can best be organized (Chapter 5), and into the extent to, and manner in which, grammar can be effectively taught (Chapter 11). However, Swan's chief contribution, as evident in this collection, appears to be in the realm of clarifying ideas. He notes in Chapter 2 that the terminology and concepts of TESOL can be confusing and complex, and in response to this a number of articles included in this collection seek to make these ideas comprehensible. This is most evident in Chapter 12, in which he offers a summary of how language courses can best be constructed, and in Chapter 14, which contains an informative discussion on how grammatical structures can be more useful to students than the teaching of pragmatic rules. In the next chapter, he writes at some length about learning strategies, critically evaluating common reading strategies (such as guessing meaning from context) and makes suggestions about which ones are most likely to be effective. These, together with a few other articles in the book (see Chapters 5, 7, 15, and 17), offer informative views on ELT methods and practice, but are perhaps less memorable than the more sceptical pieces alongside which they appear.

Towards the end of the book, we are shown some of Swan's satirical writing about TESOL and the ELT industry. The first is a discussion of the benefits of teaching through "sensory deprivation," poking fun at faddish methodologies of the 1970s and 80s such as the Silent Way. These satirical

essays are amusing and diverting, but hidden beneath each is a veiled criticism of some aspect of ELT, be it the nebulous expansion of ELT terminology and theorizing (Chapter 24) or the pretentiousness of a great deal of ELT research (Chapter 20). Of most interest to those concerned with teaching in Japan is Swan's discussion in Chapter 25 ("Learning the Piano in Fantasia"), in which he discusses the problems facing the planet Fantasia, which is seeking to make all its children proficient in the piano as a means of interplanetary communication (having been adopted as an intergalactic lingua franca). To resolve this problem, he makes a number of suggestions, such as reducing the amount of music in the syllabuses, reducing the focus on music exams as the end of the course, and making sure that the high school piano tutors are actually able to play the piano themselves. This is, of course, an obvious parody of the problematic situation in contexts such as Japan. Swan's recommendations for how to tackle these problems are sensible, realistic, and practical, and as such have been proposed many times before and are unlikely to provide any new insights to those familiar with the field. The article does, however, collect and focus the criticisms of the Japanese language education system into four easily digestible pages and may help to raise awareness among teachers and ELT professionals.

Overall, this is a book that contains insights into teaching and methodology which are useful, but which may be overshadowed by the more famous and contrarian articles reproduced alongside them. Swan argues against much of the prevailing wisdom in ELT methodology, attempts to make it more comprehensible and generally applicable to ELT practitioners, and offers some of his own advice and solutions to the problems raised. Despite these practical applications, the book is perhaps best read as an assembly of the recollections, views, and thoughts of an influential figure in the world of TESOL. It does not provide a deep focus on any aspect of methodology or practice, nor does it seek to outline a teaching philosophy or overarching pedagogical system. It does, however, provide a number of well-written, entertaining, and sceptical insights into the world of ELT and some historical perspective on how the world of TESOL has come to be as it is.

***Towards an Understanding of Language Learner Self-Concept.***  
**Sarah Mercer. Springer, 2011. xii + 220 pp.**

*Reviewed by*

James Rock

Università Cattolica del Sacro Cuore di Milano, Italy

In recent decades, there has been a surge of interest in researching self-related constructs, motivated by the widely acknowledged belief amongst psychology and applied linguistics researchers that self-beliefs can heavily influence how a learner approaches a learning task. However, studies in second language acquisition that explicitly focus on self-concept are conspicuously absent. Hence, *Towards an Understanding of Language Learner Self-Concept* is an especially insightful and timely contribution.

Using a strong, interdisciplinary approach that builds on reflections from current research in both psychology and applied linguistics, Mercer focuses on how learners form their English as a foreign language (EFL) self-concept. The nonpositivist, domain-specific exploration of the construct is particularly welcome and represents a far cry from earlier quantitative investigations. Using a grounded theory approach, Mercer describes and reports on a 2-year, in-depth, longitudinal case study undertaken with an advanced level Austrian female student learning English as a foreign language. The study is supported by several other data sources and aims to provide some valuable insight into the complexity of self-concept and the kinds of dynamic processes involved in foreign language self-concept formation.

The book comprises seven chapters that examine the theoretical nature of the EFL self-concept and the principal factors that influence its development. Chapter 1 describes the research context of the study and provides an overview of the various chapters. Chapter 2 discusses the theoretical nature and structure of the self-concept. This is followed in Chapter 3 by a thorough analysis of both the case study and other data sources to ascertain how the EFL self-concept may function within a specific context. The next three chapters focus on various factors that may affect the formation of a learner's self-concept. Chapter 4 investigates the extent to which the self-concept may be dynamic, as well as the role of demographic factors, past achievements, feedback from significant others, and social comparisons. The influential Internal/External (I/E) frame of reference model (Marsh, 1986) is also described. A modified form of this model is subsequently used in Chapters 5 and 6 as a frame of reference for understanding some of the internal and external factors that

seem to influence the EFL self-concepts of the learners in Mercer's study. The final chapter discusses the findings of the study, considers some pedagogical applications, and suggests areas for further research.

As part of the *Educational Linguistics* book series, which promotes work that is transdisciplinary, contextualised, and innovative, Mercer's study certainly fills the bill. Of particular interest in Chapter 3 is her hypothesis concerning the theoretical nature of the self-concept construct. She proposes that although self-concept seems to be multifaceted, it is highly unlikely that it is hierarchically structured, as originally claimed in an influential self-concept model developed by Marsh and Shavelson (1985). On the contrary, Mercer's qualitative data suggests a complex interconnected set of relations among various self-concepts, which indicates the potential for variation, depending on the context and the individual. Mercer introduces an innovative 3D model of self-concept, which she claims can more accurately account for individual variation. The model consists of a molecular structure with differently sized spheres and connecting lines of various thicknesses. According to Mercer, this model successfully illustrates the interrelatedness of the self-concept network and indicates which self-concept is dominant for an individual at a particular time or in a particular context. She also highlights the model's flexibility and suggests that it can easily be adapted to accommodate different individuals in specific contexts.

Mercer should also be applauded for her discussion of the factors that may affect the formation of a learner's EFL self-concept. Using an adapted form of Marsh's (1986) I/E frame of reference model ensures a sound theoretical framework, while presenting the I/E factors in separate chapters makes logical sense and reduces the potential for ambiguity in the text. Of particular interest is her discussion of several internal factors, such as the types of internal comparisons that learners make across subjects, languages, skills, and even specific tasks. Her study indicates that comparisons of this type could effectively result in a student developing an increasingly positive self-concept in one foreign language, or language skill, to the detriment of another language or skill.

Another welcome aspect of the book is Mercer's examination of belief systems about language learning in general and beliefs about each specific language when learners evaluate their own self-concepts. Language learning beliefs have been shown to heavily influence strategy use, autonomy, and motivation levels. This factor was not included in Marsh's (1986) original I/E model and its addition by Mercer is a positive development. In Chapter 5, she interestingly suggests that a learner's EFL self-concept seems to be

influenced by general foreign language learning-related domain-specific sets of beliefs, rather than by language-specific learning beliefs. Several epistemological beliefs emerged from the data, as well as beliefs about the importance of experience in learning a language and the value of a stay abroad. On a slightly critical note, the role of authority in the classroom merited greater discussion in Chapter 6 given the prominent role that teachers play in learners' lives and the influence they can have on learners' attitudes, motivation, and the types of learning strategies they use. Nevertheless, this certainly does not detract from what is a successful attempt at investigating the various factors that affect self-concept formation.

Mercer repeatedly states throughout the book that her study is exploratory and is not intended to provide an all-encompassing comprehensive description of EFL self-concept. She does not deny that there may be other factors that affect self-concept formation and does not openly state any advantage of selecting internal or external frames of reference. She is simply interested in examining the dynamic nature of the EFL self-concept and exploring the way individual learners vary both in terms of the frames of reference they select when developing their EFL self-concept and how those factors interrelate in a specific context. In this respect, she has achieved her objective and has provided substantial food for thought for researchers in the field, particularly those who are intent on developing abstract, generalised models of self-concept. From a pedagogical perspective, the recommendations Mercer makes in Chapter 7 to help improve our learners' self-concepts are a step in the right direction. Nevertheless, it is reasonable to query how feasible it is for under-pressure educators to realistically undertake the types of in-depth exploratory studies on individual learners that are suggested by Mercer.

In conclusion, I would have no hesitation in highly recommending this book to language teachers and others who are interested in self-concept or any self-related constructs. Its far-reaching conclusions and regular suggestions of areas for future research should ensure that self-concept remains an intriguing construct for further exploration.

## References

- Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, 23, 129-149.
- Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist*, 20, 107-123.

***Innovating EFL Teaching in Asia*. Theron Muller, Steven Herder, John Adamson, and Philip Shigeo Brown (Eds.). Basingstoke, U.K.: Palgrave Macmillan, 2012. xxiv + 301 pp.**

*Reviewed by*  
Colin Skeates  
Seisen University

As stated in the Preface, *Innovating EFL Teaching in Asia* was created to serve several purposes, such as to provide a forum for Asia-based teachers to explain their context and how it affects their teaching and to provide further documentation of language teaching in an Asian EFL context. Indeed, it is the opinion of contributors to *Innovating* that context determines much of what is done in the EFL language classroom. Bringing together studies from nine different countries, the book contains 20 chapters. These chapters are organized into five areas, each prefaced with an introduction that explains how the chapters within the section contribute to an understanding of teaching in an Asian EFL context.

Part A consists of chapters from China, Thailand, Indonesia, and Korea. The purpose of this part is to help define what the Asian EFL teaching context is. For example, in Chapter 1, Xi Fang uses participant observation to chart the implications of the Chinese education ministry's implementation of communicative language teaching in one high school in southeast China. Based on ethnographic data, Fang is able to illustrate how changes to language education policy are perceived and dealt with by teachers expected to carry out those changes in the classroom. Gillian Palmer and Itje Chodidjah's chapter is on a teacher training and materials creation joint project between the British Council and the Nahdlatul Ulama Pesantren Islamic Boarding Schools (NUPIBS) in Indonesia. Taking a cascade approach to teacher training and materials creation, the authors outline successes and problems that arise in trying to create materials appropriate for both relatively rich urban schools and poor rural schools with few resources. In terms of teacher training, issues include how to deal with a lack of funds, differences in language proficiency among teachers and students, and how open teachers are to changing their teaching practices from traditional grammar translation methodology to more communicative teaching approaches approved by NUPIBS. Both of these chapters provide a perspective that is clearly outlined in practice.

In Part B, the authors focus on the Asian EFL teaching context by comparing teachers' and students' views of teaching, considering factors that can impact such views. For example, in Fumiko Murase's chapter, what university teachers think about their students' attitude towards learning is compared to what students think. Murase concludes that teachers often make assumptions about their students' capacity for learner autonomy without even enquiring what students think. The sixth chapter, by Chutagarn Raktham, utilizes Holliday's (2005) distinction between "big culture" (i.e., commonly held assumptions about a group of people) and "small culture" (i.e., the behaviours observed within a particular group), to better understand the influence of national culture on student behaviour. Raktham's findings indicate that small culture dictates what students do much more than might be expected. Of particular importance in all four chapters in this section is how the authors uncover and report on new insights that run counter to stereotypical views regarding their respective EFL contexts.

In the preface for Part C, Theron Muller makes the argument that historically EFL teachers have been limited to being testers of theory generated by ESL practitioners. As a result, many EFL teachers have ended up teaching to a theory as opposed to meeting the pedagogic needs of their students. The four chapters that form this part all provide examples of how context shapes methodology and theory building. For example, Japan-based teacher-researchers Muller and Mark de Boer illustrate that a common problem among EFL practitioners is how large class sizes inhibit communication in the L2. By reinterpreting previously published studies, the authors suggest that large class sizes can work against language learning, while classes with fewer students per teacher can actually mimic more natural conversation conventions. Another interesting chapter in this section is written by Honzhi Yang and Eva Bernat, who used Activity Theory to better understand how ready teachers in China are for drastic education reforms. Using a qualitative case study approach, Yang and Bernat undertook in-depth analysis to understand the interaction among beliefs, classroom practices, and contextual factors. They used Activity Theory to label behavior according to different activity categories. The results are practical suggestions as to what can be done to make teachers more receptive to government language education policy changes, such as the recommendation that teachers initially team-teach, and a reconsideration that those educated in Western cultures might not come equipped with the necessary skills to immediately take on the classroom challenges in Asia. Including chapters from Philip Shigeo Brown



and Wendy Lam, Part C illustrates the versatility that EFL teaching requires of teachers and the necessity for teachers to adapt their methods and approaches to the demands of different students and environments.

Part D contains four chapters on an area that has seen more research recently—how to teach young learners. The authors of each chapter explore how changes taking place in their context are affecting how instructors teach young learners. For example, Yasemin Kirkgöz (Chapter 13) utilizes questionnaires, classroom observation, and interviews of teachers to better understand how the implementation of a new curriculum in primary education in Turkey is affecting teachers. Kirkgöz found that teachers require more support and training on how to teach using songs and games, as well as training on how to use technology in the classroom. In Chapter 15, Hall, Yamazaki, Takahashi, and Ishigame report on a collaborative picture book project between Iwate University and a local Japanese primary school. Working with elementary school teachers who have little or no English ability, this study highlights why top-down decision making is not sufficient to change a system; teachers need to be a part of how such changes are to be implemented if the changes are to be at all successful. All chapters in this section will be of interest to those who teach young learners.

The fifth and final section, Part E, has to do with the teaching of writing in Asia. It consists of three chapters from three different countries, plus an epilogue that outlines how *Innovating* was created. Steven Herder and Peter Clement's chapter focuses on a 2-year extensive writing course at an all-girls' high school in Japan, emphasizing that, in an EFL context, learners require more support to develop as writers. Huahui Zhao's chapter provides interesting insight into the different roles that peer and teacher feedback play in students using and understanding error correction at a university in China. This chapter makes a strong case for peer interaction when corrective feedback is provided on student writing. The final chapter, by Toshio Hisaoka, attempts to portray how research from systemic functional linguistics can be used in the writing classroom. All three chapters again emphasize the pivotal role that context plays in teaching.

Ultimately, *Innovating EFL Teaching in Asia* is about how teaching is conducted within a particular context. I enjoyed reading the stories of how EFL instructors teach in vastly different contexts. The variety of research methodologies also offer excellent references for instructors who teach on PRE-SET or INSET programs, especially for those teacher trainers based in Asia who are looking for research on teaching contexts that are more relevant to

their student-teachers. In closing, the book does not set out to define what the EFL teaching context of Asia is, but rather, why diversity in teaching in an Asian EFL context needs to be better understood.

## **Reference**

Holliday, A. (2005). *The struggle to teach English as an international language*. Oxford: Oxford University Press.

# 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 Kyoyiku Gakkai*), invites empirical and theoretical research articles and research reports on second and foreign language teaching and learning in Japanese and Asian contexts. Submissions from Asian and other international contexts are accepted if applicable to language teaching in Japan. Areas of particular interest include but are not limited to the following:

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

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

## Guidelines

### Style

*JALT Journal* follows the *Publication Manual of the American Psychological Association*, 6th edition (available from APA Order Department, P.O. Box 2710, Hyattsville, MD 20784, USA; by email: <order@apa.org>; from the website: <www.apa.org/books.ordering.html>). Consult recent copies of *JALT Journal* or *TESOL Quarterly* for examples of documentation and references. A downloadable copy of the *JALT Journal* style sheet is also available on our website at <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 not be more than 10 pages in length. *Perspectives* submissions should be not more than 15 pages in length. *Point to Point* comments on previously published articles should not be more than 675 words in length, and *Reviews* should generally range from 500 to 1000 words. All submissions must be word processed in A4 or 8.5 x 11" format with line spacing set at 1.5 lines. **For refereed submissions, names and identifying references should appear only on the cover sheet.** Authors are responsible for the accuracy of references and reference citations.

### Submission Procedure

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

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

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

within 3 months. Each contributing author of published articles and Book Reviews will receive one complimentary copy of the *Journal* and a PDF of the article (Book Reviews are compiled together as one PDF). *JALT Journal* does not provide off-prints. Contributing authors have the option of ordering further copies of *JALT Journal* (contact JALT Central Office for price details).

### **Restrictions**

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

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

Please send submissions in these categories or general inquiries to:

**[jj-editor@jalt-publications.org](mailto:jj-editor@jalt-publications.org)**

Melodie Cook, *JALT Journal* Editor

### **Japanese-Language Manuscripts**

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

**[jj-editorj@jalt-publications.org](mailto:jj-editorj@jalt-publications.org)**

Ken Urano, *JALT Journal* Japanese-Language Editor

### **Reviews**

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

**[jj-reviews@jalt-publications.org](mailto:jj-reviews@jalt-publications.org)**

Greg Rouault, *JALT Journal* Reviews Editor

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

JALT Central Office

Urban Edge Building 5F

1-37-9 Taito, Taito-ku, Tokyo 110-0016, Japan

Tel.: 03-3837-1630; Fax: 03-3837-1631

(From overseas: Tel.: 81-3-3837-1630; Fax: 81-3-3837-1631)

Email: [jco@jalt.org](mailto:jco@jalt.org) URL: [www.jalt.org](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* (6th ed.)の定める方式に合わせて下さい。不明の場合は、JALT Journalの英語論文を参考にするか、日本語編集者までお問い合わせ下さい。なお、JALT Journalの読者は現場の教師が主なので、特殊な専門用語や統計的手法は、わかりやすく定義するか説明を加えるなどして下さい。原稿: 長さは、参考文献リストも含め18,000字(書評の場合は1,500字)以内です。A4の用紙に横書きで、1行40字、1ページ30行で印刷して下さい。手書きの原稿は受け付けません。

### 提出するもの:

以下の原稿を電子メールの添付書類、あるいは郵送でお送りください。

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

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

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

投稿原稿送り先またはお問い合わせ:

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

### JALT Journal 第35巻 第1号

2012年4月20日 印刷

2012年5月 1日 発行

編集人 メロディー・クック

発行人 ケビン・クレアリー

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