Japan Association for Language Teaching

Articles
5  Directives and Assessments in Japanese Native and Nonnative Conversation
   Yuri Hosoda
   Yumi Hato
53 Awareness of Teaching Through Action Research: Examples, Benefits, Limitations
   Jerry G. Gebhard

Research Forum
71 Score Reliability and Placement Testing
   Paul A. Westrick

Perspectives
95 Short-Term Overseas Study Programs: A Survey of Private Junior High Schools in Tokyo
   Herbert E. Brauer

Reviews
121 The Handbook of Bilingualism.
   (Tej K. Bhatia & William C. Ritchie, Eds.)
   Reviewed by Debra L. Simms
123 Teaching Listening in the Language Classroom.
   (Christine C. M. Goh)
   Reviewed by Justin Falkus
125 Managing Vocabulary Learning. (Paul Nation)
   Reviewed by Andy Maggs
   (Willy A. Renandya, Ed.)
   Reviewed by Paul Lyddon
129 The Power of Context in Language Teaching and Learning.
   (Jan Frodesen & Christine Holten, Eds.)
   Reviewed by Michael Kindler

ISSN 0287-2420
¥950
Contents

May 2005
Volume 27 • No. 1

3 In This Issue
4 From the Editors

Articles
5 Directives and Assessments in Japanese Native and Nonnative Conversation
   Yuri Hosoda
   Yumi Hato
53 Awareness of Teaching Through Action Research: Examples, Benefits, Limitations
   Jerry G. Gebhard

Research Forum
71 Score Reliability and Placement Testing
   Paul A. Westrick

Perspectives
95 Short-Term Overseas Study Programs: A Survey of Private Junior High Schools in Tokyo
   Herbert E. Brauer

Reviews
121 The Handbook of Bilingualism. (Tej K. Bhatia & William C. Ritchie, Eds.) Reviewed by Debra L. Simms
123 Teaching Listening in the Language Classroom. (Christine C. M. Goh) Reviewed by Justin Falkus
125 Managing Vocabulary Learning. (Paul Nation) Reviewed by Andy Maggs
129 The Power of Context in Language Teaching and Learning. (Jan Frodesen & Christine Holten, Eds.) Reviewed by Michael Kindler

JALT Journal Information
132 Information for Contributors (English and Japanese)
All materials in this publication are copyright ©2005 by their respective authors.
Japan Association for Language Teaching
A Nonprofit Organization

The Japan Association for Language Teaching (JALT) is a nonprofit professional organization dedicated to the improvement of language teaching and learning in Japan. It provides a forum for the exchange of new ideas and techniques and a means of keeping informed about developments in the rapidly changing field of second and foreign language education. Established in 1976, JALT serves an international membership of more than 3,000 language teachers. There are 37 JALT chapters in Japan, 15 special interest groups (SIGs), and two forming SIGs. JALT is one of the founders of PAC (Pan-Asian Consortium of Language Teaching Societies), which is an association of language teachers’ organizations in the Asian area. PAC holds regional conferences and exchanges information among its member organizations. JALT is the Japan affiliate of International TESOL (Teachers of English to Speakers of Other Languages) and is a branch of IATEFL (International Association of Teachers of English as a Foreign Language).

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

The JALT International Conference on Language Teaching and Learning and Educational Materials Exposition attracts between 1,500 and 2,000 participants annually and offers over 300 papers, workshops, colloquia, and poster sessions. Each JALT chapter holds local meetings and JALT’s SIGs provide information on specific concerns. 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 enrollment in the nearest chapter, copies of JALT publications, and reduced admission to JALT-sponsored events. JALT members can join as many SIGs as they wish for an annual fee of ¥1,500 per SIG. For information, contact the JALT Central Office or visit the JALT website at <http://www.jalt.org>.

JALT National Officers, 2005

President: ................. Steve Brown
Vice President: ............ Steve Nishida
Auditor: ..................... Tadashi Ishida
Director of Treasury: ...... Peter Wanner
Director of Records: ..... Mary Christianson
Director of Programs: ... Andrew Zitzmann
Director of Membership: Hugh Nicoll
Director of Public Relations: Sayoko Yamashita

Chapters
Akita, Chiba, Fukui, Fukuoka, Gifu, Gunma, Hamamatsu, Himeji, Hiroshima, Hokkaido, Ibaraki, Iwate, Kagawa, Kagoshima, Kanazawa, Kitakyushu, Kobe, Kumamoto, Kyoto, Matsuyama, Miyazaki, Nagasaki, Nagoya, Nara, Niigata, Okayama, Okinawa, Omiya, Osaka, Sendai, Shinshu, Shizuoka, Tochigi, Tokushima, Tokyo, Toyohashi, West Tokyo, Yamagata, Yamaguchi, Yokohama.

Special Interest Groups
Bilingualism; College and University Educators; Computer-Assisted Language Learning; Gender Awareness in Language Education; Global Issues in Language Education; Japanese as a Second Language; Junior and Senior High School Teaching; Learner Development; Material Writers; Other-Language Educators; Pragmatics; Professionalism, Administration, and Leadership in Education; Teacher Education; Teaching Children; Testing and Evaluation; Pronunciation (Forming SIG); Teaching Elderly Learners (Affiliate SIG).

JALT Central Office
Urban Edge Building, 5F 1-37-9 Taito, Taito-ku, Tokyo 110-0016, Japan
Tel.: 03-3837-1630; Fax: 03-3837-1631; E-mail: jalt@gol.com;
Website: <www.jalt.org>
In this Issue

Articles
The main section of this issue contains three articles. First, Yuri Hosoda, working within the framework of Conversation Analysis, examines the pragmatic competence displayed in sequences of directives and assessments occurring in casual conversation between native and nonnative speakers of Japanese. Next, Yumi Hato critically examines the Japanese Ministry of Education’s “Action Plan to ‘Cultivate Japanese with English Abilities,’” looking particularly at the goals that have been set and the instructional time allowed to achieve the goals. Finally, Jerry G. Gebhard discusses the benefits and limitations of action research, drawing examples from three action research projects conducted in Japan.

Research Forum
Paul A. Westrick reports on issues relating to score validity uncovered during the piloting of a commercially-produced English placement test.

Perspectives
Herbert E. Brauer reports on the results of a large-scale survey on overseas study programs at private junior high schools in Tokyo highlighting several innovative programs and activities.

Reviews
In this issue we have five book reviews. In the first one, Debra L. Simms reviews a comprehensive handbook which covers both basic and advanced concepts in bilingualism and its impact on society. Our next three reviews are of books from the SEAMEO Regional Language Centre in Singapore, a publisher known for practical books geared at classroom teachers. Justin Falkus reports on a book about how to plan listening tasks and lessons, while Andy Maggs examines a book that will give teachers a basic understanding of the principles behind successful vocabulary programs, and Paul Lyddon reports on an anthology of 14 papers gathered from the 37th SEAMEO RELC International Seminar, held in April 2002. Finally, Michael Kindler, reviews a book that explores the relationship between language and context.
From the Editors

It is May. In Japan a new school year has just gotten under way. That means new teachers, new students, and new classes. Here at the JALT Journal, with this first issue of volume 27, we have something new too—a new editorial staff. With the departure of Donna Tatsuki, Sayoko Yamashita, and Kate Allen, all after several years of dedicated work, a largely new group of editors have stepped forward and taken up the editorial baton. This issue is the first for Steve Cornwell as Editor, Deryn Verity as Associate Editor, Yoshinori Watanabe as Japanese-Language Editor, and Yuriko Kite as Book Reviews Editor. Also, Cynthia Quinn is joining the proofreading staff. A special welcome goes to all the new editors and special thanks goes to the outgoing editors and all the editorial board members, proofreaders, and other volunteers who help make the JALT Journal what it is.

Conference News

JALT2005: October 7th – 10th, Shizuoka Granship
Conference Theme: “Sharing Our Stories”

Life is stories. The things our students learn, and the things we learn become stories we build on and use to grow as teachers. That’s why Sharing Our Stories is the theme of the 31st annual JALT International Conference.

Stories can be the content learners think about as they, for example, read books or do listening tasks. Or they can be the conversations students have with each other. Or the ideas they write. At the conference, we teachers will share our experience, ideas, research, successes, and challenges. All of these are types of stories. And, of course, the sharing is the key.

In addition to the workshops, short papers, forums (colloquia), and poster sessions of previous conferences, this year’s conference will include a new discussion format in which facilitators will pose questions and encourage an active discussion for us all to share our stories.

Join us at Shizuoka in October. Listen to our stories. Share your own. Explore this website to find out what we are planning and how you can join in. Events, schedules, registration, and other conference information will be added over the coming months. See you at the Granship!
Articles

Directives and Assessments in Japanese Native and Nonnative Conversation

Yuri Hosoda
Kanagawa University

Various kinds of data and methodologies have been used to investigate non-native speakers’ (NNSs’) pragmatic competence. In the past decade, attempts have been made to describe NNSs’ pragmatic abilities in naturally occurring interaction using Conversation Analysis (CA) methodology. To date, there are an increasing number of CA studies that describe NNSs’ pragmatic competence in institutional settings, but only a few in noninstitutional settings. Using the framework of CA, this study examines NNSs’ pragmatic competence displayed in sequences of directives and assessments in casual native speaker (NS)-NNS conversation in Japanese. The analysis reveals that the pragmatic competence of the NNSs and NSs is constructed out of the detail of talk and other conduct in which the participants juxtapose multiple resources such as sequential organization, speech, body, and the surrounding environment to jointly shape the sequences of directives and assessments and establish mutual understanding in ongoing interaction.

これまで非母語話者の語用論的能力の検証が多くなされてきた。ここ10年間で特に盛んに行われているのが、発話相互行為を会話分析法(conversational analysis)によって分析した記述研究である。しかしながら、これまでに行われてきた研究の多くは学校の教室内などの発話の分析であり、自然な言語使用場面の会話の記述分析はあまり行われていない。本研究は、母語話者と非母語話者の日常会話を指示(directive)および評価(assessment)の連鎖(sequence)に焦点を置き、会話分析法を使って分析した。その結果、語用論的能力は会話参与者同士が言語のみならず、会話の構造、ジェスチャー、場面などさまざまな情報に言及し判断しながら、ともに理解を構築してゆく能力であるということが示された。

JALT Journal, Vol. 27, No. 1, May, 2005
To date, various data collection procedures have been used to investigate nonnative speakers’ (NNSs’) pragmatic competence. Studies that analyze spoken data use naturally occurring interaction, conversation tasks, and interviews as data sources, while those that examine written data utilize production questionnaires, multiple choice questionnaires, rating-scales questionnaires, and diaries as data sources. In addition, think-aloud protocols that are used to elicit participants’ verbalizations of thought processes may be used in combination with other tasks such as production questionnaires. Each data collection procedure has advantages and disadvantages (for further discussion of each procedure, see Kasper, 2000; Kasper & Dahl, 1991, Kasper & Rose, 2003). Furthermore, the methods for analyzing collected data also vary depending on researchers’ disciplines. For example, data from naturally occurring interaction have attracted a number of researchers that employ microanalytic approaches such as language socialization (e.g., Ohta, 2000a; 2000b; Poole, 1992), ethnographic microanalysis (e.g., Fiksdal, 1990), and conversation analysis (e.g., Firth, 1996; Mori, 2002).

This study focuses on one way of examining naturally occurring interaction. Specifically, I will demonstrate how video-recorded naturally occurring interaction between native speakers (NSs) and NNSs can be analyzed using the framework of Conversation Analysis (CA). Into my analysis, I bring multiple resources available for the speaker (i.e., sequential organization, speech, body movements, and surrounding environment). Before I discuss the methodology and data for the present study, I will introduce some previous studies that discuss the relationships among gesture, speech, and pragmatics.

**Gesture, Speech, and Pragmatics: Psycholinguistic Perspective**

Previously, the relationships among body movements, speech, and pragmatics have been studied extensively from a psychological/cognitive perspective (e.g., Kelly, Barr, Church, & Lynch, 1999; Kita, 2003; McNeil, 1985, 1992, 2003). These authors examined body movements in terms of how they correlate with the psychological functioning of a speaker’s mind during the speaking process. In this line of research, a speaker’s private and interpersonal psychological experiences are considered to be externalized as gestures.

Some authors focused on hand movements that co-occur with speech. For example, McNeil (1985, 1992, 2003) investigated hand movements that occur during the process of speaking in narrations and dyadic interaction.
tasks. McNeil found that gesture and speech arise from a common cognitive source, and that gesture and speech together represent thought in performing semantic and pragmatic functions. Furthermore, Kelly et al. (1999) examined pragmatic functions of hand movements that co-occur with speech. They conducted a series of experiments in which subjects watched role-play videos and answered some comprehension questions. In some of the role-plays, the actors had matching hand movements and speech, and in some others, their hand movements did not match their speech. Kelly et al. found that certain gestures such as deictic (i.e., pointing) and iconic gestures have an important impact on how people comprehend and remember pragmatic functions in communication. For instance, one of their experiments showed that pointing gestures themselves can be recognized as indirect requests. The authors also demonstrated that speech and gesture interactively accomplish the meaning of communicative acts.

On the other hand, Kita (2003) emphasized the importance of paying attention to body movements such as torso movements and gaze that might otherwise be considered nonspeech related. Through experiments on direction giving, Kita found that gestures, including hand gestures, torso movements, and gaze, facilitate speech at the conceptual level and showed the possibility that all these communicative means are interlinked in the cognitive process underlying the performance of a pragmatic function.

Thus, psycholinguistic studies of gestures have shown the close ties between body movements and speech in carrying out and understanding pragmatic functions. Conversation Analysis (CA) methodology, which I will employ in this paper, also emphasizes the significance of examining both gesture and speech in understanding pragmatics. However, as I will discuss below, the CA approach is different from the psycholinguistic approach in that in CA practitioners examine gesture and body movements from the participants’ perspectives; that is, rather than having researchers or observers interpret the speakers’ gestures, CA practitioners focus on looking closely at how speakers as well as listeners publicly display the relevance of gesture and speech in the course of accomplishing actions in ongoing interaction.

**CA Methodology and Pragmatics**

CA emerged in reaction to mainstream American sociologists whose discipline imposed a priori theorization of social phenomena. From its earliest stage, CA analysts strongly cautioned against a priori theorization and ad hoc analytical categorization of social interaction. Instead,
through repeated examination of tape recordings and transcripts of authentic spoken interaction, the analysts focused on revealing participants’ displayed orientation to making sense of interactions, an orientation that is embodied in the detail of their talk and other conduct.

In this methodology, in attempting to understand what actions the speaker is performing in the talk (i.e. “speech acts” such as requesting, apologizing, complaining, complimenting, and offering), we are cautioned not to isolate each utterance from the sequence but rather to observe the organization of courses of action realized jointly by and for the parties of interaction through sequences of turns. Thus, we need to observe and understand how and why a certain action has come to be deployed by the speaker at that very moment, how the recipient shows understanding (or nonunderstanding) of the deployment, and what consequence the deployment and understanding (or nonunderstanding) have in the subsequent interaction. Participants’ pragmatic competence to perform various actions is considered to be publicly displayed through the details of turns taken and yielded by the participants themselves in constructing the ongoing talk.

When we examine video-recorded naturally occurring interaction, we also need to consider the nonverbal signals sent, because in face-to-face communication “interactants are normally visible for one another” (Streek, 1993, p. 275) and interactants use not only language but other resources such as gaze, posture, and local environment to accomplish actions in face-to-face interaction (Goodwin, 2000, 2003). Traditionally, conversation analysts have relied on audio recording for collecting data of talk-in-interaction. Therefore, much of the earlier work has been done on telephone conversation, which automatically eliminates interactants’ orientation to each other’s nonverbal behavior. However, ever since the development of technology made video cameras available to everyone, some CA researchers have used video cameras to collect data. Those studies which used video-recorded interactions as data revealed how speech and nonverbal behavior such as gaze and body movements are oriented to by the interactants with respect to the structural organization of talk-in-interaction (e.g., Goodwin, 1981; 2000, 2003; Goodwin & Goodwin, 1986, 1987; Schegloff, 1984, 1998, 1999). It is important to note that even when interactants’ nonverbal behavior is examined, CA’s focus is still on the management of turn allocation and turn construction, and nonverbal features of speakers and hearers are examined only when the participants themselves have displayed their orientation to the nonverbal features in the interaction.
CA has originally been applied only to conversation between NSs and has generated a number of studies that revealed NSs’ pragmatic competence in ongoing interaction (e.g., Drew, 1984 on invitations; Goodwin & Goodwin, 1987; Pomerantz, 1978 on compliments; Pomerantz, 1984 on assessments; Schegloff, 1988 on complaints). Some CA practitioners have recently started applying CA to the analysis of talk by NNSs, especially those in institutional settings such as educational settings (e.g., Carroll, 2000; Markee, 2000; Mori, 2002) and business encounters (e.g., Firth, 1996; Wagner & Firth, 1997). However, there is still only a small number of CA studies that have described NNSs’ pragmatic competence in noninstitutional settings (e.g., Wong, 1994; 2000a; 2000b).

In the remainder of this paper, I will use the CA framework to demonstrate how NNSs and their NS interlocutors display their pragmatic competence in collaboratively establishing mutual understanding through sequences of directives and assessments.

Methods

The data analyzed for this study are based on two video-recorded NS –NNS conversations between friends speaking in Japanese involving four individuals in all. The three sets of video-recorded data (two in conversation 1 and one in conversation 2) come from a larger set of data that consists of 15 sets of NS-NS conversation and 15 sets of NS-NNS conversation. Both conversations introduced in this study were recorded for approximately 30 minutes.

Participants

Of the four participants, two were Americans and the other two were Japanese. All were male; their ages ranged from the late 20s to early 30s. As the base language for the conversations was Japanese in the data, the two Americans were considered to be NNSs and the two Japanese were considered to be NSs. The two NNS participants, Bill and Gregg (pseudonyms), were native speakers of English who were advanced speakers of Japanese. Both of them had completed courses in teaching Japanese conducted in Japanese and offered in the graduate school of an American university in Tokyo. Both Bill and Gregg had been living in Japan for six years. Bill was a high school English teacher and Gregg was a college English teacher. One of the NS participants, Koma (pseudonym), was an English teacher at a language school, and the other NS participant, Taro
(pseudonym), had just come back from a one-year working holiday program in Australia.2

Procedures

The researcher prepared for collecting the data by asking the American participants and their Japanese friends for permission to audio- and video record their conversations beforehand and then visiting the place where each dyadic conversation took place. The researcher arrived at the place before the conversation began in order to set up the camera. The conversation between Bill and Koma was recorded at Bill’s office in the high school in November, 1999, and the conversation between Gregg and Taro was recorded at Gregg’s residence in December, 1999. The participants were not given any topics prior to the conversation and were encouraged to talk freely about whatever they would normally talk about. They were, however, asked to remain within view of the camera. The camera was on a tripod placed about five feet away. In order to avoid the influence of the researcher on the setting or the participants (known as “reactivity,” Maxwell, 1996), the researcher was not present while the conversations were being recorded. As for the influence of the video camera, the presence of the camera may not affect the naturalness of the data to a great degree. As Goodwin (1981) notes, the participants in face-to-face conversation never interact as if they are not observed; they organize their talk and other conduct in terms of their coparticipants’ behavior. Moreover, the focus of CA studies is not the surface content of the interaction but the structural organization. Therefore, even when the participants talk about the recording procedure and equipment, the underlying structural organization of the interaction remains unaffected.

The two conversations were transcribed and analyzed by the researcher. Transcription and translation conventions were adapted from Jefferson (1984) and Maynard (1997) (see Appendix A for transcription conventions). In the discussion of the data below, I will focus my analysis on three directive sequences I found in the data.

Results and Discussion

In the data, the people who utter the directives are the NNSs of Japanese, Bill and Gregg. The three directives are all V (verb)-mite “Try V-ing” form of directives. The V-mite utterance is a directive in the sense that it is an utterance that a speaker uses to get someone else to do something
and it expresses what the speaker wants (Searle, 1979), and its grammatical form is the imperative. However, my concern in this paper is not to examine the directive utterance in isolation, but rather to look at all the resources the participants orient to, and examine what understanding it took for the speaker to produce the directive, how the directive is understood by the recipient, and what consequences the recipient’s understanding had in the subsequent turns.

**Overview of the Data**

The first directive sequence comes from the conversation between Bill and Koma. (See Appendix B for the full transcription of the three sequences.) As mentioned above, Bill and Koma conversed at Bill’s office. The directive occurs after Koma mentions the heaviness of his notebook computer. Bill utters “*Kore motte mite* (Try lifting this up)” and Koma lifts up Bill’s computer on Bill’s desk. The second directive sequence also comes from the conversation between Bill and Koma. The directive occurs when Koma mentions the track pad on his notebook computer. Bill produces “*Kore tsukatte mite* (Try using this)” and Koma tries using the track pad on Bill’s computer. The third directive sequence comes from the conversation between Gregg and Taro at Gregg’s residence. The directive is produced by Gregg when Taro shows Gregg a musical instrument Taro brought home from Australia. Gregg utters “*Yatte mite* (Try doing)” and Taro starts playing the instrument.

An initial look at the three directive sequences reveals that there is a pattern in all three: the Japanese-NS recipient of the directives immediately follows the directive uttered by the NNS. In addition, some kind of assessment followed each directive and the compliance with the directive. In the first directive sequence, after Koma lifts the computer up, assessments concerning the heaviness of the computer are made. In the second directive sequence, while Koma is using the track pad on Bill’s computer, Bill makes an assessment of the easiness of using the device and Koma makes a comment. In the third directive sequence, after Taro complies with the directive, some assessments concerning the sound of the instrument are made. In all three instances, the speaker and the recipient seem to have understood each other and a set of actions was carried out smoothly. How did they establish mutual understanding through the directive sequences? In order to address this question, I will look next at each instance in more detail.
Uttering the Directives

In all three cases, the conversation leading up to the interaction and the immediate setting in which it takes place are relevant to the directive, and the relevance is shown in the directives themselves. In the first directive sequence, at the beginning, Koma is complaining about his Mac computer, which he thinks is too heavy to be called a notebook computer. After listening to Koma’s complaints about the heaviness of his computer, Bill produces an utterance that directs Koma to lift up his computer which is behind Koma. In the second directive sequence, after Koma tells Bill that the track pad of Koma’s notebook computer is not easy to use, Bill tells Koma to try using the one in his computer. In the third directive sequence, prior to the directive, Taro and Gregg are talking about the instrument Taro has brought to Gregg’s residence, and just before Gregg tells Taro to play the instrument, Taro brings the instrument close to his mouth. Therefore, in all three cases, the speakers’ directives are dependent on what was already being talked about and the immediate settings in which the talking took place (i.e., the availability of the objects at the moment of the utterances). In other words, the NNSs’ directives show their competence in understanding what has been said thus far and the immediate environment in which the interaction is going on.

Recipients’ Understanding of the Directives

As discussed above, the directives produced by the NNSs arose out of the preceding interaction and the immediate setting, and this may have been one of the main factors of the recipients’ understanding of the directives. However, closer examination of the data revealed that the speakers not only drew upon these key elements, but also used their bodies to facilitate the recipients’ understanding of the directives. Therefore, in this section, in order to discuss exactly how the recipients reached the understanding of the directives, I will show the process of making directives and complying with directives in detail. Consider the first directive sequence with nonverbal features.

(1) [Bill-Koma:3:59-62] Bill=NNS; Koma=NS
1. Koma: _omo sugite nootobukku pasokon to wa ii gatai_=
   heavy too notebook computer as Top say difficult
2. =tokoro ga aru n da yo ne.=
place Nom exist NR Cop IP IP
“It has got an element to say that it is too heavy to say it’s a notebook computer.”

Figure 1. Bill starts preparing for a pointing gesture before he produces his directive.

Figure 2. By the time Bill finishes producing motte, Koma starts turning his body orientation to the referent.

Uh-huh this have try
“Uh-huh. Try lifting it up.”
Bill: |((turns gaze to the object))
Bill: |((raises his arm))
Bill: |((gazing and pointing at the object))

|------|------
Koma: |((turns his gaze to the object))
Koma: |((turns his body to the object))

   this how much exist IP
   “How much does this weigh?”
   |------------|-------------
Koma:|((turns his body completely facing the object))
Koma: |((moves both arms toward the object))

   |------------|-------------
   Bill:|((points at the object))|((brings his arm down)

   quite heavy
   “It’s quite heavy.”
   |-----------
Koma:|((leaning toward the object))

6. (.) ((Koma lifts the computer up))

In line 1, before Bill starts uttering a directive, he turns his gaze to the referent (i.e., his computer) while aligning with Koma’s complaint by producing u::n at the turn-initial position. Bill then starts raising his arm to point to the referent while he is making a sound “tt”: Bill starts preparing for a pointing gesture before he produces the directive (Figure 1). This finding is consistent with findings from previous studies on native speakers’ speech and gestures in that the preparation for the gesture slightly precedes the coexpressive utterances (e.g., McNeil, 1992; Schegloff, 1984). At the point Bill produces the directive, he executes the pointing gesture: he extends his arm by leaning slightly forward and actually points to the object. In this way, he makes clear what he refers to through gaze and pointing. In other words, Bill uses both gestures
and speech to establish a shared focus of reference. In response to Bill’s action, Koma makes some movements that show his understanding of what is being talked about. At the beginning of line 3, Koma’s eye gaze and his body orientation is directed at the speaker. However, at the point Bill finishes uttering *kore*, Koma turns his gaze to the referent, and by the time Bill finishes producing *motte*, Koma starts turning his body orientation to the referent (Figure 2). Thus, before Bill finishes uttering the directive, Koma’s gaze is secured at the referent, and his body begins to face the referent. This shows that their mutual orientation to the referent had been established by Bill’s gesture before his directive utterance was completed. In line 4, while asking the question *kore donogurai aru wake?* (“How much does this weigh?”) of Bill, Koma turns his body completely to the referent and moves both his arms toward the referent. Then in line 5, while Bill is answering Koma’s question *kekkou omoi* (“It’s quite heavy”), Koma leans forward toward the referent. Finally in line 6, Koma complies with the directive: he lifts the referent up.

Almost the same phenomena can be observed in the second directive sequence (2).

![Figure 3. While Koma is producing “track (padget) track board,” Bill starts looking at the object and starts raising his arm to point to the referent.](image.jpg)

2. Koma: >track (padget) track board.<
   “Track pad, track pad.”
   |--------

Bill: |((gazes at the object and raises his arm to point to the object))}
Figure 4. At the point Bill starts producing the directive, Koma begins turning his gaze to the referent.

   This use try very use easy |
   “Try using this. It’s very easy to use.” |
   |----------------|------|
   Bill: |((pointing)) |((withdraws his arm)) |
   |----------------|------------------|
   Koma: |((starts turning his gaze to the object)) |
   |((brings his both arms to the object)) |
   |((uses the referent with his hand)) |

In line 2, while Koma is producing his line, Bill starts looking at the object and starts raising his arm to point to the referent (i.e., the trackpad on his computer) (Figure 3). Therefore, at the point Bill starts producing the directive, Koma begins turning his gaze to the referent (Figure 4). When Bill actually produces the directive in line 3, he extends his arm to point to the object. In other words, he displays his orientation to the referent by a gaze and pointing gesture. By the time Bill finishes the directive, Koma’s hands are about to touch the referent. At the point Bill finishes the directive, Koma starts complying with the directive.

In the two examples above, Bill points to exactly the same object, his computer, which is on his desk along with various other things, and he refers to the referent using the same demonstrative pronoun kore (“this”)
instead of saying *bokuno konpyuta* (“my computer”), which names the referent more explicitly. Moreover, grammatically speaking, Bill’s use of *kore* is not appropriate. In Japanese, *kore* is used to refer to something close to the speaker, whereas *sore* is used to refer to something close to the addressee. In the examples above, Bill’s computer is placed behind Koma and it is close to Koma but not close to Bill. Therefore, *sore* is more appropriate in this situation. However, in spite of Bill’s ambiguous way of referring to the referent, Koma shows his understanding of Bill’s directives by lifting up the correct referent in (1) and using the correct referent in (2).¹⁴

Similar phenomena can be observed in the third directive sequence, which is taken from the second data set.

---

**Figure 5.** Before Gregg starts producing the directive, his gaze is already at the referent and Taro also turns his gaze to the referent by the end of the pause.

(3)

5. (1.8) ((Gregg begins to sit down next to Taro while looking at the instrument. Taro turns his gaze from Gregg to the instrument and brings it close to his mouth.))

6. Greg: °**yatte mi te.** °
   do    try
   “Try doing.”
   |------|---
Greg:    |((looks at the instrument))
7. Taro: bwhohh bwhoohh ((plays the instrument))

In line 5, during a 1.8-second pause, Gregg is looking at the referent (i.e., Taro’s instrument) Taro is holding. Therefore, before Gregg starts producing the directive, his gaze is already at the referent. Taro, who was at first gazing at Gregg, also turns his gaze to the referent by the end of the pause (Figure 5). As Gregg starts producing the directive, Taro brings the referent closer to his mouth, and as soon as Gregg finishes producing the directive, Taro starts complying with the directive: he starts playing the instrument. What is intriguing in this example is that Gregg’s directive in line 6 is uttered softly and it may not have been heard clearly by Taro. In addition, Gregg’s directive sentence “try doing” does not have an object and it does not specify what Gregg wants Taro to do. Nevertheless, Taro shows his understanding of Gregg’s directive by complying with the directive correctly.

In sum, in the three cases above, the following common phenomena were observed. First, before the speaker actually produced the directive, the speaker started gazing at the referent and prepared to point to the referent when there was a pointing gesture that accompanied the directive. In response to the speaker’s gaze and gestures, the addressee immediately turned his gaze and body orientation to the referent. Thus, the speakers and the addressees established a shared focus of attention through gaze and their bodies before the speakers actually finished producing the directives. Second, although the speakers’ ways of naming the referents were not explicit, the addressees displayed their understanding of the directives by complying with the directives immediately. The mutual understanding by the NSs and NNSs is also shown in the subsequent turns, in which they make some assessments on the referent, which I will address below.

**Making Assessments in Subsequent Turns**

The activity of doing assessments has been found to be the “central resource available to participants for organizing the perception and interpretation of what is being talked about” (Goodwin & Goodwin,
In the present data, the assessment sequences following the directives as well as compliance with the directives display the interactants’ interpretation and understanding of the previous talk. Consider (1)’ again with the focus on the latter part of the conversation.

(1)’

   Uh-huh this have try
   “Try lifting it up.”

   This how much exist IP
   “How much does this weigh?”

   Quite heavy
   “It’s quite heavy.”

6. (.) ((Koma lifts the computer up.))

7. Koma: ah kore: niitengo kiro gurai aru n zyanai?
   Oh this 2.5 kilogram about exist NR Cop:Tag
   “Oh, it weighs about 2.5 kg, doesn’t it?”

8. Bill: wakan nai kedo:
   Know NEG but
   “I don’t know but,”

   Heavy such thing have walk if work able Neg
   “It’s heavy. If I carry this around, I can’t work.”

10. Bill: shigoto dek(h)ih=
    work able:Cont
    “You can’t work,”
In this example, the participants, especially the recipient of the directive, make comments about a computer (i.e., the referent) in lines 7 to 11. It is about the heaviness of the computer and it is contingent on Bill’s directive in line 3 as well as the talk before the directive. By doing assessments, Koma, the recipient of the directive, once again displays his understanding of the directive and surrounding talk. Therefore, these lines of assessment show that Bill (NNS)’s, directive was successful in making the weight of the computer visible and salient. Furthermore, Koma’s action in line 11 demonstrates the inseparable nature of body and language. In line 11, when Koma says his arm muscles will be tight and strokes his arms (Figure 6), he is talking about the effect of the weight of the computer on the arms of a person carrying it. In other words, in making this comment, he is conceptualizing the interaction between the weight of the object and the human body through language.

In (2)’, although the assessment sequence is short, the interlocutors’ mutual understanding can be seen.
   This use try very use easy
   |---------------|------|
   Bill: |((pointing to the object))
   Bill: |((withdraws his arm))
   |--|---|----------------
   Koma: |((turns his gaze to the object))
   |((moves both his arms to the object))
   |((uses the referent with his hand))
   “Try using this. It’s very easy to use.”

4. Koma: |honto da::
   True Cop
   |---------------|
   |((using the object with his hand.))
   “You’re right.”

In line 3, Bill starts making an assessment on the referent (i.e., a track board on his computer) immediately after the directive. Koma, while using the referent with his hand, shows his understanding of Bill’s directive and comment by agreeing with Bill. His agreement token honto da:: (“you are right”) is strengthened with an emphatic stress at the end. In this example, the interlocutors are making an assessment on how easy it is for human hands to use the object. Therefore, again, they are expressing the interaction between the body and the object through language.

Finally, consider (3)’ again with a focus on the part in which the interactants make assessments.

(3)’

   Do try
   “Try doing.”

7. Taro:  bwhohh bwhoohh ((plays the instrument))

8.  chotto matte,
    a little wait
    “Wait a second.”
Following Gregg’s directive in line 6, Taro starts playing the instrument in line 7. Then in lines 12 and 14, Gregg makes a comment \textit{kakko ii} (“It’s cool”). In line 15, Taro, who had stopped playing in line 13 makes a request for confirmation \textit{ii oto suru desho;} (“It makes a nice sound, doesn’t it?”). This comment on the sound of the instrument shows Taro’s understanding that Gregg’s directive in line 6 is a request to make a sound with the instrument and has nothing to do with its shape or weight. As a request for confirmation, this utterance by Taro makes agreement or disagreement relevant. In line 16, Gregg agrees with Taro. Thus, in this example, the interlocutors again express what the referent does to human bodies, specifically, to the ears.

In sum, through making assessments, the interactants displayed their understanding of the directives as well as what was going on in the prior turns. In addition, in making assessments on the referents, the interlocutors virtually transformed the weight, usability, and sound of the referents to the human body and thereby expressed the inseparable nature of the human body, objects, and language.
Conclusion

This paper has demonstrated one way of looking at naturally occurring conversation to investigate NNSs’ (as well as NSs’) pragmatic competence and examined sequences of NNSs’ directives that are jointly shaped by the NNSs and their interlocutors. First, it was shown that the NNSs’ directives were dependent on previous talk and immediate settings, and the directives displayed the NNSs’ competence in understanding the previous talk and the immediate environment in which the interactions were going on. Second, it was found that the NNSs and the NSs established a shared focus of attention through gaze and body movement before the NNSs actually finished uttering the directives. Third, even when the NNSs’ ways of naming the referents were ambiguous, the NSs displayed their understanding of the directives by complying with the directives immediately. Accompanied gestures as well as sequential context helped to make the referent clear. This finding suggests that actions an utterance brings about are not coded solely by the linguistic forms of the utterance but rather, they are understood in the emerging sequential context in which bodily actions and language are crucially dependent on each other. Fourth, the NSs’ compliance with the directives showed the NNSs’ competence in carrying out the directive. Fifth, subsequent assessments also demonstrated the interactants’ understanding of directives as well as what was going on in the prior turns. Finally, in their assessments the participants expressed the inseparable nature of the human body, objects, and language. In the directive sequences, the participants themselves demonstrated the relevance of speech, body, emerging sequential contexts, and the surrounding environment in producing and understanding directives.

In earlier studies, NNS’s pragmatic competence was judged based on the NNSs’ speech. However, as this initial study shows, in naturally occurring interaction, “competent” interlocutors juxtapose multiple resources with sequential organization, body, surrounding environment, and so forth’ joining with speech to make meaning in interaction, and these nonlanguage resources may sometimes count for more than the linguistic appropriateness. Furthermore, the NNSs’ pragmatic competence is displayed in sequences of actions that are realized collaboratively by both the NNSs and their interlocutors. In the future, research in NNS pragmatics needs to consider the collaborative nature of interaction and the range of resources the interactants deploy.

Finally, some teaching implications: In naturally occurring interaction, as participants use not only language but also other resources to
accomplish actions and make meaning, second-language speakers may not focus their attention closely upon linguistic forms. Thus, the context of naturally occurring interaction may not be an ideal place for second-language speakers to practice appropriate linguistic forms. Therefore, language classrooms are needed to provide learners with opportunities to focus on linguistic forms. On the other hand, learning a second language only in classrooms, which lack interactional contexts or real objects may not be enough to develop the competence necessary to deal with naturally occurring interaction. In order to equip themselves with sufficient pragmatic competence in a second language, learners may need to practice and learn the language in both language classrooms and in settings with naturally occurring interaction.

Yuri Hosoda, Ed. D. is an assistant professor at Kanagawa University. Her research interests include Conversation Analysis and second language teaching and learning.

Notes

1. For details on the larger set of data, see Hosoda (2002, 2003).
2. Therefore, although in the data the participants spoke mostly in Japanese, the two NSs of Japanese were advanced speakers of English.
3. In transcribing nonverbal features, | | is used to indicate overlapping of nonverbal behavior; and ----- is used to indicate continuation of the nonverbal feature. Nonverbal features of interlocutors are shown in lines below each sentence.
4. Koma’s understanding may result from the fact that pointing gestures by themselves can be sufficient to constitute directives (Kelly et al., 1999).
5. Moreover, as second assessments were found to be upgraded (Pomerantz, 1984), in this example, the assessments are gradually upgraded. It starts from kekkou omoi (“It’s quite heavy”) by Bill (line 5), followed by omoi (“It’s heavy”) (line 9) and konna no moti arui tara sigoto dekinai (“If I carry this around, I can’t work”) by Koma, and finally ends with ude ga panpan ni natyau (“my arm {muscles} will become tight”) by Koma.

References


Appendix A

Abbreviations and Transcription Conventions

**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP</td>
<td>Interactional particle (e.g., ne, sa, no, yo, na)</td>
</tr>
<tr>
<td>Nom</td>
<td>Nominative (-ga)</td>
</tr>
<tr>
<td>Acc</td>
<td>Accusative (-o)</td>
</tr>
<tr>
<td>Gen</td>
<td>Genitive (-no)</td>
</tr>
<tr>
<td>Top</td>
<td>Topic marker (-wa)</td>
</tr>
<tr>
<td>Cont</td>
<td>Continuing (nonfinal) form</td>
</tr>
<tr>
<td>QT</td>
<td>Quotative particle (-to, -tte)</td>
</tr>
<tr>
<td>Q</td>
<td>Question marker (ka and its variants)</td>
</tr>
<tr>
<td>Cop</td>
<td>Copula (be-verb)</td>
</tr>
<tr>
<td>NR</td>
<td>Nominalizer (e.g., no, n)</td>
</tr>
<tr>
<td>Tag</td>
<td>Tag question</td>
</tr>
<tr>
<td>ONO</td>
<td>Onomatopoetic expressions</td>
</tr>
<tr>
<td>Neg</td>
<td>Negative</td>
</tr>
</tbody>
</table>

**Transcription Conventions**

- [ ] overlapping talk
- = latched utterances
- (0.0) timed pause (in seconds)
- (.) a short pause
- co:lon extension of the sound or syllable
- co::lon a more prolonged stretch
- . fall in intonation (final)
- , continuing intonation (nonfinal)
- ? rising intonation (final)
- ; intonation between a period and a comma
a rise stronger than a comma but weaker than a question mark

CAPITAL loud talk
underline emphasis

↑ sharp rise
↓ sharp fall
° ° passage of talk that is quieter than surrounding talk
< > passage of talk that is slower than surrounding talk
> < passage of talk that is faster than surrounding talk.

hh audible aspirations
*hh audible inhalations
(hh) laughter within a word
(( )) comment by the transcriber
( ) problematic hearing that the transcriber is not certain about
“ “ Idiomatic translation of Japanese utterances

In idiomatic translation,
{ } words or phrases which are not explicitly stated in the Japanese versions.

Appendix B

(1) [Bill-Koma:3:59-62] Bill=NNS; Koma=NS
1. Koma: "omo sugite nootobukku pasokon to wa ii gatai=
   heavy too notebook computer as Top say difficult

2. =tokoro ga aru n da yo ne.=
   place Nom exist NR Cop IP IP
   “There is a good reason to say that it is too heavy to call it a notebook computer.”

3. Bill: =u::n. tt kore motte mite.
   Uh-huh this have try
   “Uh-huh. Try lifting it up.”
   This how much exist IP
   “How much does this weigh?”

   Quite heavy
   “It’s quite heavy.”

6. (.)((Koma lifts the computer up.))

7. Koma: ah kore: niitengo kiro gurai aru n jyanai?
   Oh this 2.5 kilogram about exist NR Cop:Tag
   “Oh, it weighs about 2.5 kg, doesn’t it?”

8. Bill: wakan nai kedo:
   Know Neg but
   “I don’t know,”

   Heavy such thing have walk if work able Neg
   “It’s heavy. If I carry this around, I can’t work.”

10. Bill: sigoto dek(h)ih=-
    work able:Cont
    “You can’t work,”

11. Koma: =mou (0.4) ude ga panpan ni natyatte
    Well arm Nom ONO to become
    “Well, my arm (muscles) will become tight.”

(2) [Bill-Koma:5:100-101] Bill=NNS; Koma=NS
   ((Just prior to this segment, Koma mentions about a track board of his notebook computer.))
1. Bill: track (padget) desyo?
   Track (padget) Cop:Tag
   “Track (padget), right?”

2. Koma: >track (padget) track board.<

This use try very use easy
"Try using this. It’s very easy to use."

4. Koma: honto da:
   True Cop
   “You’re right.”

(3) [Gregg-Taro: 1:1-16] Gregg=NNS; Taro=NS

1. Taro: miru no hajimete;
   Look NR first time
   “Is it the first time you have seen it?”

2. Greg: hazimete. ah miru no wa jya nai kedo=
   First time oh look NR Top Cop Neg but
   “First time. Oh, it is not the first time I have
   seen it.”

3. Taro: =a: hon[tou.]
   Oh really
   “Oh, really.”

   First time Cop Neg IP
   “It is not the first time.”

5. (1.8) ((Gregg begins to sit down next to
   Taro while gazing at the instrument. Toro
   turns his gaze from Gregg to the instrument
   and brings it close to his mouth.))

   Do try
   “Try doing.”

7. Taro: bwhohh bwhoohh ((plays the instrument))

8. chotto matte,
   a little wait
   “Wait a second.”

9. (1.5)
10. bwhoooo[oooooooooooooooooooooooooooooooooooooooooo]
11. ooooooooooo]ooooooooooooooohh

12. Greg: [hhuh hhuh hhuh kakko ii nhhuh nhhuh nhhuh
           Style good
           “It’s cool.”

13. hhuh hhuh]

14. Greg: kakko ii=
           Style good
           “It’s cool.”

15. Taro: =ii oto suru desyo?:
           Good sound do Tag
           “It makes a good sound, doesn’t it?”

           Good sound do IP
           “It makes a good sound.”
Problems in Top-Down Goal Setting in Second Language Education:
A Case Study of the “Action Plan to Cultivate ‘Japanese with English Abilities’”

Yumi Hato
Fukui Prefectural University

This study critically examines the “Action Plan to Cultivate ‘Japanese with English Abilities,’” which the Japanese Ministry of Education has implemented as part of its reform of English education. Specifically, the paper appraises on the basis of up-to-date research findings on L2 learning the attainment goals the Ministry of Education through its Action Plan has set for junior and senior high school students. In this regard, it is shown that there is no empirical data to justify the Action Plan’s adoption of particular standardized tests into the definition of these goals, and that the goals defined in terms of English proficiency cannot be achieved within the available instructional time. This study thereby identifies flaws in the Action Plan which are caused mainly by the lack of input from those who are acquainted with the reality of L2 learning (i.e., teachers and researchers). The study also suggests possible ways for improving policy making and specifies the types of research that would be instrumental in formulating realistic and effective educational policies.

JALT Journal, Vol. 27, No. 1, May, 2005
33
As English has come to be used more extensively in various areas of professional and social life, there is a growing perception that people’s ability to communicate in English is closely related to both individual and national economic success. In this context, however, it is necessary for the authorities in charge of educational policies to base their decisions on the actual conditions under which students are learning the second language (L2) and not require results that cannot be achieved under those conditions. Focusing on this point, the present study critically examines the “Action Plan to Cultivate ‘Japanese with English Abilities’” (hereinafter, the Action Plan), which the Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT) has implemented to guide the ongoing reform in English education (Ministry of Education, 2003). Specifically, the study investigates the attainment goals which the Action Plan has established for junior and senior high school students, and identifies problems in those goals on the basis of up-to-date research findings on L2 learning. By exploring more effective ways of defining learning objectives, the study also makes some suggestions for future research and policy making.

The Action Plan and Its Evaluation

As part of the “Human Resource Strategy” included in the Government’s “Basic Policies for Economic and Fiscal Management and Structural Reform 2002” (Cabinet Office, 2002), MEXT officially announced the Action Plan on the last day of the 2002 fiscal year (March 31, 2003), and put it into effect the following day. The Action Plan is a five-year project. Its aim is to establish a system whereby the goals to cultivate Japanese with English abilities, defined in its opening chapter, can be attained at each level of formal schooling by the end of the 2007 fiscal year. The goals for senior and junior high school levels are specified in the following way:

- English language abilities required for all Japanese nationals
  “On graduating from junior high school and senior high school, graduates can communicate in English”
  • On graduation from a junior high school, students can conduct basic communication with regard to areas such as greetings, responses, or topics relating to daily life. (English-language abilities for graduates should be the third level of the Society for Testing English Proficiency (STEP) on average.)
On graduation from a senior high school, students can conduct normal communication with regard to topics, for example, relating to daily life. (English-language abilities for graduates should be the second level or the pre-second level of the Society for Testing English Proficiency (STEP) on average.) (Ministry of Education, 2003, Chapter 1)

Assuming these abilities can be acquired by the time of graduation from senior high school, the Action Plan defines the goal for the university level as ensuring that “graduates can use English in their work” (Ministry of Education, 2003, Chapter 1).

The subsequent part of the Action Plan is devoted to detailed specifications of “[how] to improve English education” (Ministry of Education, 2003, Chapter 2). The enumerated measures are concrete and comprehensive, concerning diverse aspects of L2 education such as teaching methods, teacher training, learner motivation, and high school and university entrance examinations. The Action Plan is, in fact, the first government-directed campaign launched for the specific purpose of improving the national standard of English education and thereby English proficiency.

As its implementation has come to involve a substantial number of personnel in local schools and boards of education, the Action Plan has been generating various views and opinions. Regrettably, however, those views and opinions are mostly based on individuals’ experience, intuition, or beliefs concerning L2 education. Tanabe (2004), for example, regards the Action Plan’s specificity in its descriptions for implementation as an advantage, and predicts that “it [the Plan] undoubtedly provides Japanese with opportunities to grow” (p. 7). Focusing on the same feature of the Action Plan, however, the Japan Society for the Improvement of Foreign Language Education (2003) denounces the central authorities’ excessively rigid control over teaching practice. The Society also argues that any educational plan that prioritizes a particular foreign language has a corrupting effect on students’ values and should therefore be retracted.

The exchange of such speculative opinions is unlikely to lead to valid conclusions that can readily be incorporated into future policy making. The Action Plan requires more rigorous research-based evaluation. Above all, its goals should be carefully examined since it is the ends rather than the means that is more likely to reveal the true nature of the ongoing reform in English education.
Problems in Examination-Oriented Definition of Goals

Prominent in the definition of the goals of the Action Plan is the decision to equate success in STEP, a test designed and implemented by an incorporated foundation outside MEXT, with the achievement of the target levels of proficiency. In offering a brief explanation of these goals, the Action Plan also refers to other external examinations when it states, “it is important for all Japanese people to aim at achieving a level of English commensurate with average world standards based on objective indicators such as STEP, TOEFL, and TOEIC” (Ministry of Education, 2003, Chapter 1).

Such an examination-oriented aspect of the Action Plan is open to criticism because MEXT has provided no rationale for incorporating ready-made external tests into its goals or for selecting particular tests from among others. Given the broad and cursory way of describing the target levels of proficiency, students and teachers may well assume success in those tests to be the main objective. Moreover, such officially recognized clear-cut criteria will necessarily be used to assess teachers, schools, and educational regions and will bring about competition regarding students’ test results. According to Morizumi (2003), some regional boards of education have already inquired into the number of students who have passed the third level of STEP in each junior high school under their jurisdiction.

With such inculcation of exam-based norms by the authorities, teaching and learning will necessarily be adjusted to the nominated tests, and the quality of those tests will therefore exert a great influence on the formation of Japanese learners’ L2 competence. If that is MEXT’s intention or if MEXT utilizes those external tests as a tool for disseminating its policies, MEXT should demonstrate the correlation between the competence which it assumes should be fostered in students and the competence those tests are to assess.

Japanese learners’ lack of communicative ability in English is often attributed to the overemphasis placed on grammatical knowledge and reading ability in high school and university entrance examinations. MEXT seemingly accepts such a view since the Action Plan specifies concrete ways of improving entrance examinations so that they will conduct “an appropriate evaluation of communication abilities” (Ministry of Education, 2003, Chapter 2, Section 4). Ironically, however, there is a strong possibility that the adoption of particular tests into the goals of the Action Plan will cause different problems of the same type. It is very
difficult, if not impossible, to create a test that does not distort the nature of communication abilities and their development, especially when the test has to be implemented on a large scale. In view of this, MEXT should at least disclose its reasons for judging that the adopted external tests offer a more “appropriate evaluation of communication abilities” than the entrance examinations in their current state.

Studies investigating the use of tests from a social or political perspective (e.g., McNamara, 1997; Shohamy, 2001; Spolsky, 1997) demonstrate that language tests can serve as powerful instruments for imposing authorities’ educational policies. On the other hand, psycholinguistic research into the influence of test quality on teaching and learning (e.g., Alderson & Wall, 1993) indicates the insufficiency of the washback effects of tests as an agent for educational reform, and suggests that concurrent improvement in teacher education and materials is necessary. In any case, if the authorities insist on exploiting the power of tests as a means for actualizing their policies, the prerequisite should be a careful adaptation of the test content to match the aim of those policies. In other words, the authorities should rigorously examine the possible washback effects of those tests to have a clear idea of the outcomes. Without assuming such responsibility, MEXT has hastily chosen readily available external tests. In response to such a hit-or-miss decision by the central authorities, English education in Japan is now becoming more and more attuned to those examinations.

Strangely, having adopted particular external tests into its goals, the Action Plan indicates the necessity of research to clarify the relation between the results in those tests and the degrees of proficiency required at each school level (Ministry of Education, 2003, Chapter 2, Section 7). This post hoc specification of necessary research suggests that MEXT’s first priority in forming the Action Plan was to respond to the Government’s “Basic Policies for Economic and Fiscal Management and Structural Reform 2002,” which explicitly required MEXT to establish by the end of the 2002 fiscal year an action plan for improving English education. Thus, the injudicious reliance on readily available external tests reflects the makeshift nature of the educational plan that was drawn up as part of economic policy.

Problems Relating to Feasibility of Goals

Ideally, success in examinations should be seen as a natural result brought about by the attainment of the target proficiency levels. However,
the attainment goals that the Action Plan defines in terms of proficiency also have a fundamental problem. Specifically, it has not been verified that the degrees of proficiency required at junior and senior high school levels can actually be achieved within the class hours allotted for English in each setting. As MEXT appears unconcerned about this matter, in the following section of this study, an attempt is made to evaluate the feasibility of those goals with regard to available instructional time.

**Absence of Context-Based Criteria for Assessing Feasibility**

From a psycholinguistic viewpoint, determining what should be learned (as in the Action Plan) fundamentally differs from prescribing what to teach (as in the Course of Study, the ordinance that MEXT issues and periodically revises). Teachers can strictly follow the Course of Study and present students with the listed items in the stipulated manner and sequence. However, as demonstrated by Corder (1967, 1978), Selinker (1972), and many other second language acquisition (SLA) researchers, L2 development is governed primarily by the students’ internal mechanisms and cannot be controlled by teaching. Specifically, students will not necessarily acquire the items of linguistic knowledge offered by the teacher (i.e., become able to use them in actual communication) at the time they are taught or in the order they are taught. In order to reasonably define attainment goals for particular students, therefore, empirical data are needed as a basis for predicting the level of proficiency that is attainable under the particular conditions and within the available time for learning.

However, current SLA research does not contribute much to L2 education in this area. A number of studies (e.g., Doughty & Williams, 1998; Krashen, 1985; Long, 1983; Schmidt, 1990; Swain, 1995) have investigated the effects of “negotiation of meaning,” “comprehensible input/output,” “noticing,” or “attention to form” on L2 development and have explored how opportunities for each of these can be enhanced. The results of such investigations can have implications in the search for ways to increase the rate of acquisition or decrease the time required to attain higher levels of proficiency. Current SLA research, however, does not offer much empirical data that can be referred to in predicting the time required for reaching specific levels of proficiency.

As one of the key issues to be addressed by future research, Lightbown (2001) raises the question of “how much time is required for most students to reach specified levels of proficiency in a variety of classroom settings” (p. 599). The time required to attain a particular level of pro-
ficiency varies considerably depending on many factors related to the similarity of the L2 to the language(s) already learned, the intensity of instruction and exposure, and the quality of the instruction. Context-specific research is therefore needed to determine the time required to achieve specific levels of proficiency in a particular setting. With regard to the bilingual immersion programs offered in Canada, extensive investigation of this kind (e.g., Swain, 1981; Turnbull, Lapkin, Hart, & Swain, 1998) has been undertaken, exerting an important influence on policy makers’ decisions. Some of the studies on minority language children (e.g., Collier & Thomas, 1989; Krashen, 2001) also deal with the time factor from this perspective. However, with regard to many other educational contexts, including Japanese secondary education, such an enterprise has not been attempted. Therefore, the goals of the Action Plan are not based on any empirical data. For the same reason, there are no context-based criteria for evaluating the feasibility of those goals objectively. The feasibility, therefore, can only be assessed indirectly by referring to the information obtained from outside the Japanese context.

Available Class Hours

As learners of English in Japan generally have very little exposure to English in their daily life, the feasibility of the goals of the Action Plan should basically be evaluated in terms of classroom instruction. The quality of instruction, together with learner motivation, will be enhanced to a greater or lesser degree if the measures prescribed in the Action Plan are steadily put into practice. However, the Action Plan does not make any reference to the quantity of instruction (i.e., allotted class hours). What is examined in this section, therefore, is whether improvement in teaching and learning alone can make the attainment of the goals possible.

Class hours allotted for each subject in junior high school are determined by MEXT and stipulated in the Course of Study. According to the latest Course of Study for Foreign Languages for Lower Secondary Schools (Ministry of Education, 1999a), the instructional time allotted for the subject “Foreign Language” is three 50-minute classes per week over three years with 35 weeks per year. The “Foreign Language” in this context can be seen as synonymous with English, as the new Course of Study stipulates that English should basically be chosen from among other languages. Thus, the total time for instruction a junior high school student receives per year is some 90 hours, which amounts to 270 hours over three years.
Class hours in senior high school cannot be calculated so simply. The latest Course of Study for Foreign Languages for Upper Secondary School (Ministry of Education, 1999b) places very few constraints on individual schools’ decisions concerning the selection of subjects and allocation of credits. As a result, the number of English classes a senior high school student attends varies considerably in accordance with the school’s educational objectives and particular conditions. In ordinary or academic high schools, a student typically earns 16 to 22 credits in English subjects upon graduation. One credit being equivalent to one 50-minute class per week for one year (i.e., 35 weeks), the accumulated time for instruction such a student receives in three years is approximately 470 to 650 hours. On the other hand, in some vocational or technical high schools, students earn fewer than 10 credits (some 290 hours) within the same time period, while in the high schools or departments of high schools that attach greater importance to English education, a student typically acquires around 25 credits (some 730 hours). The Action Plan sets an identical goal for all these students who are studying English under such varied conditions. This fact clearly demonstrates MEXT’s lack of concern about the time required to attain the goals it defines.

**Indirect Assessment of Feasibility**

The Action Plan requires total beginners of English to learn to “conduct basic communication with regard to areas such as greetings, responses, or topics relating to daily life” through 270 hours of instruction delivered in junior high school for three years, and then become able to “conduct normal communication with regard to topics, for example, relating to daily life” within the class hours accumulated in junior and senior high school for six years, for example, 740 to 920 hours for graduates of an academic high school (Ministry of Education, 2003, Chapter 1). The purpose of this section is to assess the feasibility of these goals by referring to the relevant information obtained from outside the Japanese context.

Cleveland, Mangone, and Adams (1960) examined various types of “training for overseasmanship” offered in the US at the time and created a table of “Time Requirements for Foreign Language Achievement” (pp. 250-251). According to the table, the minimum time required for American adult learners of Japanese with “average aptitude and positive motivation” to attain “sufficient proficiency in speaking to satisfy routine travel requirements” is either (a) nine months with a one-hour class per
day plus two hours of practice drills performed in a language laboratory with the aid of instructors, or (b) six months with a two-hour class per day plus four hours of drills. In the former case, the total class hours are 810, including the time spent on drills, while in the latter case, they amount to 1,080 hours.

The researchers also estimated that the minimum time required for the same type of learners to obtain “fluency and accuracy in speaking with sufficient vocabulary to meet any ordinary requirements” is either (a) 30 months with a one-hour class per day plus two hours of drills (total class hours: 2,700), or (b) 24 months with a two-hour class per day plus four hours of drills (total class hours: 4,320).

Apart from the linguistic distance between the L1 and L2, the context for which these estimates were made and the setting of the Action Plan appear to have very little in common. Furthermore, the ability “to satisfy routine travel requirements” does not correspond to the ability to “conduct basic communication with regard to areas...or topics relating to daily life,” the latter possibly allowing the learner to deal with a wider range of topics and situations. The ability “to meet any ordinary requirements” and to “conduct normal communication with regard to topics...relating to daily life” could be argued to be at similar levels, but no further explanation is available for either of these definitions. Nevertheless, a rough and ready comparison indicates that the Action Plan requires students to make tremendously rapid progress in terms of instructional time.

Another piece of relevant information is cited in Swain and Lapkin (1982):

> The basic level...should enable the student to acquire fundamental knowledge of the language, the ability to participate in simple conversation, the ability to read simple texts, and the ability to resume the study of [English] in later life. The middle level...is expected to enable the student to read newspapers and books of personal interest with occasional help from a dictionary, to understand radio and television, to participate adequately in conversation, and to function reasonably well in [an English]-speaking community after a few months' residence. (p. 14)

The definition of the “basic level” could well be an additional description of the target level of proficiency that the Action Plan specifies for junior high school students. The definition of the “middle level” could also account for what is meant by the ability to “conduct normal com-
munication with regard to topics...relating to daily life,” which is the goal the Action Plan sets for senior high school students. These definitions, however, have nothing to do with the Action Plan. They are the criteria presented for bilingual immersion programs in a document published by the Ontario Ministry of Education, and the target language was therefore French. According to the criteria, “the basic level is considered to be achievable through at least 1,200 hours of French instruction during the student’s school career,” and “the middle level is considered achievable through at least 2,100 hours of French instruction” (Swain & Lapkin, 1982, p. 14).

There are, of course, major differences between French immersion programs in Canada and EFL teaching in Japan. Depending on the context, therefore, one class may differ considerably in the significance it holds for learners’ L2 development. Nevertheless, it would be illogical to assume that the L2 ability requiring 1,200 hours for students of immersion programs can be acquired by Japanese students within 270 hours. It would also be unreasonable to expect Japanese students who spend no more than 740 to 920 hours in classes to attain a level of proficiency comparable to that of Canadian students who receive instruction for 2,100 hours. Thus, the Action Plan is inordinately optimistic in assuming that EFL teaching in Japan can accomplish such feats by performing two to four times better than Canadian immersion programs in terms of students’ achievement in the long-term.

A number of studies (e.g., Collier & Thomas, 1989; Cummins, 1979; Krashen, 2001) demonstrate that students in immersion programs and minority language children in majority language classrooms need several thousand hours to acquire an adequate command of the L2 and demonstrate age-appropriate performance in that language. Accordingly, in those contexts where students naturally have considerable exposure to the L2 outside the classroom, class hours are usually talked about in four-figure numbers, while in Japan, where such support for learning cannot be expected, no more than 1,000 hours are spared for English instruction through junior and senior high school education. It is therefore not sensible to discuss attainment goals for Japanese students on the same level as for students who are learning under such favorable conditions. The ability to conduct “normal communication” on a variety of topics related to “daily life” is, in fact, aimed at by large numbers of ESL and EFL learners in the world. However, given the limited class hours currently made available for Japanese students at junior and senior high schools, such an ability cannot be expected of them.
If the linguistic distance between the L1 and L2 is taken into account, the goals of the Action Plan appear even more impractical. In studies focusing on language transfer, German is often referred to as one of the languages that is relatively “close” to English (e.g., Odlin, 1989). However, in Germany, students at the middle-school level typically attend four hours of English classes per week and by the time they reach the age of 15 have received instruction amounting to 400 hours (Milton and Meara, 1998). In contrast, as mentioned earlier, their counterparts in Japan attend three 50-minute classes per week and have received 270 hours of instruction throughout their three years of junior high. Thus, despite the Action Plan’s claim that “it is important for all Japanese people to aim at achieving a level of English commensurate with average world standards” (Ministry of Education, 2003, Chapter 1), the prerequisite for attaining that goal has not been fulfilled.

Carroll (1975) examines the influence of various factors on the performance of learners of French as a second language and concludes that “the primary factor in the attainment of proficiency...is the amount of instructional time provided” (p. 276). Furthermore, in an article entitled “Are the British really bad at learning foreign languages?” (Milton & Meara, 1998), it was found that in terms of vocabulary growth per hour of tuition, British learners of French outperformed their counterparts studying English in other European countries. On the basis of these findings, the authors suggest that low levels of language performance in Britain may well be due to the comparatively small amount of time devoted to L2 instruction. In addition, there is substantial evidence (e.g., Stenett & Earl, 1984a, 1984b; Stern, 1985) to support the common sense view that accumulated instructional time closely correlates with students’ achievement in L2 proficiency. Without giving due consideration to such findings of empirical research, the Action Plan sets impractical goals for students and teachers that cannot be achieved within the time available.

**Washback Effects of Impractical Goals**

The impractical goals of the Action Plan will very likely generate cynicism among teachers as a result of their dissatisfaction or disappointment at the central authorities’ disregard for the reality of students and their learning conditions. Unable to come close to realizing the goals they are pressured to achieve, teachers will also experience unnecessary stress and frustration, and may even lose confidence in English education.
Furthermore, the unrealistic goals of the Action Plan cannot be expected to solve the major problem facing English education in Japanese secondary education. Japanese learners of English have a reputation for suffering from problems of communication in the L2, while having a good knowledge of grammar and vocabulary. This is often attributed to teaching methods, entrance examinations, or the shyness of students. From a psycholinguistic perspective, however, it can also be accounted for by the lack of classroom interaction through which students can obtain the “right” kinds of input and output. Students often supplement schoolwork with self-study or cramming at evening schools, but that also does not offer much opportunity for interaction. This deep-rooted problem cannot be overcome by setting unrealistic goals for students.

Research in SLA indicates the significance of interaction in L2 development from different points of view. Some researchers (e.g., Krashen, 1985; Long, 1996) argue that “comprehensible input” obtained though active interaction is the key variable for L2 acquisition, while others (e.g., Izumi, 2003; Swain, 1995) suggest that input alone is not sufficient for successful acquisition, and contend that “comprehensible output,” often gained through interaction, plays an important role in L2 development. The extent to which form-focused instruction (FFI) can contribute to SLA is still controversial (e.g., Krashen, 1992; Norris & Ortega, 2000). However, even the researchers who attach great importance to FFI (e.g., Ellis, 2001, 2002) presuppose a sufficient amount of interaction when they attempt to determine where and how FFI can come into play. FFI cannot therefore be expected to play a major role in L2 development, or to compensate for the lack of interaction or deficiencies in the quantity and quality of input or output. On the contrary, it is demonstrated by a number of studies (e.g., Schmidt, 1990; Skehan, 1998) that the explicit knowledge obtained through FFI can promote SLA (e.g., facilitate “noticing” or “restructuring”) only when learners use that knowledge in actual communication. In view of this, the key to improving English education in Japanese secondary education is to create a learning environment in which opportunities for interaction can more readily be provided for students.

In specifying measures to improve teaching methods, the Action Plan attaches great importance to the role of interaction, and stipulates the following:

...in English classes, instruction mainly based on grammar and translation or teacher-centered classes are not recommended.
Through the repetition of activities making use of English as a means of communication, the learning of vocabulary and grammar should be enhanced, and communication abilities in “listening,” “speaking,” “reading,” and “writing” should be fostered. (Ministry of Education, 2003, Chapter 2, Section 1)

However, in the definition of its goals, the Action Plan does not give due consideration to making it easier for teachers and students to move forward in that direction. For teachers who are pressured by time constraints regarding the highly ambitious goals to be achieved, it is not easy to take time away from FFI, through which they can easily obtain immediate tangible outcomes, and allocate that time to classroom interaction that has no explicit target knowledge to be learned and hence no instantaneous perceivable effects. The students who recognize the lack of time for achieving such lofty goals may make even more efforts in self-study or at cram schools. Interaction, however, cannot be expected to occur in self-study, and cram schools by their nature tend to pursue immediate visible effects of instruction even more eagerly than formal schools.

The goals of the Action Plan are therefore inconsistent with the direction it requires teaching methods to move toward. The time necessary for transmitting knowledge about linguistic forms could be reduced greatly through the efforts of the teacher and students concerned. However, communicative ability, or the ability to use those forms in actual communication, cannot be promoted as rapidly as the Action Plan expects. Thus, the Action Plan, while seemingly accepting the view that L2 acquisition is primarily promoted by the communicative use of that language, takes little account of the time naturally needed to acquire an L2. Owing to such inconsistency in educational policy, students and teachers are now puzzling over the distribution of time and effort they should divide between grammatical knowledge and communicative ability.

**Suggestions for Future Policy Making and Research**

Examinations conducted on a large scale cannot faithfully reflect the nature of communication abilities and their development. Examination-based attainment goals, therefore, inevitably have the effect of distorting the L2 competence of the learners who work toward them. In view of this, attainment goals defined in terms of L2 proficiency seem to have a better influence on teaching and learning. If those goals provide precise descriptions of what learners at the target level know and are able to
do and if such descriptions are based on updated knowledge about the course SLA typically follows and the empirical data on the characteristics of L2 development of the learners in that context, teachers can consult those descriptions in conceptualizing what their students need to learn and integrate ideas based on that data into instructional and assessment tasks. In other words, with such realistic and practical goals, teachers can adjust what they teach according to how students progress.

Language standards (also called bandscales, benchmarks, or curriculum frameworks) which have been developed around the world provide precise and comprehensive descriptions of the knowledge and ability which learners at different levels possess. For example, the Common European Framework of Reference for Languages (Council of Europe, 2001), which was established for English as well as other European languages, provides a large number of descriptors that can be used to determine learners’ L2 proficiency and needs. The Australian ESL Bandscales, which were also developed to guide teaching and learning, present detailed descriptions of ESL learners’ progress as they develop proficiency in each of the four skills of reading, writing, listening, and speaking (for more information on the language standards, see McKay, Coppari, Cumming, Graves, Lopriore, & Short, 2001). These language standards are the result of research conducted by a number of applied linguists and educational specialists on learners in context and repeated consultations with language teaching professionals. As most of the standards were completed fairly recently, there is not much empirical basis on which to argue for their values or benefits. Nevertheless, it is clear that the absence of such endeavors in the Japanese context is directly reflected in the ill-defined and impractical goals of the Action Plan, which appear to have little value in guiding teaching or assessing learning.

Attainment goals, if they are set, should be realistic and practical so that they can effectively guide teaching and learning without putting unnecessary strain on those responsible for achieving them. Indispensable in formulating such goals is a valid framework for estimating the time needed for students in that particular learning environment to reach various levels of proficiency, or the level of proficiency that they can attain in a given period of time. Therefore, context-specific research should also be devoted to this aspect of learning in order to obtain empirical data that can be used as a base for establishing such a framework. The SLA literature has not attached much importance to context-specific practical studies. In particular, longitudinal investigation focusing on L2
development of classroom learners in context has been neglected in the Japanese context. However, attainment goals defined at various educational levels (e.g., national, regional, school, and departmental levels) can only be justified when they are based on the objective data obtained through such empirical inquiries.

This study has highlighted the fact that sufficient time is an important requirement for L2 acquisition. It has also emphasized that in making educational policies, the time naturally needed to learn an L2 should not be underestimated. However, this should not be taken to imply that more time should be devoted to English instruction, which relates to a different topic that needs to be discussed in relation to other school subjects, and is therefore not the object of the present study. What this inquiry is meant to suggest is: (a) policy makers should make realistic decisions, giving due consideration to the restrictions imposed by the actual learning conditions, and (b) various types of context-specific research need to be conducted to find ways of minimizing the restrictions imposed by the context or maximizing learning opportunities for students.

For example, when students receive English instruction of no more than 270 hours in junior high school, their learning objectives, not to mention the target level of proficiency, should naturally differ from those of students who spend a longer period of time in classes. In other words, learning objectives for Japanese junior high school students should be defined in terms of how those 270 hours can most effectively be spent. With regard to what form English education at the junior high school level should take, some people argue that explicit grammar teaching should be totally abandoned while basic interpersonal communication skills should be intensively promoted through communicative L2 use. Others, however, contend that instruction in junior high schools should focus on providing students with a good grounding in grammar with which they can first develop reading and writing abilities and that students will acquire oral communication abilities if and when they come to have sufficient exposure to the L2. At present, however, none of these claims is valid since there is as yet no supporting evidence in terms of available class hours.

In addition to defining learning objectives, the issue of when to begin instruction and how to distribute instructional time should also be considered for the purpose of making optimal use of the limited amount of time available. In Japan, there has been a heated debate about the best time to begin English instruction (e.g., Otsu & Torikai, 2002; Tosu, 2002), especially because MEXT has been moving toward introducing English
education into the elementary school curriculum. There is, however, some evidence that the additional time gained by an early start may be less important than sustained exposure and instruction as students get older (Lapkin, Hart, & Harley, 1998; Turnbull et al., 1998). A traditional view of the distribution of instructional time is that language learning should be constant, with classes regularly provided over a long period of time. However, some research has shown that shorter periods of concentrated instruction at the primary level are more effective than “drip-feed” exposure (Lightbown & Spada, 1994, 1997). Again, all of these findings originate from outside Japan, and it is not clear to what extent they are applicable to the Japanese context. To enhance its accountability, MEXT, in cooperation with researchers in related fields, should explore the most effective ways of utilizing the limited instructional time available for Japanese students.

Conclusion

Strevens (1978) suggests that most of the origins of “relative failure” in L2 learning are outside the classroom and not the responsibility of individual students or teachers. Among those external causes, Strevens cites “unattainable objectives” and “insufficient time” (pp. 198-199). English education in Japan will certainly be enhanced to some extent if the measures prescribed in the Action Plan are steadily put into practice. Nevertheless, the Action Plan and other educational policies that may be introduced in the future will never succeed in eradicating the “relative failure” in English education in Japan as long as those policies are formulated on the basis of “unattainable objectives” and “insufficient time.” Accordingly, students and teachers will have to continue suffering undue blame for the failure, and the Japanese as a whole will continue to have an ingrained inferiority complex about their English abilities. Whether to lower the goals or to secure the necessary time to achieve the goals is an issue that should be addressed seriously, as is also the matter of maximizing available instructional time.

Acknowledgements

I am grateful to Prof. Michael Byram at University of Durham and Dr. Tony Ridgway at Queen’s University Belfast for their helpful information that allowed me to embark on this study. I would like to express my sincere gratitude to my colleague, Dr. Brian Morren as well as to the
editors and anonymous reviewers of the JALT Journal for their insightful comments and suggestions on earlier drafts of this paper. Any remaining errors or omissions are mine.

Yumi Hato teaches at Fukui Prefectural University. Her research interests include the possibilities and limits of focus-on-form(s), and time-related variables in L2 learning and teaching.

References


(Received February 17, 2004; accepted September 28, 2004)
Awareness of Teaching through Action Research: Examples, Benefits, Limitations

Jerry G. Gebhard
Indiana University of Pennsylvania

In this article I describe three action research projects conducted by teachers at Teachers College, Columbia University, Tokyo MA TESOL Program. I discuss the benefits of doing action research (how it helps us to make more informed teaching decisions; gain skills at posing and solving teaching problems; expand reflective skills; create a forum to discuss teaching issues and beliefs), as well as its limitations (a focus only on problems; a narrowly defined linear process). I then highlight other avenues to expand awareness, such as “exploring to see what happens” by trying the opposite, exploring “what we actually do” as opposed to “what we think we do,” considering “what we believe in relation to what we do,” and exploring to “gain emotional clarity.”

There are a variety of activities teachers can use to develop their teaching beliefs and practices. For example, Bailey, Curtis, and Nunan (2001), Gebhard (1992, in press), Gebhard and Oprandy (1999), and Richards and Farrell (2005), discuss how teachers can develop their teaching through the use of case studies, peer observation,

JALT Journal, Vol. 27, No. 1, May, 2005
self-observation with video, team-teaching, mentoring, coaching, teaching portfolios, learning a second language, talking with supervisors, teaching journals, and action research. This article focuses on one of these teacher development activities—action research.

At the most basic level, action research is a problem-posing cyclic process, through which teachers can identify, investigate, and try to solve problems in their teaching. The cyclic process begins with a search for a teaching problem, often based on intuition and informal observations. A next step is to learn more about the problem. For example, the teacher might videotape classroom interaction or ask students to complete a short questionnaire. After investigating the problem, the teacher can reflect on what has been learned, as well as guess what might happen if changes are implemented. Next, the teacher plans changes that aim at solving the problem, then implements the plan, observes what happens, reflects on the consequences, plans the next actions, and continues the cycle.

However, as Crookes (1993) and van Lier (1993) point out, action research is more than this. Rather than just trying to solve classroom problems, teachers can work through a process of problem posing to explore aspects of teaching not only within the classroom, but also in the school and larger community that might affect what happens in the classroom. In addition, as Burns (1996) and Crookes (1993) point out, the cyclic process is enhanced when teachers have chances for collaboration with others, such as colleagues, administrators, and parents, as they work through the process.

Those who have published their action research projects on EFL or ESL teaching (for example, Cornwell, 2001; Curtis, 2001; Sitler & Tezel, 1999; Stewart, 2001; Thorne & Wang, 1996) or focus attention on the process and benefits of doing action research (Burns, 1996, 1997; Burns & Hood, 1997, Farrell, 2001; Gebhard, 1999; Kemmis & McTaggart, 1982; Wallace, 1998) all directly or indirectly indicate the following: By using action research, we can gain skill at posing problems in our teaching, as well as work at solving the problems, and this process can make us more aware of our teaching. Likewise, action research can expand our reflective skills and provide a forum through which we can discuss teaching issues, pose problems, and address teaching beliefs and practices.

However, as I shall bring to light in more detail, action research focuses noticeably on problems in our teaching and uses a rather narrowly defined linear process, and this may limit the kind of awareness we can generate from just doing action research. As such, in addition to doing
action research, I encourage teachers to look beyond the problem-posing process of action research, and make use of other avenues to teaching awareness.

With this introductory statement in mind, I shall (a) describe three action research projects done by teachers enrolled in the Teachers College, Columbia University, Tokyo MA TESOL program, (b) discuss the benefits of using action research, and (c) highlight how action research possibly places limitations on our awareness, as well as show how we, as teachers, can go beyond action research by exploring different avenues to expand our awareness.

**Action Research: Three Teachers’ Problem-Posing Processes**

**The Context of the Action Research Projects**

Teachers studying in a course titled Observation, offered in the Teachers College MA TESOL program in Tokyo, did the three action research projects described. The MA TESOL program accepts mostly experienced teachers who will continue teaching while also taking its courses. The idea is for them to use their teaching settings as places where they can apply what they are learning in the program, and this idea is a central part of the Observation course.

In the Observation course seminar the teachers read and talk about the cyclic process of doing action research, including how to identify and pose problems, do a preliminary investigation, reflect, plan an action, implement the plan, observe the outcome, reflect again, and plan the next action. We also discuss the social implications of action research, such as how the cyclic problem-posing process can reach beyond the classroom and into the school and larger community, as well as the need to collaborate with others while doing action research. In this regard, students in the observation course form collaborative-support groups. With the backing of fellow-members of the support group, each teacher is asked to work through the cyclic action research process, using his or her own teaching setting as the context for the project.

Teachers in the Observation course have generated a number of action research projects, as the partial list of topics given below shows. To further illustrate the action research projects, following this list I provide descriptions of three of the teachers’ projects. Keep in mind that I do not consider any one project to be more valuable than another; each was selected simply to illustrate different teaching settings and uses of action research with them.
Example Action Research Topics

- Getting students to speak
- Increasing students’ fluency
- Lowering student anxiety
- Using praise to motivate students

- Making instruction in English more comprehensible
- Motivating students to read English for pleasure
- Improving students’ comprehension of the teacher’s questions in English

- Getting students to read for meaning
- Changing disruptive students’ behavior
- Decreasing the amount of spoken Japanese in the classroom
- Decreasing teacher talk and increasing student talk
- Making grammar rules comprehensible to students

An Action Research Project: How Can I Get Students to Get the Main Idea?

Sachiko Kimura teaches English at a high school where there is a lot of emphasis on preparing students to pass university entrance exams. The emphasis is mostly on reading and grammar, and the school administration asks teachers to use Grammar-Translation methodology in their teaching. Through her use of Grammar-Translation, Sachiko believed that many of the students had trouble comprehending her lessons. For example, she observed that most of the students could not tell her, even in Japanese, the main idea of the reading passage she had just translated for them.

Through consultation with her Teachers College collaborative-support group, Sachiko decided to learn more about the students and interaction in the classroom, and she did this in three ways. First, she audio-taped the class and studied the interaction; second, she had the students complete a questionnaire; third, she kept a reflective journal on what went on during the lessons. By listening to the audio tape and writing in her reflective journal, she further realized that she lectured a lot in Japanese and followed a common teacher-solicit, student-response, teacher-react way of teaching (Bellack, Kliebard, Hyman, & Smith, 1966; Fanselow, 1977). She additionally realized that most of the time was spent going over the meaning of words and that her questions to students were
often about vocabulary. She also realized that some students had trouble paying attention. Many looked bored; a few even fell asleep.

She also learned something about the students’ perceptions about the class from their responses to her questionnaire. All the students “got relief from translation”; most said they think about the content when they read, but they cannot fully comprehend the meaning.

Based on discoveries from her explorations, Sachiko set two specific goals: (a) to teach students how to identify the main idea of reading passages, and (b) to make her reading lessons interesting to the students. Instead of lecturing in Japanese, she decided to talk less, have students study a handout that explained the structure of a paragraph, including how to locate the main idea, and to write down answers to questions about the meaning in each paragraph. When she implemented the lesson, she again taped the class and reflected on it in her journal. She also asked students to complete another short questionnaire.

She recognized that she spoke much less, as planned, and that students were no longer sleeping. In addition, some students asked her questions to check their understanding of the reading. All in all, the class was much more attentive. Students said they liked doing the handout activity more than listening to her translation and grammar explanations. However, Sachiko also discovered that many of the students still could not identify the main idea in a passage.

Sachiko thought about how to solve the comprehension problem, and after discussing it with her classmates at Teachers College, she decided to continue to use handouts and have students complete worksheets, as well as write summaries of the paragraphs in both English and Japanese. She would also encourage students to ask questions, as she believed that when students inquire about meaning, they are learning. This rarely happened when she lectured, translated, and asked questions.

A Second Action Research Project: The Use of Japanese and English

At the time of this action research project Ikuko Matsumoto was teaching English in a junior high school in Yokohama. After talking with her support group at Teachers College, she felt that she was using too much Japanese in her teaching, and as a starting point to improve students’ speaking abilities, she decided that she wanted to use more English with her students.

To learn more about her use of Japanese and English, Ikuko decided to videotape classroom interaction. She showed the tape to her Teachers College support group, and together they studied the interaction on the
tape to see when she and the students used Japanese and English. They found that Japanese was used 75% of the time. Ikuko used Japanese to give instructions, explain points of grammar, vocabulary items, and homework assignments, as well as to translate passages and to discipline students. The students used Japanese to answer the teacher’s questions and to ask the teacher questions. The teacher used English to greet, to praise, and to read passages and ask questions from the text. The students used English to greet the teacher, read from the text and to answer questions about the content of the text, such as giving a quote from an author.

Ikuko felt personally surprised and bothered by how much Japanese she used, as this conflicted with her belief that her job was partly to provide chances for students to experience the use of English to communicate their thoughts and feelings. She knew that many of the students would not feel comfortable speaking English with her because they all share Japanese as their first language. She also felt that some students would not be able to comprehend her use of English, for instance, to give grammar explanations and instructions, as they were accustomed to hearing these in Japanese. However, she believed that given enough time and success with comprehending her English, the students’ communicative abilities would improve.

After collaborating with her support group about possible ways to bring English into the class, Ikuko planned to make two small changes in her use of English in the classroom. First, she decided to give classroom activity instructions in English. Second, she decided to ask students questions that are based on the content of the readings, but not those in the textbook, which students could simply read. For example, she would ask, “We read about food in England. If you went to England, what would you like to eat?”

As Ikuko taught the lesson, she used a small digital camcorder to record the class, focusing the camera on the students so she could capture their reactions to her instructions and questions. After the class she viewed the recording, made short transcriptions and analyzed the English used in her instructions. She also studied how the students reacted to these English language instructions, including their nonverbal reactions, such as their facial expressions and gestures. She also made short transcriptions and analyzed the interaction when she asked students questions in English.

Ikuko discovered that the majority of the students looked as if they were following the instructions, but most could not start the task (to silently read a passage in their text, answer six questions, as well as write
down one question of their own) without first confirming this in Japanese. She also discovered that when she asked questions in English that were not in the text, most students, as she predicted, avoided eye contact with her and did not volunteer to answer. However, there were also surprises. It took much more time than she had anticipated for the question-answer session. She was also surprised that two students who were usually quiet volunteered to answer questions and did so in fairly fluent English.

Ikuko was also able to see that the number of students (40), the room itself, and the way students were sitting (in rows) limited the kind of question-answer communication she was hoping to have with them. She noticed, for example, that she stood in the front of the room and most interaction took place with students sitting in the front left side of the room.

Ikuko reflected on what happened. She believed that she became much more aware of how much Japanese is actually used in her English classes; she also felt a stronger commitment to using more English. Also, she realized that giving instructions and asking a few questions (not in the text) in English was only the beginning. She knew that she would have to take the students gradually through a process of switching from Japanese to using English, and that this would not be easy. One main reason for this, she thought, was because of the limitations placed on her and the students due to traditional beliefs within the exam-oriented Japanese educational system. As Ikuko explained: “Teachers and students, and even parents, believe that to improve English, students should analyze complex grammar rules and be good translators...[This is] because these skills are useful to pass exams.”

In this regard, Ikuko felt discouraged about the prospect of successfully getting students to use English as the language of communication in the class. However, she decided to continue to try. She next planned to write key words from her instructions on the board and to point to these words as she gives activity instructions orally in English. She also planned to give students time to write down answers to her questions before asking them to answer them orally.

**A Third Action Research Project: Adult Learners’ Interaction and Autonomy**

When Shelly Jordon was working on this action research project, she was teaching English to adults at a private language school, and she was especially interested in improving her teaching in her Travel English...
course. Her class included ten students between the ages of 40 and 60. These students planned to travel abroad and needed to gain competence in listening to and speaking in English when traveling.

Shelly sensed that she was talking much more than the students. After talking with her Teachers College support group members, she was able to ask questions like, “How much do I talk? How much do the students talk to me? How much do they talk to other students? When do they talk in a natural direct way without my encouragement? Do the students actively listen to each other?”

Shelly videotaped classroom interaction during pair work activities and whole-class discussion. She then met with her support group, and together they transcribed (including nonverbal behaviors) and analyzed random sections of the pair work and whole-class discussions. Shelly discovered that during pair work, the students talked 0% and she 30%. Further, through her analysis of nonverbal behaviors, she saw that the students were not displaying “natural” animated interaction when they talked to each other. For example, one student wrote throughout the entire session, seldom looking at the teacher or other students. Another student kept looking at the teacher while her classmate was talking. Shelly also discovered that during whole-class discussions, she spoke 46% of the time and students 54%. However, although it seemed like the students were talking more, Shelly pointed out that students often took a long time to express an idea, including some rather long pauses. When Shelly spoke to students during the whole class discussion and pair work activity, the interaction followed a traditional teacher-solicit, student-response, and teacher-react pattern.

Shelly wanted to decrease the amount of talking she did and increase student animated talk, as well as provide an atmosphere for students to have natural animated discussions in English. To do this she decided to rearrange the seats so that pairs could easily see each other. She also decided to stand where she was not the center of attention and to limit her English to giving verbal instructions. As a result, the students spoke close to 100% during the pair work activities, maintained consistent eye contact with each other, and used a variety of gestures, body language and facial expressions while speaking. For example, they pointed to themselves and used their hands to indicate size or an action.

Shelly said she learned how to be less dominant in the class simply by not talking, arranging chairs so students could see each other, and by standing outside “center stage.” By doing these things, students seemed to have more animated natural conversations with each other in English,
although they did use Japanese at times. She decided to continue the cycle of action research by having students complete a questionnaire about the kinds of activities they would like to do in the class, then implement these activities to see what happens. She also decided “to dig deeper into the analysis of transcripts” and to identify patterns in the way she and the students interact with each other while doing these activities.

**Benefits of Action Research**

*More Awareness of Our Teaching*

There are at least three benefits to doing action research. To begin, continuous posing of problems and cycling and recycling of behavioral changes to better understand our teaching can provide much more awareness of the consequences that our teaching has on interaction in the classroom. For example, Sachiko realized students could not summarize the meaning of the reading selection even in Japanese after she had explained it. Ikuko found that most of the students did not have confidence in their abilities to understand her English instructions as most could not start the reading activity without first asking her to explain the task in Japanese. Shelly became aware that by changing seating arrangements and repositioning herself to a less central place in the room, she enabled students to have animated conversations.

*Advancement of Our Reflective Skills*

A second benefit from doing action research is that teachers have chances to develop their reflective skills. Scholars (Bartlett, 1990; Dewey, 1933; Farrell, 1999, 2004a, 2004b; Greene, 1986; Murphy, 2001; Richards & Lockhart, 1994; Stanley, 1998; and Zeichner & Liston, 1996) all point out that reflective teaching includes thoughtful persistent consideration of beliefs or practices, often in relation to the problems teachers face in their teaching. Richards and Lockhart (1994) add that a part of reflective teaching includes “collecting data about teaching, examining their attitudes, beliefs, assumptions, and teaching practices, and using the information obtained as a basis for critical reflection” (p. 1). Greene (1986) adds that reflective teaching also involves intuition, passion, and emotion.

As a part of doing action research, teachers have opportunities to develop their reflection-in-action and reflection-on-action (Schön, 1983, 1987) skills. This is what “reflective practitioners do when they look
at their work in the moment or in retrospect in order to examine the reasons and beliefs underlying their actions and generate alternative actions for the future” (Stanley, 1998, p. 585). Stanley adds that “such reflective thinking and examination either during or after the fact can lead to greater awareness on classroom teachers’ part in relation to their knowledge-in-action, or theories, ideas, metaphors, and images they use as criteria for decision making in their teaching practice” (p. 585).

The reflective nature within the cyclic process of action research can take the teacher beyond classroom interaction to a new awareness of issues at the school and community levels that can affect what happens in the classroom. For example, Ikuko was able to reflect on and gain a renewed awareness of why students could not answer questions in English about something they read. She could more clearly see that the emphasis on preparing students to pass exams, with much more focus on analysis of language rather than on the use of English for communicative purposes, was highly valued by colleagues and administrators in her school. Ikuko also realized that parents in the community understand the importance of teaching English in a way that helps students successfully pass college entrance exams even if they don’t agree with it. What Ikuko reflected on reminds me of something Peter Strevens wrote some years ago: “The public is the creator of general levels of expectations, within a community, for the achievement in language of its citizens” (1987, p. 13).

Ikuko further realized through reflection that she cannot go against public expectations and that she needed to creatively work within the established Japanese educational system to give students communicative experience with English while at the same time meeting the expectations of her colleagues, school officials, students’ parents, and others who defined how she was expected to teach.

A Forum to Talk about Teaching

Another benefit of doing action research is the emphasis on creating a forum to talk through issues related to teaching problems, to communicate teaching ideas, and to voice concerns at the classroom, school, and even national levels. During their course experience, Sachiko, Ikuko, and Shelly had chances to attend weekly seminars to talk with other teachers about their own and classmates’ action research projects. As Burns (1999) points out, teachers can also create forums for discussion by disseminating their action research projects. For example, it is pos-
sible to report on action research at conferences (group and individual reports, poster sessions) and in teacher-centered journals.

Limitations of Action Research and Other Avenues to Awareness

A Focus on Problems and Best Ways to Solve Them

There are at least two limitations to action research. First, action research is focused on problems, and when we constantly are looking for ways to understand a problem and ways to solve it, we miss chances to explore teaching ideas outside of this problem (Fanselow, 1987, 1988). For example, Ikuko posed the problem of using considerable Japanese in her class, and she decided to use more English by giving verbal instructions in English to introduce reading activities. She then became aware that many of the students could not fully understand these instructions. As such, she started to look for a better way to give instructions, such as writing key words from her verbal instructions on the board.

Although Ikuko’s problem posing process seems reasonable, if there is no focus on a problem and the subsequent search for a best way to solve the problem, Ikuko could be free to see her teaching differently by trying new things simply to see what happens. For example, Ikuko could follow her interest in giving instructions in English, not from a problem-solving point of view by looking for a better way to give instructions, but rather simply to explore what might happen. She could try out a number of ways of giving instructions, such as giving the instructions as a dictation, or as a cloze-dictation (in which students listen, read, and fill in every fifth word), or by having students read instructions on an overhead or handout, or having students listen and translate, or listen to a song with the instructions, or chant them with the teacher, or listen to the teacher whisper the instructions. Again, the idea is that if there is no focus on a problem, the teacher can feel free to explore teaching possibilities simply to see what happens, and this can generate considerable awareness.

A Narrowly Defined Process

When we consistently follow a single process, we also place limits on how we can explore our teaching and on the kinds of awareness we can gain from our exploration. Action research includes identifying a problem, doing a preliminary investigation, reflecting on what happened, planning an action, implementing the plan, observing the outcome,
reflecting, planning the next action based on reflection, and moving to the next cycle of re-identifying a problem, and so on. As useful as this process can be, by only following this linear process, we miss chances to gain awareness of our teaching outside of this process.

Although the action research process makes sense and is certainly worth doing, we can go beyond this process by exploring a variety of other avenues to awareness outside the problem-posing one of action research. One of these avenues is to explore simply to see what happens. To do this, Fanselow (1987, 1988, 1992a, 1992b, 1997) suggests we try the opposite of our usual classroom behavior. For example, if we are aware that we say “very good” after most student responses, we can be silent, and then describe what happens. If we usually have students sit in rows, we can have them form a semicircle. If we always teach from the front of the classroom, we can try teaching from the back. If students read aloud in every class, we can ask them to read silently. If we ask students to form groups by themselves, we can assign students to particular groups. And we can do so creatively. For example, we can give out pieces of candy. Students with cherry-flavored candy form a group, lemon another group, and so on. The idea is to discover what we normally do and to try the opposite to see what happens.

Although the three teachers gained awareness from doing their action research projects, they could have gained considerable awareness simply by trying the opposite of their normal practice just to see what would happen. For example, instead of verbally analyzing the grammar of a reading passage, Sachiko could have had students read a passage silently several times while doing a different task each time, such as underlining all past tense verbs, circling subordinate clauses, highlighting main ideas, and answering multiple choice and true-false questions. Likewise, Ikuko could have tried writing down the instructions for students to read one day, then give them orally the next day, then have a student read the instructions aloud to the class the next, continually trying out something different to see what happens.

Another avenue to awareness is exploring to see what is. One way to do this is to explore what we actually do in our teaching as opposed to what we think we are doing. For example, if a teacher thinks she has designed group work activities that keep students talking in English and staying on task, she could tape record students’ group work interaction. By doing this, she could analyze the interaction to determine if students are talking in English and staying on task. All three of “our” teachers could have explored what they actually did in relation to what they be-
Gebhard believed they were doing in their classrooms. For example, Sachiko could have asked, “Am I actually teaching students to read as a process of understanding meaning?” Ikuko could have asked, “Do students perceive that I am giving instructions in English so that they will become more competent in comprehending English?” Shelly could have asked, “Am I actually moving off center stage in the classroom when I position myself away from the students?

We can also explore our teaching by considering what we believe as teachers in relation to what we do. Do our beliefs match our actual practices? For example, some teachers who stated that they don’t believe in correcting students’ oral errors, constantly corrected them anyway (Jimenez-Aries, 1992). Again, all three teachers could have gained awareness by considering their beliefs in relation to what they were doing as teachers. For example, Sachiko could have asked “Does the way I teach reading match my beliefs about teaching reading as a process of comprehending meaning?” Shelly could have asked, “What are my beliefs about teaching students to converse in English? Do I believe that students are genuinely learning to converse in English when I do not participate in conversations with them—when I have them talk in pairs or small groups?”

We can also explore to gain emotional clarity. By exploring our feelings, we can gain awareness about things we feel deeply about, or don’t really care about, or are ambivalent about. As Jersild (1955) and Oprandy (1999) emphasize, we can pay attention to the affective side of teaching, including how we feel about the students, teaching, and ourselves as teachers. For example, by focusing on exploring her emotions, Ikuko could have gained more emotional clarity about her inner conflict between teaching students to analyze English to pass exams and teaching them to communicate in English.

**Concluding Remarks**

As the three examples given illustrate, action research can provide a way for us to identify problems in our teaching, as well as ways to solve these problems, and this process can provide us with much awareness about our teaching. Action research can expand our reflective skills and provide a lens through which we can discuss teaching issues, problems, and teaching beliefs and practices. However, action research focuses distinctly on problems in our teaching. It also uses a rather narrowly defined linear process, which if followed uncritically can constrict the
awareness we seek when we undertake action research in the first place. As such, in addition to doing action research, we can benefit from using other avenues to awareness. In addition to problem-solving, we can explore our teaching simply to see what happens, for example, by trying the opposite of what we usually do. We can also explore to see what we actually do as opposed to what we think we do, as well as what we believe in relation to what we do in the classroom. We can also take the exploration avenue of trying to gain emotional clarity about our teaching.

I close with this thought. Exploring our teaching through action research, as well as going beyond this problem-posing and solving process by using other avenues to awareness, can be embodied in the words of Akiko Ueda-Montonaga, who took time to observe her teaching:

I have sought alternatives in teaching and found them. After I found that I have alternatives, I felt freer and securer about deciding on activities for the students... I will continue observing and investigating classroom interaction to find what is going on in my teaching...I will make a wish, make a dream to see teaching clearly and differently, and imagine all the possibilities of teaching! (Gebhard and Ueda-Motonaga, 1992, pp. 190-191)

Jerry G. Gebhard is Professor of English and Director of Graduate Studies in Composition and TESOL at Indiana University of Pennsylvania. He also teaches courses for Teachers College, Columbia University, Tokyo and has had extensive experience teaching in Asia.

Note
1. To protect the identities of the teachers studying in the MA TESOL program, I use pseudonyms in this report.

References


(Received July 12, 2004; accepted December 1, 2004)
Research Forum

Score Reliability and Placement Testing

Paul A. Westrick
Kyuushu University

This study examines the piloting of a commercially-produced test of English, the Quick Placement Test – Pen and Paper Test (QPT-PPT). In consecutive administrations of two versions of the test with 161 first-year students at a Japanese university, the test results failed to discriminate among students of varying proficiencies. Narrow ranges, low score reliability estimates, and large standard errors of measurement characterized the results. Item analysis revealed that most of the test items did little to separate high and low scoring students. The data also suggests that test anxiety, familiarity with the test format, and test-taking skills were important factors in the test scores.

Throughout Japan, university and college teachers of English frequently complain that their institutions do not place students in classes based on English proficiency levels. Because creating a placement test requires some knowledge of testing principles and can be time consuming, purchasing a commercially-produced proficiency test may appear to be an easy solution. On the other hand, commercially-produced tests are often expensive, and scoring them can take time. For
these reasons, it was decided to independently pilot test the Quick Placement Test – Pen and Paper Test (QPT-PPT), an inexpensive commercially-produced placement test, created by the University of Cambridge Local Examinations Syndicate (UCLES) and published by Oxford University Press (OUP).

Research Focus
Placement Tests

There are a variety of reasons why teachers want placement systems at their institutions. One is that mixed ability classes are hard to teach effectively. Without a placement system, teachers typically struggle to aim instruction in ways that do not simultaneously befuddle lower-level students and bore higher-level students. In an attempt to reach all students at all ability levels, teachers may try to create different activities for different groups within a single classroom, “mini-lessons” that try to please everyone but often please no one. In contrast to this inefficiency, Redondo (2000) points out that, “Teachers can use materials, as well as teaching styles and pedagogical approaches more effectively when classes share a greater similarity of learning qualities and characteristics” (p. 126). Besides classroom teaching efficiency, another problem of mixed-level classes is grading. Teachers often face the dilemma of whether to lower standards to pass hard-working low-level students or to maintain standards and fail low-level students who try their best but simply cannot compete with their classmates. Unless the instruction is concentrated on their needs, the weakest students invariably fall further behind their peers. Another issue intertwined with the others, and probably the most important, is the amount of learning that takes place. As Chauncey and Frederiksen (1951) recognized long ago, for a student to advance, that student should “take courses which are neither too difficult nor which involve wasteful duplication of earlier-learned content” (pp. 108–109). Because of the problems associated with mixed ability classes, many teachers, as Brown (1996) observes, tend to prefer having students grouped by ability levels, bringing together those with similar strengths and weaknesses. With such a class, the teacher can concentrate on addressing their weaknesses in ways that are appropriate for those students.

These teachers generally want students to be sorted with a placement test, and here it is necessary to differentiate between placement tests and proficiency tests. Though they are similar in that they are norm-
referenced tests (NRTs), in which students’ performance is compared to that of other test takers, they are not identical, as Brown (1996, pp. 11–12) makes clear:

Experiencing the similarities and differences between proficiency and placement testing will help to clarify the role of placement tests. To begin with, a proficiency test and a placement test may look very similar because they are both testing fairly general material. However, a proficiency test tends to be very, very general in character, because it is designed to assess extremely wide bands of abilities. In contrast, a placement test must be more specifically related to a given program, particularly in terms of the relatively narrow range of abilities assessed and the content of the curriculum, so that it efficiently separates the students into level groupings within that program.

Hughes (2003) expresses similar thoughts and advises that placement tests should be developed by the users themselves so that they specifically meet their needs. As for commercially-produced tests, he states that:

Placement tests can be bought, but this is to be recommended only when the institution concerned is sure that the test being considered suits its particular teaching programme. No one placement test will work for every institution, and the initial assumption about any test that is commercially available must be that it will not work well. (p. 16)

Ideally, an institution’s placement test should be connected with its curriculum, but institutions still choose commercially-produced proficiency tests as placement tests for a variety of reasons, some more defensible than others. Starting with the least defensible, one reason given for choosing a commercially-produced test is status. Often administrators, and unfortunately some language teachers, believe that they can improve the image of their school by using a famous test or a test produced by a famous institution. Another poor reason (or excuse) given is that creating a placement test is too difficult. A lack of enthusiasm or confidence often leads to the decision that it is best to defer to the experts, the producers of commercial tests. At some institutions, a dedicated group of teachers may win agreement or approval for a placement test, but are stifled in their struggle for the development of a true curriculum. “Curricula” that consist of little more than course titles are, sad to say, not uncommon. Without institutional guidance, instructors teach whatever
they find appropriate with little or no regard for course titles or what other instructors teach. In such a situation, a test that emphasizes specific language targets may be no better than a commercially-produced general proficiency test. Probably the most defensible reason that can be given for using a commercially-produced test is that such a test is used only as a temporary measure until it can be replaced with a customized test that matches a true curriculum, a curriculum that has clearly defined goals and objectives (for a more detailed discussion of testing and curriculum development, see Brown, 1995, 1996).

For those institutions that decide to implement a placement system, one last critical factor is time. As noted in Poel and Weatherly (1997), and Fulcher (1997), the brief time between when students arrive and when classes begin often poses a problem for placement personnel, and the amount of time allotted for placement testing may be less than optimal. Ideally, the placement test should be quick and easy both to administer and to score. The initial attraction of the QPT-PPT is that it can be administered in 30 minutes, and the answer sheets can be scored on campus with transparent overlays. The trade off is that with a commercial test there is the possibility that many of the items are not written at an appropriate level for the examinees at a given institution. In Classical Testing Theory (CTT), generally speaking, longer tests produce scores that are more reliable than short tests (Brown, 1989; 1996). In Item Response Theory, test length may not be a critical factor, but even in CTT it is recognized that if a short test contains only good items that clearly separate the examinees, the score reliability estimate may be high. But if a short test is filled with items so easy that almost all the students can answer them correctly or items so difficult that almost all the students answer them incorrectly, there may not be enough remaining high-quality items to separate the students and reliability will suffer.

**Item Analysis**

Item quality has an effect on the amount of score variance, and the amount of score variance influences the reliability of the scores. If a placement test contains items that all examinees answered correctly or items that all examinees answered incorrectly, those items can be removed from the test without altering the distribution of scores or the ranking of examinees; in other words, the amount of variance would remain the same. Such items fail to separate examinees and provide no useful information for placement decisions; therefore, they can be eliminated.
Instances where all examinees answer an item correctly or all answer an item incorrectly are rare, but at what point do we decide that an item is functioning well, that is, separating students and creating score variance? Item analysis is not an exact science, but over the years testers have created some general guidelines for analyzing the quality of items. One tool test writers have is the item facility (IF) index, the proportion of examinees answering an item correctly. Brown (1996) states that IFs between .30 and .70 are generally acceptable for an NRT, and McNamara (2000) states that testers should be content with IFs between .33 and .66. Another evaluation tool is the item discrimination (ID) index, the difference between the IF for the highest-scoring third of the examinees and the IF for the lowest-scoring third of the examinees (ID = IF_{upper} - IF_{lower}). According to Ebel’s guidelines (1979, cited in Brown, 1996), an item with an ID at or above .40 is considered very good. Those from .30 to 39 are considered reasonably good, those between .20 and .29 marginal, and items in both categories may be improved. Items below .19 are considered poor, and may be either rewritten for improvement or discarded. In sum, a test needs items that are neither too difficult nor too easy, and the items should discriminate between the higher- and lower-scoring students. Such items are needed to generate score variance, which in turn affects score reliability.

**Score Reliability and Validity**

Reliability, the precision and consistency of measurements, and validity, measuring what one claims to be measuring, are vitally important. A scale that says a person weighs 50 kilos one minute and 100 kilos the next is not giving reliable measurements. Even if a scale does give reliable measurements, those measurements are only valid measurements of the person’s weight; they would be invalid measurements of the person’s height.

Discussing reliability and validity in detail is beyond the scope of this paper, but teachers should keep a few points in mind. First, a test cannot be considered reliable or unreliable because reliability is a characteristic of scores, not of the measure. For this reason, Thompson (2003a) stresses that testers and researchers should be clear by talking about the reliability of scores, not the reliability of tests. Two very different groups may produce widely different scores on the same measure, as Brown (1989) demonstrated with a cloze test. The second point to remember is that, generally speaking, heterogeneous groups tend to have more variance.
in their observed scores and therefore produce more reliable score estimates than homogeneous groups do. In the words of Thompson (2003b, p. 93, emphasis in original):

Reliability is driven by variance – typically, greater score variance leads to greater score reliability, and so more heterogeneous samples often lead to more variable scores, and thus to higher reliability. Therefore, the same measure, when administered to more heterogeneous or to more homogeneous sets of subjects, will yield scores with differing reliability.

Third, while reliability estimates may change from group to group, there are standards to be met. This point is especially true with high-stakes tests, and as there are times when placement tests are high-stakes tests, these standards should be met. For important tests, score reliability coefficients of .90 are considered the minimum and reliability coefficients of .95 or higher are preferred (Hopkins, Stanley, & Hopkins, 1990; Nunnally & Bernstein, 1994). Considering that the highest possible reliability estimate for scores is 1.0, having a reliability coefficient of .90 appears to be quite good, but even with a reliability coefficient of .90, the standard error of measurement will be nearly one-third the size of the standard deviation (Nunnally & Bernstein, 1994).

The fourth point is that there can be no score validity without score reliability, for it is difficult to make sound judgments when the measurements of what is claimed to be measured are unreliable. As Thompson (2003a, p. 6) explains,

Perfectly unreliable scores measure nothing. If the scores purport to measure something/anything (e.g., intelligence, self-concept), and the scores measure nothing, the scores (and inferences from them) cannot be valid. Scores can’t both measure nothing and measure something. The only time that perfectly unreliable scores could conceivably be valid is if someone was designing a test intended consistently to measure nothing.

While zero reliability is the extreme, we know that we want a high reliability estimate and a small standard error of measurement (SEM), but do most teachers really understand why? Teachers generally understand percentiles and have seen distribution curves, but understanding band scores and the percentages that fall within a standard deviation (SD) in a normal distribution (a bell curve) can clarify the importance of high reliability.
In CTT, the SEM is used with an examinee’s observed score to create band scores (Bachman, 1990; Hughes, 2003) to estimate the examinee’s true score. We determine the SEM by multiplying the SD by the square root of one minus the reliability estimate \((r_{xx'})\). That is, \(SEM = SD \sqrt{1 - r_{xx'}}\). The higher the reliability estimate is, the smaller the standard error of measurement. Using one SEM above and below an examinee’s observed score, we create a band score, and we can estimate with about 68 percent certainty that an examinee’s true score falls within that band. Using two SEMs, we can estimate a true score with about 95 percent certainty, and with three SEMs we can estimate a true score with about 99 percent certainty. Turning now to the normal distribution, the bell curve, we know that roughly 68 percent of the people will fall within one standard deviation from the mean – 34 percent above and 34 percent below. A student with an observed score at the mean is at the 50\(^{th}\) percentile; a student one SD below the mean is at the 16\(^{th}\) percentile; and a student one SD above the mean is at the 84\(^{th}\) percentile.

Now let’s put this together. For example, we have a 60-item placement test, and we know that the score reliability estimate is .90; the average score is 30; the SD is 10 points; the SEM is 3.16; and the distribution is normal. A student with an observed score of 30 is at the mean, thus at the 50\(^{th}\) percentile. Under such circumstances, using band scores that extend out three SEMs, we can estimate with 99% confidence that the student’s true score is in a band between 20.52 and 39.48, a band that covers nearly two standard deviations. That band does not quite stretch all the way to the 16\(^{th}\) and 84\(^{th}\) percentiles, but it does come rather close. If making fair and defensible decisions with score reliability estimates at .90 is difficult, making fair and defensible decisions with score reliability estimates below .90 is more difficult, if not impossible.

**Past Studies of Placement Tests**

There have been a number of studies of placement tests (Blais & Laurier, 1995; Fulcher, 1997; Poel & Weatherly, 1997; Wall, Clapham, & Alderson, 1994), but most of these studies concern customized tests developed at the institutions where they were used. Studies of commercially-produced tests have been harder to find. In one study, Culligan and Gorsuch (1999) explored the use of a commercially-produced proficiency test (the Secondary Level English Proficiency test or SLEP test) as a placement test at a Japanese university and determined that it did not work very well for them. In their analysis of the test results, they discov-
erected that most of the test items did not discriminate between the high- and low-scoring students, and the test had a high SEM. Furthermore, and more importantly, the test did not match well with their curriculum.

Although that study suggested that a commercially-produced test was inappropriate for an institution with a defined curriculum, the use of such a test at an institution with a less clearly defined curriculum could possibly be worthwhile. In a very small piloting of the QPT-PPT for review purposes at a Japanese university with liberal admission standards, the test separated students much as the instructor expected (Westrick, 2002). In two classes with a total of 27 students, the scores of most Japanese students were clustered at the low end of the scale, just above the few Chinese students who had had no formal English education, and below the Chinese and Korean students who had had formal English instruction in their home countries. Though it was not covered in the published review, removing the Chinese and Korean students decreased the range of observed scores and the score variance. The narrow range of scores for the Japanese students who made up the majority of test-takers warranted a larger piloting of the QPT-PPT.

**Research Questions**

The primary objective of the current study was to test the test. Ideally, the results of the administrations of the QPT-PPT would be analyzed with the students’ scores from the Center Test and university entrance exam, but this was not possible. All information regarding student scores on the Center Test and the university’s entrance exam is considered private and was unavailable for the study. Operating under this constraint, the research questions were as follows:

1. Would the QPT-PPT separate university students who have already been sorted by the Center Test and a university entrance exam?

2. As placement tests have serious implications for students, would the reliability coefficients of the test scores for these students be at or above 0.90?

3. As there are “two parallel (photocopyable) versions” (QPT-PPT user manual, p. 2) of the QPT-PPT, how high would the correlation between student scores on Version 1 and Version 2 be?
Information regarding the cutoff points for those tests would have been useful in regard to the first research question, and a study that correlated the students’ QPT-PPT scores with their Center Test and/or entrance exam scores would have been ideal for establishing parallel-forms reliability in the third question, but as stated earlier, the students’ Center Test and entrance exam scores were unavailable due to privacy concerns.

Method

Participants

With one exception, all of the participants in the piloting of the test were first-year technology or economics students at a national university (one law major was enrolled with the economics majors). They were in three different class sections: one was for economics majors, and two separate sections were for technology majors. Section sizes ranged from 53 to 60 students, and 161 of the 167 students enrolled in the first-year English classes took the tests. Of those students, 159 were from Japan, one was from the People’s Republic of China, and one was from South Korea. New students are admitted to their department within the university largely based upon their performance on the national Center Test and the university’s entrance examination. Once admitted to their department, students are placed in jun order (an “alphabetical” order based on the pronunciation of their family names in the Japanese phonetic system), assigned a student number, and then placed into sections within their department in this alphabetical/numerical order. All students within each section are required to take a general English course with the other students in their section regardless of their scores on the Center Test and the university’s entrance examination. There is no initial placement system.

Materials

Created by UCLES and published by OUP, the QPT-PPT is, at first glance, quite easy to use. It is legally photocopiable, so an unlimited number of tests can be reproduced, and it can be scored locally with transparent overlays that come with the test package. The QPT-PPT tests reading, grammar and vocabulary skills only. The user manual suggests that the QPT not be used as the sole instrument to evaluate and place students, that speaking and writing skills be assessed, and that with the PPT, a listening component should be added (the QPT-Computer Based
Test, the CBT, has listening items built into the test). UCLES does not assert that the QPT is the only thing teachers need in order to make placement decisions; it is but one resource that English language programs can utilize when making placement decisions.

There are two versions of the test, Version 1 and Version 2, and each version has two parts. Part 1 has 40 multiple-choice items, and all test-takers must complete this portion of the test. (Examples of items similar to those found on the actual tests can be seen in Appendix A.) Items 1 through 5 provide test-takers with notices (signs, postings, etc.) and ask where these notices would be seen. Following the notices are three cloze passages with five items each (items 6–20); test-takers choose the best word to fill each space in the passage. Items 21 through 40 are individual sentences with a word or short phrase missing; students choose the best word or phrase to complete each sentence. For items 1 through 10, there are three possible answers for each item, and for items 11 through 40 there are four possible answers for each item. Part 2 of the test consists of an additional 20 multiple-choice items that are designed for more advanced learners. It starts with two cloze passages with five missing words or phrases for each passage (items 41–50); followed by individual sentences with a word or short phrase missing (items 51–60). All 20 items in Part 2 have four possible answers. It should be noted that at the beginning of Part 2 there are written instructions for students not to do Part 2 unless told to do so (which is important because it caused problems in the first administration in this study).

Test administrators have the option of giving their students only Part 1 of the test if they believe that their students are below the advanced level according to the Association of Language Testers in Europe (ALTE) descriptions of users’ language abilities, which are provided in the QPT-PPT user manual. For administrators unsure of their students’ levels, the user manual advises the use of both parts of the test (all 60 items). Test scores from the 40 item and 60 item tests can be matched to the level chart in the user manual. For the 60 item test, students with scores between 0 and 17 are at the beginner level; those between 18 and 29 are at the elementary level; those between 30 and 39 are at the lower-intermediate level; those between 40 and 47 are at the upper-intermediate level; those between 48 and 54 are at the advanced level; and those between 55 and 60 are at the very advanced level. Again, the user manual suggests that the QPT be used with other forms of assessment to help make decisions about students who score within one SEM (±4 on the 60 item test, ±3 on the 40 item test) of a cutoff point on the ALTE scale.
It should be noted that at the time of this writing, the QPT-PPT packet with answers could be purchased from booksellers on the Internet, making security a concern. The QPT-CBT may be safer because the computer checks item responses without giving the answers, but the possibility remains that someone could purchase a CBT packet and record the questions while taking the test multiple times. In this study, it appears that none of the students had access to copies of the tests prior to the day of the administrations.

Procedure

The participants took the QPT-PPT in the first semester of their first academic year. Each class section was told in English that they would take two versions of the test, Version 1 and Version 2, and they were told to do both Part 1 and Part 2 of each test. They were told that half of the class, Group 1, would take Version 1 first, and the other half, Group 2, would take Version 2 first, and that they would have 30 minutes to complete the test (This first test administration is hereinafter referred to as Administration A). After the first test was completed, the tests were collected and the second test was given. Students in Group 1, those who had taken Version 1 first, then took Version 2, and students in Group 2, those who had taken Version 2 first, then took Version 1 (This second test administration is hereinafter referred to as Administration B). They again had 30 minutes to complete the test. These instructions were also written on the chalkboard in English. Students were told to mark their answers on the answer sheet, and the instructor modeled on the chalkboard how to mark answers on the answer sheet. Students were asked if they had any questions. There were no questions.

During Administration A, the instructor walked about the classroom observing the students as they took the tests. In each of the three class sections, the instructor observed that some students stopped after completing Part 1 of their tests. When this was observed, the instructor told those students to continue on to Part 2, and a general reminder was made to the entire class that they were to do Part 2 after finishing Part 1. After thirty minutes, the tests and answer sheets were collected. After all tests and answer sheets were accounted for, students received the second test (Version 2 for Group 1 and Version 1 for Group 2) and a new answer sheet. After the second test administration, Administration B, was finished, tests and answer sheets were again collected. When tests and answer sheets had all been accounted for, the students were released.
Total administration time varied between 80 and 85 minutes for each class section. Answer sheets were hand scored using the transparent overlays provided in the test packet. (This took one marker much longer than the time needed to administer the tests. Scoring the tests with an optical scanning machine would have made this task much quicker and easier.) Scores were double checked when each student’s answers were entered into a spreadsheet for analysis.  

Results

The data gathered from the back-to-back test administrations revealed interesting changes in how 64 students (nearly 40% of the participants) approached the test during the second administration. In Administration A, four students did not do Part 2 after completing Part 1 despite each class being told repeatedly to do so; two students apparently started with Part 2 of their tests, struggled, and were then unable to complete Part 1. It appears that these six students did not understand the teacher’s instructions. Additionally, 8 students did not attempt to answer all the items on their tests. They apparently were not as test-wise as the other 97 students, who either paced themselves so as to have time to read every item or realized there was no harm in guessing on difficult items and continuing onward through the test. In Administration B, these differences largely disappeared. All students did Part 1 then Part 2 of the test, and all but 13 students attempted to answer all 60 items. Consequently, the descriptive statistics for Administration B are quite different from those for Administration A.

In Table 1, in Administration A, the students’ means and medians for Version 1 (AV1) and Version 2 (AV2) were very close. Unfortunately, the ranges of scores on both tests were below 30, meaning more than half of the scale went unused. The reliability of test scores for Version 1 was 0.61 while the reliability of test scores for Version 2 was slightly lower at 0.59. For the test scores on both versions, the SEMs were more than half the size of the SDs. It also appeared that students scored better on Version 2 than on Version 1.

In Administration B, it appears that all the students started with Part 1 of their tests and attempted Part 2 after completing Part 1, that 148 of 161 provided answers for every test item, and that the 13 who did not answer every item came very close to doing so. As a result, the variance in test scores decreased. Score ranges became even narrower (20 for Version 1 and 18 for Version 2); the reliability coefficients for the test scores fell to
0.46 for G2-BV1 and 0.41 for G1-BV2; and the gaps between the SDs and SEMs narrowed even more dramatically. It again appeared that students scored better on Version 2 than on Version 1, and the average scores on both versions of the test rose in the second administration.

In both groups shown in Table 1, students’ scores on both versions are combined. This spread students out further than just using one version of the test. Despite the wider ranges of scores, the ranges were still narrow in relation to the scale, and the widest range of scores was 36 (Group 2), meaning only 30 percent of the scale was utilized. With regard to score reliability, the estimates for the combined scores were higher than those for a single administration, as would be expected, but for neither group did the reliability estimates approach .90.

The item analysis summaries in Tables 2 and 3 help explain the low reliability estimates. Very few test items were at the appropriate level for the students and not many items separated high and low scorers. Fewer than half of the items on both versions of the test had IFs between .30 and .70. That is, more than half of the items on both versions of the test appear to have been too easy or too difficult for the students. Furthermore, even fewer items had IDs of .40 or higher. The vast majority of IDs fell below .19, and in all but one instance there were more negative IDs than there were IDs above .40.
Two other observations about the individual test administrations should be made. First, the correlations (Pearson product moment correlations) between the two versions of the test were low for both groups even though they are supposed to be parallel forms, and the order in which the tests were taken seems to be a factor. Second, there may have been regression toward the mean in Administration B. Predicting the amount of regression depends on the reliability of the test scores (Hopkins, et al., 1990). With perfect score reliability ($r_{xx'} = 1.0$) there should be no regression; with perfectly unreliable scores, all examinees should be
expected to regress to the mean. With low score reliability in Administration A, some regression toward the mean in Administration B should have been anticipated. It seems that regression to the mean did occur, as scores in Administration B were more tightly clustered around the mean than in Administration A. However, a closer look was warranted because of the differences in approaches in Administration A, so students were divided by their approaches to see how much their scores differed between the two administrations (see Table 4). Removing confusion and adding experience clearly made a difference.

### Table 4. Changes in Total Scores Based on Approach

<table>
<thead>
<tr>
<th>Students in Administration A who ...</th>
<th>N</th>
<th>Mean score Administration A</th>
<th>Mean score Administration B</th>
</tr>
</thead>
<tbody>
<tr>
<td>answered all 60 items</td>
<td>97</td>
<td>36.70</td>
<td>37.12</td>
</tr>
<tr>
<td>did Parts 1 and 2 but did not answer all 60 items</td>
<td>58</td>
<td>34.05</td>
<td>36.81</td>
</tr>
<tr>
<td>did Part 1 but not Part 2</td>
<td>4</td>
<td>30.50</td>
<td>39.25</td>
</tr>
<tr>
<td>did Part 2 but did not finish Part 1</td>
<td>2</td>
<td>23.00</td>
<td>39.00</td>
</tr>
</tbody>
</table>

Note: The 58 students who did not answer all the items in Administration A answered an average of 8.14 more items in Administration B.

Aside from the changes in scores, the differences in how the students approached the test revealed additional information concerning reliability, correlation, and regression to the mean. Regarding reliability, removing the six who were confused by the instructions (one from Group 1 and five from Group 2) lowered the reliability estimates for G1-AV1 0.61 to 0.55 and for G2-AV2 from 0.59 to 0.55, and it decreased the ranges from 28 to 20 and from 26 to 20 respectively. Their removal improved the correlation coefficients for the two groups, from 0.35 to 0.40 for Group 1 and from 0.49 to 0.54 for Group 2, but the correlations were still low. Even when looking at only the 97 students who approached the test consistently answering all 60 items in both administrations, the score correlations were 0.48 for Group 1 (N=52) and 0.65 for Group 2 (N=45). Regarding regression toward the mean, practice and experience (Brown, 1988) could explain the improvements of 64 students and the decrease in test score variance, yet for the 97 “consistent” students, the SD for Group 1 (N=52) fell from 4.36 in Administration A to 3.71 in Ad-
administration B, and for Group 2 (N=45) it fell from 4.97 in Administration A to 4.45 in Administration B, suggesting that regression to the mean did occur in Administration B.

**Discussion**

Though not part of the initial research focus, test anxiety and the practice effect must be addressed. Regression toward the mean can partially explain the differences in the descriptive statistics for the first and second administrations of the QPT-PPT, particularly the decline in the reliability estimates, but the increase in the number of items answered and the slight rise in the means for both groups in both tables suggest that the students were unsure of how to approach this test the first time they took it. As mentioned earlier, a review of the data revealed that six students apparently misunderstood the directions, and 58 students apparently were not as test-wise as other students were. With experience and practice, it appears that the students became more comfortable with the test format, and the average scores for these students rose in Administration B.

Without practice and experience, many students (if not most) have some degree of anxiety before and during a test administration. For those familiar with English education and testing in Japan, it is common to see students initially paralyzed when they encounter their first “English-only” test. The students in this study had taken two English-only listening tests already—they had taken a diagnostic test on the first day of the semester, and they had taken their first test of record five weeks later. Many were visibly distressed when they took the diagnostic test, but they were much more relaxed and confident when they took their first test of record five weeks later. It seems that they made the same adjustment with the QPT-PPT within a much shorter time frame. When the students learned on the day of the test administrations that the tests were made by the University of Cambridge Local Examinations Syndicate and published by Oxford University Press, one could not miss the visible and audible responses of anxiety. The test had a high degree of face validity with the students, and many were manifestly intimidated, but their anxiety levels declined as they progressed through the tests, particularly during the second administration.

Anxiety may also have been a factor in the apparent confusion six students had with the instructions. How often have teachers looked at a student’s test, seen that the student did not understand the instructions and performed poorly as a result, and then said something along
the lines of, “Well, that tells me something.” Yes, it tells the teacher something—the student did not understand the instructions, and as Japanese students, generally speaking, do not ask questions because they do not want to appear foolish, such a student would not ask for clarification. It does not necessarily tell the teacher that the student is significantly less proficient than the other students in whatever intended construct the teacher is trying to measure. As a case in point, the “worst” student in Administration A transformed into an “above average” student in Administration B. He gained 22 points not because his English skills improved in under an hour; it is just that with a bit of experience, he learned how to take the test. That student became test wise, as did many others in this study.

Using the data from Administration A to make judgments on English reading, grammar, and vocabulary proficiency is probably inappropriate. Anxiety, confusion, and lack of experience created variance in the scores, but it was error variance because it was “not related to the purpose of the test” (Brown, 1996, p.188). Reducing this error variance caused the range and reliability of the scores to fall in Administration B. The data collected from Administration B, in which all the students understood the test format, had prior experience with the test, and were relaxed, probably provides a more trustworthy measurement of the intended construct. One could argue that another option would be to combine the scores from both administrations, mitigating the negative effects of the first administration for some students without throwing away the other students’ time and effort, but this would still be unfair to the 64 students who were not as test wise as their peers.

The disappointing results from this pilot testing do not necessarily mean that the QPT is a bad test. Again, the expression “the reliability of the test” should be avoided; “the reliability of the test scores” should be used. Closely linked to this was the fact that a test may produce reliable scores with one group of test participants but produce unreliable scores with a different group of test participants. A placement test should separate students along a continuum, and the range of scores should be wide enough to make distinctions, but with this population of students, the QPT-PPT did not separate students very well, and the reliability estimates of the scores were quite low. The QPT-PPT may very well distinguish among students who have widely different educational backgrounds. In the QPT-PPT user manual (UCLES, 2001, p. 14), it says that, “prior to publication the QPT was validated in 20 countries by more than 5,000 students.” At the end of the description of the two validation
phases of the QPT, the manual concludes with this sentence: “By investigating the reliability of the test scores as well as the tests themselves, we have produced a test which is both reliable and practical” (p. 14). Considering the diversity of the participants in the development of the test, it would come as no surprise that students’ test scores were broadly dispersed along the continuum and that the reliability of the test scores for both versions was acceptable. (Though SEMs are given in the user manual, reliability coefficients are not given.)

EFL situations tend to be quite different from each other, as this piloting of the QPT-PPT with the first-year university students suggests. Broadly speaking, this pilot study supports Culligan and Gorsuch’s (1999) observation that there is typically not much diversity as measured by global proficiency tests among the incoming student populations at Japanese universities. In Japan, students follow a national curriculum through junior and senior high school, and for those going on to university, many are then sorted first by a nationwide test (the Center Test) and again by university entrance exams. If anything, the QPT-PPT simply confirmed that students selected for admission had been screened by the Center Test and possibly the university entrance exam with regard to their proficiency in reading, grammar, and vocabulary. In this study, an additional general test of English reading, grammar, and vocabulary, with low score reliability estimates and SEMs approaching the size of the SDs, provided little information that could be used for placement decisions.

This leads to the final point made in the section on score reliability, which is that without score reliability, score validity is cast into doubt: if the measurements cannot be trusted, sound inferences cannot be made based upon those measurements. Using the data for all students in Table 1, one can see that a student with an observed score of 37 on test Version 1, Administration B, would be at the 50th percentile. Going out three SEMs ± one can estimate with 99% confidence that the student’s true score is between 28 and 46. Looking at the observed scores, which ranged from 27 to 46, and then looking at the student’s band score, one could estimate that the student could be somewhere between the 2nd and 99th percentiles. Again using the data from for all students in Table 1, one can estimate that a student who scored 38 on test Version 2, Administration B, is just slightly above the 50th percentile. Going out three SEMs ± one can estimate with 99% confidence that the student’s true score is between 29 and 47. This student’s band score is actually larger than the range of observed scores. Observed scores of 30 and 47 put students at the 1st and 99th percentiles respectively. There is a tremendous amount of
overlap when using band scores at the 99% confidence level, and though some distinctions can be made at the extremes, the “average” students in either group may very well be among the strongest or weakest students in their respective groups. Learning that a student is between the 1st and 99th percentiles after the test is not much better than knowing the student stood between the 1st and 99th percentiles before the test. Though these scores are not perfectly unreliable, they are unreliable enough to have come close to measuring nothing.

Conclusion

While score reliability matters, it is a characteristic of the group, not the test. Though the QPT may have been validated in 20 countries with over 5,000 students, the 161 students in this study came from only three countries, and 159 of them came from just one country, Japan. Other commercially-produced tests that focus on reading, grammar, and vocabulary that were validated worldwide would probably produce no better results with this population. The students in this study had been sorted by their performances on the Center Test and the university entrance exam, but their scores on these two tests had no impact on their class assignments. There may be differences between these students in regard to their English language abilities, but administering the QPT-PPT did not generate data that could be used to sort these students in a fair manner. It would probably be much better for these students to be separated by their abilities in other, as yet untested, skills. Better yet, the skills tested should be linked to the goals and objectives of the curriculum.

Sadly, there are teachers and administrators who never question the reliability of test scores and instead accept raw and converted scores as perfect reflections of students’ abilities. Cut points are set and strictly followed, and decisions are made without any second thoughts. It is worrisome to think that some institutions may blindly use the results of commercially-produced tests like the QPT-PPT for placement decisions. Naive users of a test may ignore the guidance from the user manual, which advises institutions to use other assessment tools with the test, and may instead simply take the scores obtained by the administration, match the scores to the ALTE chart, and then separate students accordingly.

Even in the best situation, if the scores are reliable and other assessment tools are used, what information do the QPT-PPT scores really provide? What are the examinees’ strengths and weaknesses regarding vocabulary, grammar, and reading? What do they already know? What
do they need to learn? A commercially-produced test cannot provide the answers to these questions, but an in-house placement test based upon an existing curriculum would provide those answers and make the assignment of students to specific classes more logical and defensible.

Directly linking the placement test to the curriculum would also confer the advantage of having items written at the appropriate level for the students. Using a one-size-fits-all commercially-produced test is similar to buying one of the grab-bags available at Japanese department stores at the New Year. Upon opening the bag you usually find an item or two of value, but you also get items that you consider of little or no value. Some of the test items in this study were good, but more than half the test items were either too easy or too difficult for the students, and doubling the length of the test by combining both versions did not help much. While a longer test produced more reliable scores, it was not enough to ensure high reliability. Teachers would be better off writing a large number of test items, piloting them, analyzing the data, keeping the best items, and creating a short test filled with quality items instead of buying a test containing items of unknown value.

What is clear from this study is that test anxiety, poor test-taking skills, and unfamiliarity with the format of a placement test may result in the true ability of many students being underestimated, and that decreasing the influence of these variables can reduce score reliability. Teachers and researchers use tests to measure constructs, but often they measure constructs other than the intended constructs. In Administration A, the intended construct to be measured was general English reading, grammar, and vocabulary proficiency, but it amply appears that in fact the students’ test anxiety levels, test-taking strategies, and their familiarity with the test format were also being measured. Removing the data of just one confused student in Group 1 lowered the reliability estimate from 0.61 to 0.55. Providing instructions in the students’ native languages and model items with answers in the test booklets to show test takers how to answer the items might have prevented much of the confusion.

More studies on the use of commercially-produced tests and in-house tests for placement purposes at other Japanese colleges and universities are needed. Creating an effective placement test involves developing test items related to a true curriculum with clear goals and objectives, piloting the tests items, analyzing the data, and revising the tests to ensure that the scores are reliable and sound placement decisions can be made. This requires hard work, but it must be done if fair and defensible placement decisions are to be made.
Paul Westrick holds an M.Ed. in TESOL from Temple University. He is currently employed at Kyushu University in Fukuoka, Japan

Notes

1. In this study, CTT analysis is used instead of Rasch Analysis. Although many, if not most, language testers consider Rasch superior to CTT, the average language teacher may find CTT less intimidating and easier to understand. It also does not require expensive, specialized software. The focus of this study is score reliability and placement testing, and although Rasch analysis was conducted on the data, it is beyond the scope of this paper.

References


(Received December 29, 2003; accepted July 19, 2004)
Appendix A

The following are examples of items similar to those found on the actual tests. None of these are taken from either test.

**Items 1 – 5: Notices (signs, postings, etc.)**

- A. at a restaurant
- B. at a bank
- C. at a theater

**Items 6 – 20 and 41 – 50: Cloze passages**

**Crows**

Crows are very troublesome birds, but they are also very (6) ............ . They can be found throughout the country, and they are well known for (7)........... noisy calls. Sometimes they even attack people. People especially dislike their nasty habit of tearing (8)............. trash bags on the streets. While we (9).............. do not like them, we must admit that they are rather intelligent creatures. They have been known to (10)............ tools, something few animals can do.

6. A. clever  B. clevers  C. clevered
7. A. their   B. there   C. they’re
8. A. open   B. close   C. part
9. A. generally B. stealthily C. accidentally
10. A. use    B. uses    C. using

**Items 21 – 40 and 51 – 60: Individual sentences with a word or short phrase missing**

21. John went to Tom’s apartment to ............ the boxes and take them to Kim’s house.
   A. get    B. has    C. be    D. doing
Perspectives

Short-Term Overseas Study Programs: A Survey of Private Junior High Schools in Tokyo

Herbert E. Brauer
Hosei University, Second Junior High School

This survey research is the first large-scale study to provide a description of short-term overseas study programs implemented by private junior high schools in Tokyo. In addition to fundamental quantitative parameters, this comprehensive survey returned descriptive data from 84% of the private junior high schools in the Tokyo region with programs in 2001 and 2002. This descriptive data included types and details of activities undertaken, program and activity objectives, integration between the overseas study programs and school curricula, follow-up activities, and program evaluation. The survey revealed several innovative programs and activities implemented by these schools and identified areas that might benefit from further research.

In 1996, the Japanese Ministry of Education conducted a survey of international exchange and study abroad programs and activities implemented by public and private senior (although not junior) high schools in Japan. That survey examined only the broadest parameters of the programs, such as the numbers of schools and pupils partici-
pating in longer-term (3 months or longer) or shorter-term (less than 3 months) sojourns, the countries visited, and the disposition of Japanese pupils returning from the longer-term programs (whether they repeated a year, continued to the next year, or graduated) (Mombukagakusho, 1998). However, neither independent researchers nor the Ministry appear to have undertaken any large-scale descriptive studies focusing on such aspects of these programs as objectives, extracurricular activities, English as a second language (ESL) classes (content and hours), and accommodation (e.g., single or double homestay, dormitory stay).

Some researchers in Japan have conducted studies on the effects of particular short-term study abroad programs. Most of these studies have examined college programs and concerned themselves with linguistic proficiency, attitudes to the target language or language community, personal changes, and cultural understanding. For example, Higuchi, Saito, Lamarche, Shelangouski, and Kikuchi (1982), Kitao (1993), and Nozaki (1987) examined the effects of short-term (approximately one month) study abroad programs on Japanese junior college students. Higuchi et al. focused on changes in interpersonal values and images of native speakers in the target culture, Nozaki considered changes in attitudes towards the target culture, and Kitao examined attitudes towards, and images of, host nationals as well as linguistic proficiency. All three studies provided evidence suggesting that even short-term sojourns had measurable effects upon student attitudes. A study by Geis and Fukushima (1997) also examined a six-week intensive language study program arranged between a Japanese junior college and a major university in the United States. Because the students stayed in dormitories with their compatriots, the group as a whole remained relatively isolated and cohesive; it was not until the end of the sojourn that the Japanese students finally began to make friends with English-speaking students. The researchers concluded that any observed effects were more likely the result of intensive English lessons than the result of contact with the target culture and language community. On the other hand, among the few studies that have considered secondary school pupils, some have either looked at long-term (one-year) programs (e.g., Yashima & Viswat, 1991) or focused on individual student exchange within the framework of an international exchange program (e.g., Stitsworth & Sugiyama, 1990).

There does not, however, seem to have been any research at all into short-term “summer language programs” implemented for junior high school pupils 12 to 15 years of age. Although not as numerous as those offered at senior high school or college levels, programs for pupils in
the final year of junior high school may provide a particularly good opportunity for improving proficiency or attitudes towards English, which could, in turn, have a timely and positive impact upon language study in senior high school.

Some researchers have argued against the notion that there is either significant linguistic or cultural benefit from short-term overseas study programs (OSPs) (Day, 1987; Grove, 1983; Milleret, 1990). In particular, Grove claimed that programs lasting less than one month (a time frame that includes almost every short-term program implemented by the junior high schools in the current study, as well as the majority of programs at the senior high school level) not only fail to achieve their intended results, but may even be counterproductive and lead to undesired results. “Research and informed opinion suggest strongly that, in the case of a VSP (very short exchange program) (a) some important positive outcomes are not attainable, and that, (b) some undesirable outcomes are a likely result” (1983, p. 1). Moreover, in a previous paper, Grove discussed American Field Service programs, which include short-term sojourns for high school pupils (Grove, 1982). Therefore, it is reasonable to infer that his criticism would extend to VSPs organized for junior high school pupils. While Grove’s remarks refer primarily to changes in long-lasting attitudes towards people in the target culture and intercultural perspectives, both Day (1987) and Milleret (1990) refer specifically to circumstances that limit or restrict linguistic effects. In particular, Milleret, in her discussion of the difficulties involved in evaluating the outcomes of summer study programs, observes that “the living situation and location of the homestay may affect the opportunities each student has to experience linguistic and cultural growth” (1990, p. 486). Day reiterates Abrams’ (1960) observation about the short length of summer study programs, which “may entail superficial contact with the host culture [and] inadequate language practice” (1987, p. 261).

Educators in Japan offer a somewhat different perspective. Although their observations are subjective, most of the teachers who responded to the current survey were very positive about the linguistic and cultural effects of these programs. Echoing a similar belief, Shiozawa writes: “There’s no end to the benefits of overseas study or short-term study abroad trips. [Students] cultivate an objective perspective on Japan...develop confidence from relying on themselves...take notice of their insufficient communicative skills and return to Japan motivated to study English” (2001, p. 16) (translation by author).
Purpose

The purpose of this study is to describe short-term OSPs administered by private junior high schools in Tokyo. To the best of my knowledge, there are no formally published materials about how these programs actually operate, what they attempt to achieve, or how many participants visit which countries and for how long. Nor are there descriptions of the activities these programs offer and how these activities are organized to provide opportunities for naturalistic as well as instructed exposure to the target language.

The present study describes a survey conducted in May and June of 2002 of OSPs that were implemented the previous spring (March 2002) or summer (July, August 2001) by private junior high schools in Tokyo. While these schools may not be representative of private junior high schools throughout Japan, this is an extremely populous region (approximately 12 million people) with a correspondingly large student population (over 312,000 junior high school pupils). Moreover, since short-term OSPs conducted by junior high schools are almost exclusively the domain of the private institutions, and over 32% (75,711 in Tokyo, 233,553 nationwide) of private junior high school pupils in Japan attend schools in Tokyo (Mombukagakusho, 2001), the significance of those schools surveyed increases.

Instrument and Procedures

The questionnaire (Appendix I) contained 34 items arranged into six parts:

I. Respondents and schools
II. Time, duration, and destination of the OSP
III. Participating pupils
IV. Objectives, activities, and preparations for the OSP
V. Respondent’s opinions about the OSP and participating pupils, and
VI. Follow-up and evaluation of the OSP.

To conserve space, Part V questions, which concerned respondents' opinions rather than specific program features and characteristics, have been omitted from this paper.
The questions were created to take into consideration the major objectives and program features described in a variety of sources that have addressed the subject of study abroad for language learning. For example, Drysdale and Killelea (1982) provided a model for the organization of short-term language immersion programs for high school pupils and stressed the importance of proper orientation (predeparture) as well as follow-up activities. Milleret (1990) also mentioned orientation and reentry debriefing as well as the importance of homestays, and language and culture classes provided at the local institution as major program features. Brecht and Walton (1994) claimed that “the primary local conditions under which learning takes place in an SA [study abroad] program [are] preprogram preparation, in-country learning, and reintegration into the home institution” (p. 220). Within the context of “in-country learning,” they referred not only to formal instruction, but also, especially, to the opportunities that can arise from informal, unstructured interaction. They were also critical of inadequate integration and follow-up, noting that “(a)n aspect of SA that has been virtually ignored is how the SA experience can be enhanced upon reentry into the domestic program” (p. 223). Coleman (1997) examined study abroad programs among tertiary institutions and found shortcomings in key aspects such as preparation, curriculum integration, support while abroad, and overall program evaluation (p. 3). In an article describing the benefits of, and criteria for, a successful study abroad program, Shiozawa (2001) referred to the importance of clear program objectives, predeparture orientation, and the need to make the most out of accommodation arrangements. These observations and concerns of other practitioners and researchers in the field informed the creation of the questions that comprised the present survey.

The questionnaire was originally drafted in English. A nonspecialist Japanese native speaker translated the questions into Japanese and a senior administrator of a Japanese junior high school checked them for appropriateness of style and expression. The questionnaire was then piloted by three colleagues (Japanese teachers of English with experience in chaperoning pupils on overseas study programs) who provided feedback on wording as well as on the overall amount of time it took to complete the questionnaire. Items were reworded accordingly; however, the length of the questionnaire was not reduced, even though it was recognized that a complete response might require about 30 minutes.

Private junior high schools in Tokyo currently implementing OSPs were identified by referring to the Chugakko Juken Annai 2002 Ny-
Due to the relatively small number of schools with study abroad programs, it was decided not to sample this population but to request each of the 73 schools with programs to participate in the survey. Initially, either the principal, vice-principal or head teacher (kyotosensei) from each school was contacted by telephone to seek permission to mail the questionnaire. At that stage, only seven schools declined to participate, citing lack of time (four schools) or grounds of principle (three schools), and of the remaining 66 schools, five failed to return the questionnaire for various reasons. Thus, ultimately, 61 schools returned completed questionnaires. This yielded an extremely high response rate of 92 percent of questionnaires sent out and an overall participation rate of 84 percent of all schools among the target population.

Respondents, 27 of whom were full-time or part-time administrators (principals, vice-principals, head teachers) and 34 of whom were members of the English department (including 5 kokusai koryu tanto [students] who were officially in charge of their international exchange programs), returned the questionnaires by facsimile. In a few instances, follow-up contact was made to clarify ambiguous responses and to elicit more complete information.

Validity

Nearly all of the respondents (93%) had participated in SA trips on at least one occasion and a large majority (74%) on more than one occasion. They were therefore likely to be well acquainted with the planning and the outcomes of the programs. The existence of well-informed respondents in positions to provide expert opinions and observations about their OSPs supports an assumption of internal validity for this questionnaire. In other words, the data are meaningful and truly representative of the programs sampled, or to paraphrase Nunan, we are observing precisely what we think we are observing (1992, p. 81). As far as external validity is concerned, as mentioned above, no claim is made that the observations or conclusions of this survey research can be generalized beyond the defined population (namely, private junior high schools in Tokyo). Nevertheless, in the absence of significant differences between
private junior high schools in Tokyo and in other metropolitan centers in Japan, it seems not unreasonable to assume that similar results might be obtained from other private junior high schools in Japan. Content validity is supported in that most of the questions are directly related to OSP features that have been frequently mentioned in the literature. Further reference to those sources is made in the discussion section below.

Summary of Results

Respondents and Schools

The first group of four questions examined respondents’ experiences regarding OSPs, as well as some general characteristics of the schools, such as the number of pupils and the existence of special curricula (e.g., “English,” “international communication”) and whether there was any relationship between the existence of a special curriculum and an OSP. As mentioned above, results showed that respondents had participated in the OSPs on at least one occasion in 93% of the cases, and on several occasions in 74% of the cases. There was an average of 378 pupils (from a range of 26 to 964) in the schools with OSPs and 482 in schools without programs. Moreover, a frequency comparison between larger and smaller schools (more or fewer than 426 pupils respectively) yielded a 5.71 Chi-square (significant at \( p < .05 \)) indicating that smaller schools were more likely to have an OSP. A similar examination of schools by gender (boys, girls, co-ed) showed no significant difference in likelihood of having an OSP (Chi-square = 3.81). Most (95%) of the private junior high schools with OSPs implemented only a standard curriculum; however, one of the three schools with special curricula (international/English) was the school with the longest study abroad program (60 days).

Time, Duration, and Destination of the OSP

The second group of four items examined the broader “time and place” parameters of the program such as the number of years a program had been in existence, the destination countries, length of sojourns, and the times of year in which they took place.

Program histories averaged 12.4 years in length and ranged from 1 to 31 years (SD = 8.4 years). Of the 71 programs (nine schools had more than one program) only 15 had been in existence for more than 20 years, while 25 programs (35%) had been created within the preceding 5 years. The most popular destinations were the U.S., New Zealand, and the U.K.,
with 21, 16, and 15 programs, respectively, followed by Australia with 9 and Canada with 7. There was also one OSP to each of Thailand, Ireland, and the U.S. territory of Guam, respectively. Sojourns lasted an average of 16.5 days (SD = 5.4 days) and ranged in length from 5 to 29 days, with the exception of one program that lasted 60 days. The majority (70%) of excursions took place in the summer (50 trips), while 13 trips occurred in the spring (18%). The remaining 8 trips were in May, June, September, October, and November.

**Participating Pupils**

The five questions in this group of items addressed the age of the participants (school year), previous exposure to English (hours of instruction), whether English as a foreign language (EFL) conversation classes were part of the regular school curriculum, the number of pupils participating in the OSP, and the existence of a special selection process for participation.

In 47 out of 71 cases (66%), participation was limited to third-year students (aged 14 and 15 years). Sixteen OSPs (22%) combined 2nd- and 3rd-year pupils, and in one exceptional case with a high percentage of “returnees,” a school implemented the OSP exclusively for first-year pupils. The numbers of participating pupils varied from as few as 5 to as many as 164 (AVG = 38, SD = 39). In only five schools was there a selection process to screen students for participation on the basis of English proficiency.

Every school offered “conversation classes” but, depending upon the school year of the participants, time of departure, and school curriculum, there was a lot of variability in estimated hours of total EFL instruction (grammar, reading, and conversation) before departure, ranging from 234 to 800 hours (with an average of 462).

**Objectives, Activities, and Preparations for the OSP**

This section comprised eight questions concerning classification and objectives of the OSPs, predeparture orientation, integration between the OSP and the EFL program (in Japan), school visits, homestay arrangements, ESL lessons (in the host country), and support for pupils while they were in a homestay situation.

According to the Obunsha Guide to Junior High Schools, OSPs are classified as “international exchange” (kokusai koryu), “study abroad”
(ryugaku koryu), “language study” (gogaku kenshu), “overseas study” (kaigai kenshu), and “graduation trips” (shugaku ryoko). While the first two are fairly general terms as far as short-term junior high school study abroad is concerned, language or overseas (culture) study imply more specific objectives, and graduation trips are intended primarily to cap off and commemorate the junior high school experience. The results for classification showed that “language study” was the official classification for 37 (52%) of the OSPs surveyed (20 programs were classified as “language study” only, while 17 programs combined this with another classification). Eight schools classified their OSPs as “graduation trips,” which is, at least ostensibly, related to a specific objective such as the learning of culture or language. However, when queried about specific objectives, respondents considered “language ability” the number one priority in only 29% of the cases, “cultural understanding” most often (58% of the cases), and “school experience” hardly at all (1.2%).

Most OSPs included visits to schools in the host country (48 out of 71 cases), and among these, 38% offered participation in Japanese as a foreign language (JFL) classes, 40% arranged special ESL classes, and 33% assigned “buddies” (multiple responses were possible, therefore the total does not add up to 100%). Forty-three percent of schools claimed to integrate their EFL and OSP programs, employing host country-related materials, both authentic and specifically classroom-designed, such as photos taken by pupils on previous trips and audio materials, as well as planning lesson content relevant to the OSP.

With regard to predeparture orientation, 46 out of 61 schools (75%) indicated that they provided some form of orientation to OSP participants, although five schools restricted this support to extra conversation practice, leaving only 41 out of 61 schools (67%) providing extra-linguistic predeparture information. Specifically, orientation to life in home-stay families (in applicable cases), explanation of local customs, or familiarization with local dialects and pronunciation were omitted in approximately 30% of the cases.

Homestay was a feature in 59 of 71 OSPs (83%), with 86% of respondents citing cultural, and 51% linguistic benefits as the main reasons for including homestay (multiple response). Ease of supervision or cost savings were reasons in only 4% and 6% of the cases, respectively. Pupils were billeted or housed singly in 73% of homestays, reflecting the preference of the home institution in most cases. Reasons for preferring to billet their students singly rather than in pairs were fairly evenly distributed among “personal growth,” “culture learning,” and “language learning.”
Only about half (52%) of the homestay OSPs reported efforts to provide special guidance to host families, although 14% “didn’t know” (suggesting that this might have been organized by the hosting side without the knowledge of the Japanese schools).

Special ESL lessons were provided in the host country in 58% of all OSPs, but in 75% of those OSPs officially classified as “language study” trips. These classes were an average of 75 minutes a day (including weekends; weekday averages would have been proportionately higher) with considerable variation; one OSP arranged more than three-and-a-half hours per day, while some others only scheduled two hours (possibly only one or two sessions) during a 20-day sojourn. Moreover, in most cases (68%), ESL lessons were “not at all” integrated with the EFL program in Japan.

Follow-up and Evaluation of the OSP

The final section of the survey asked about what happened after the pupils returned home to Japan. These questions considered pupils’ follow-up activities as well as the organizers’ own program evaluations. The last three questions invited comments regarding recent changes to the OSPs, special characteristics or achievements of the OSPs, or any notable problems or difficulties.

Upon their return to Japan, pupils completed “follow-up” assignments or activities in 66 of the 71 OSPs surveyed. Communicating with their homestay families (via the post or e-mail) and writing reports (in English or in Japanese) about their experiences were the most commonly cited (80% and 79%, respectively). In 48% of the cases, the trip was featured in the school’s annual culture festival, although it is not clear to what extent that was a teacher- or pupil-centered activity. In a further 29% of cases, teachers reported using the OSP experience either as the basis for a lesson plan or for reviewing the OSP in class.

“Evaluation” is less a feature of the OSP itself than of its administration. This item was included on the questionnaire because evaluations can provide schools with valuable data about how successful the programs are in meeting stated objectives and therefore better enable administrators to make informed decisions about improving their programs. Furthermore, the very existence of evaluations themselves might reflect a sincere desire on the part of administrators to derive the greatest benefit for the participating pupils and possibly for other stakeholders (e.g., schools, chaperoning teachers, host families) as well. As reported above, 66% of
the OSPs conducted some type of program evaluation, most commonly student and parent feedback, both of which could be said to belong to the “consumer satisfaction” variety of evaluation. These were reported in 79% and 55% of the cases, respectively (multiple responses) in which an evaluation took place. Such feedback is important because the survival of the program ultimately depends on whether it satisfies the participants and other stakeholders. Feedback from host families was obtained in only 19% of the OSPs with homestays. It may be the case, however, that the Japanese schools entrust the gathering of feedback from host families to affiliated schools or organizations in the host country.

Apart from participant feedback evaluations, two types of analytic evaluations were reported: pupil pre- and post-linguistic proficiency tests and attitude surveys. Of these, attitude surveys were reported in 50% of the cases that implemented evaluations. Linguistic proficiency tests were only reported by 4 out of 49 schools (8%) that conducted evaluations and for only 6% of the 71 programs overall, certainly not in keeping with the 59% of OSPs that claimed to put a high priority on linguistic objectives.

**Discussion of Major Results**

The following discussion will focus on the “primary features” of OSPs, which are directly related to the amount and variety of linguistic or cultural exposure that the pupils receive, and which are therefore liable to have a direct effect upon the achievement of linguistic or cultural objectives. These four features are addressed in Question 8 (duration of the sojourn), Question 16 (visits to a local school), Question 19 (homestay), and Question 20 (ESL classes and their content integration).

**Duration of sojourn**

Duration of the sojourn, in particular, is an issue that strikes at the very heart of summer study exchange programs. Indeed, some scholars may question whether programs of short duration have any educational (linguistic or cultural) merit in the first place. Grove (1983), in reference to exchange programs that last shorter than one month, asserts, “Such exchanges . . . hold little promise of accomplishing those high goals that most of us in the business of intercultural exchange hold dear” (p. 1). These concerns are echoed by Milleret (1990), who explains, “shortness of the summer study experience can limit student contact with the host
culture [and]...the opportunity for language practice” (p. 483). Milleret, it should be noted, seems to be referring to programs lasting 5 to 6 weeks. Elsewhere, as cited above, Day (1987) also observes that one of the drawbacks of short programs is insufficient language practice or contact with the host culture.

However, for pupils in junior high schools surveyed for this study, the duration of summer OSPs is seldom a matter of 6 weeks, nor even an entire month; with one notable exception, every one of the OSPs examined lasted less than four weeks, and the average length was only 16.5 days. There was also much variability, with 12 programs lasting 10 days or less, and 25 programs lasting 3 weeks or longer. This raises several interesting questions: What accounts for the variability? Is there a relationship between participant satisfaction and duration? What sort of objectives can be achieved with shorter programs and is there a relationship between changes in attitude, for example, and length of sojourn even within the context of these very short OSPs? What are the costs and benefits of increasing program duration? While various constraints will inevitably continue to restrict the duration of summer study abroad programs for junior high school students, it is important to examine these programs more closely in order to discover what objectives can be reached, and how better to achieve these objectives within the time frames available.

**School visits**

Attending school in the host country, either on the basis of broader sister-school relationships or other arrangements with a host institute, is a common feature. While observing regular classes at a school can be very challenging for limited English proficient pupils, there is clearly potential for various types of linguistic exposure. In their study of Australian long-term (one year) high school exchange students in Japan, Marriott and Enomoto (1995) observed that “the school domain also provides...various opportunities for daily...communication and interaction” (p. 69). Elsewhere, Marriott has commented on the effects upon language learning that accrue to students who “participate in focused instruction in class as well as being immersed in the natural environment” (1995, p. 197). It is not clear that identical opportunities will exist for pupils on short-term visits, but certainly increasing the variety and amount of exposure to the language can only contribute to a richer, and therefore more beneficial, experience.
It is not simply a question of whether or not to visit schools and observe classes, but of which classes and what sort of interaction occurs there. Marriott and Enomoto (1995) explain that while pupils had great difficulty coping with the content of most subjects, there were fewer difficulties in subjects such as English, physical education, or mathematics. For Japanese pupils overseas, classes in JFL, if they exist, might offer opportunities for active participation. The present survey revealed one interesting approach in which the visiting (Japanese) pupils acted as teaching assistants in JFL lessons for the pupils of the hosting school. It also revealed another innovative feature; some pupils visited an elementary school because it was believed that the linguistic challenge would be more suitable for their level of proficiency. Such program innovations should be explored in greater detail. Comparative program evaluations could be undertaken for the benefit of all those concerned with improving existing OSPs or planning new ones. Case studies utilizing qualitative research methods, including participant observation or interviews with teachers and program designers, could shed light upon activities that might be arranged to supplement or support classroom observation and what their potential benefits might be.

The survey also revealed a relationship between visits to southern hemisphere countries and the existence of school visits as OSP features. Currently, many of the excursions to the U.S., Canada, and the U.K. (according to this survey, 33 out of 42, or 79%) take place in the summer when schools in the host country are generally not in session. This suggests that scheduling excursions to northern hemisphere countries for the late winter or early spring would afford additional opportunities to coordinate OSPs with school visits and thus help achieve program objectives. By that time, pupils would also have had more hours of English instruction and there would be fewer conflicts with club activities, particularly sports clubs, many of which have important tournaments in the summer, as one of the respondents noted. While this may be of less interest to researchers, program administrators might wish to consider such alternative schedules.

**Homestays**

In an evaluation and review of summer study programs in 1990, Milleret reports that homestays are among “the major features of a summer language study abroad program that have been standardized through wide acceptance and implementation at both the high school and col-
While there is, in fact, research that questions the value of a “homestay effect” (Freed, 1990; Rivers, 1998), learners themselves often feel there is benefit from the homestay context. Referring to Australian high school exchange pupils in Japan, Marriott and Enomoto (1995) noted that “exchange pupils themselves commonly claim that it is in the home context where their opportunity to acquire Japanese is maximized” (p. 68). In yet another survey of Australian learners of Japanese, Hashimoto (1994) reports that “eleven out of twenty pupils chose host families as the most helpful factor in their language acquisition, and seven pupils out of the remaining nine rated them second in importance. Consequently, most pupils considered the homestay to be very influential in relation to their language development” (p. 4).

Homestay is present in 83% of the programs examined in this survey. The extent to which this crowning feature is effective in promoting increased linguistic proficiency, improved attitudes to the target language or target language community, or greater cross-cultural understanding, particularly with regard to very short-term programs for junior high school students, deserves further attention. The author is currently engaged in a study employing quantitative and qualitative methods to examine the effect of homestay upon listening comprehension and to describe the characteristics of that environment and the opportunities it provides for verbal interaction. That study will also address the potential for affecting interaction between homestay pupils and their host families through the appropriate use of communicative homework tasks. While it may turn out that a salient feature of this environment is its inherent variability, this may argue all the more for the development of homestay tasks or assignments designed to stimulate a certain basic level of verbal interaction. The development and analysis of such “homestay communicative tasks” itself is a potentially useful avenue for research.

**ESL lessons**

Just over half of all of the OSPs, but 75% of those that were officially classified as “language study,” featured ESL classes. Milleret (1990) identifies language and culture classes taught at a local institute as one of the six standard features of summer language study abroad programs (p. 484). But the issue is not so much classroom learning itself as it is the role and function of these ESL classes within the specific context of study abroad, and particularly as a means of maximizing the opportunities that pupils may derive from naturalistic exposure to language resulting from home-
stay. There should be a synergy between the more structured approach of classroom learning and the opportunities for genuine communication that arise out of homestay or other out-of-class environments in the host country. In other words, “a more porous classroom is required for foreign pupils studying in-country, one that guides more structured language learning and practice outside of the classroom...[and] the classroom should present opportunities to build upon the learning taking place outside” (Brecht & Walton, 1994, p. 222).

Integration of the OSP, in general, and of ESL lessons (those provided in the host country) in particular, with the EFL program in Japan, are issues raised in Questions 17 and 20. This is not to suggest that these programs cannot be valuable and effective in their own right, rather, it is a question of missed opportunities. While the importance of adapting ESL lessons to the levels and needs of the pupils is beyond question, how much more effective might these lessons be if they maximized the opportunities for naturalistic exposure that the pupils encounter in their homestays? How much more practical, meaningful, and interesting would EFL lessons be if they integrated themes, needs, and situations arising out of OSPs that the pupils know their seniors have participated in and in which there is a distinct possibility that they, or their fellow pupils, might also join, in the future? Such improvements to EFL classes would benefit all the students, not only those who, for various reasons, were fortunate enough to participate. In this respect, it is disappointing to observe that the majority of schools (57%) reported no integration at all between the OSPs and EFL programs and that with respect to integration of ESL lessons with EFL programs, 8% of the OSPs showed no integration at all.

At the other extreme, ESL classes that merely repeat classroom experiences that pupils have already had in their home country, albeit with a different teacher and with a group of (hopefully) more highly motivated pupils, may serve relatively little purpose. If the lessons are tiresome, they may even be counterproductive. If they fail to provide unique learning opportunities, to stimulate verbal interaction within or reflection upon the homestay environment, and if their major value lies only in their face validity, then some important opportunities are being squandered. The evaluation of ESL classroom experiences that are offered as a part of OSPs also deserves study. How do these classes differ from the EFL classes that are offered in Japan? What can the teachers of those respective classes learn from each other? How can they reinforce each other’s efforts? Is there a way to apply the experiences of the OSP to the EFL classroom so that all students can benefit from more interesting and meaningful classes?
Conclusion

This comprehensive survey of junior high school SA programs in the Tokyo metropolitan area reveals a number of salient characteristics: Most of these programs are directed towards northern hemisphere countries, involve mainly 3rd-year pupils, take place during the summer months (July and August), and have an average duration of slightly more than two weeks. The stated objective for these programs is usually linguistic, although most informants feel that cultural objectives are at least, if not slightly more, important. Homestay, which is believed to support both linguistic and cultural objectives, figures prominently as the preferred style of accommodation. Program evaluations are largely limited to attitude questionnaires and pupil feedback and there is little effort to measure linguistic effect or cross-cultural awareness.

This paper has focused upon SA programs provided for junior high school pupils. As pointed out in the introduction, this age group has not been the subject of much research in the past. It is an important age group to study for at least three reasons. First, the relatively low level of English proficiency suggests that it may be easier to both achieve and measure changes in linguistic proficiency as a result of the SA experience than with older age groups. Providing evidence of a proficiency effect has been a difficult challenge for short-term study abroad, but there may be opportunities with junior high school OSPs, precisely because the participants are less proficient and thus are facing a steeper learning curve. Second, because these young pupils are at an impressionable age, there may be a greater potential to observe beneficial changes in attitudes towards the target language and culture, as well as increased integrative orientation. Whether these effects actually occur, and whether they are lasting effects, might also be fruitfully researched. Crosssectional attitude surveys comparing participants with non-participants over a 6-year period spanning junior and senior high school might reveal interesting results. Finally, this may be a particularly opportune time to bring about improved attitudes or integrative motivation since the pupils are at the threshold of their final three years of required English education. If the impact of this experiential education is such that the pupils approach those important remaining years with greater confidence, interest, and enthusiasm, then this result will go far towards justifying the continuation and improvement of these programs.
Herbert Brauer has taught EFL in Japan for over 15 years and is a doctoral candidate in Applied Linguistics at the University of Birmingham.

References


(Received August 30, 2003; accepted September 1, 2004)
Appendix I

Questionnaire to private junior high schools with OSPs

PART I: ABOUT THE RESPONDENT AND THE SCHOOL
(1) How often have you accompanied the pupils on a study abroad trip to an English speaking country?  
   ___ (a) once   ___ (b) more than once   ___ (c) never
(2) How many pupils are there in your school?  
     __________ boys   __________ girls
(3) What type of course or curriculum do you offer pupils at your school? (multiple response)  
     (a) standard  (b) international / English  (c) other _____________
(4) In what course or curriculum are pupils who participate in overseas study programs enrolled?  
     (a) standard  (b) international / English  (c) other _____________

PART II: TIME, DURATION, AND DESTINATION OF THE OSP
(5) When did your school first implement this study abroad program?  
    Year ______________
(6) What was the destination country of your most recent trip?  
    ________________________
(7) In which month does the overseas program take place?  
    ________________________
(8) What is the duration of the trip? _______________________ days

PART III: PARTICIPATING PUPILS
(9) What school year are these participating pupils? (choose all that apply)  
     (a) first year  (b) second year  (c) third year
(10) How many hours of English instruction do your pupils receive per week (grammar, reading, conversation).
(a) in first year _____ hours  (b) in second year _____ hours  (c) in third year _____ hours

(11) Does your regular school program offer conversation classes? 
YES  NO

(12) How many pupils participated in this program last year? _______

(13) Is there a special process for selecting the pupils?
___ (a) No. All pupils are accepted.
___ (b) Only if number of applicants exceeds spaces.
___ (c) Yes, to ensure a minimum level of language skills.

PART IV: OBJECTIVES, ACTIVITIES, AND PREPARATION FOR THE OSP

(14) Which classification best describes your overseas program?
   (a) overseas study   (b) linguistic study (c) graduation trip
   (d) international exchange (e) sojourn exchange

(15) What are the program objectives? (rank as: 1 for most important, 2 for next most important, etc.)
___ (a) to provide a memorable school experience
___ (b) for better understanding of foreign culture
___ (c) to improve English ability
___ (d) for personal growth and development of the pupils

(16) Do your pupils visit a junior or senior high school at the destination?  YES  NO

(16a) IF YES, which of the following are true (check all that apply)
___ (a) your school has a formal "sister school" relationship with this school
___ (b) your pupils observe (audit) some regular classes at that school
___ (c) your pupils attend (participate) in some Japanese language classes
(d) your pupils attend special ESL classes given at that school
(e) your pupils are assigned a "buddy" from the host school
(f) your school sponsors reciprocal visits to Japan for pupils from that school

(17) Is your regular English program integrated with the study abroad program?  YES  NO
(17a) IF YES, in which ways: (check all that apply)
(a) audio materials for listening practice related to the study abroad destination
(b) pictures / writings in classroom or halls from pupils who participated
(c) videos of school, lifestyles, culture etc. related to the study abroad destination
(d) lesson content is related to the host country or study abroad
(e) authentic materials (maps, menus, etc.) from host country used in lessons
(f) other ________________________________

(18) Do you provide special orientation for pupils selected to participate in the visit?  YES  NO
(18a) IF YES, do you... (Check all that apply)
(a) explain family, social customs of the host country?
(b) familiarize pupils with features of typical English pronunciation or dialect at the destination
(c) prepare individual pupils about their particular host families (in case of homestay)?
(d) give extra conversation practice by reading dialogues
(e) give extra conversation practice by memorizing words or phrases
(f) give extra conversation practice by doing role play or simulations

(19) Does your program provide arrangements for homestay visits?
   YES     NO

(19a) If yes, why do you provide such arrangements?
   (main reason - choose one only)
   ___ (a) more economical
   ___ (b) difficult to arrange other types of accommodation
   ___ (c) better for learning English
   ___ (d) better for learning culture
   ___ (e) easier to supervise pupils
   ___ (f) other ______________________________________________

(19b) If yes, how many pupils stay with a single family?
   ___ (a) one      ___ (b) two
   ___ (c) combination ___% 1 pupil; ___% 2 pupils

(19c) If yes, which type of accommodation arrangements do you think is best, and why?
   (a) homestay (1 pupil) (b) homestay (2 pupils)  (c) dormitory
   (d) other __________
   Reason:_____________________________________________________

(19d) If yes, do host families receive guidance on communicating with low English proficiency guests?
   (a) YES   (b) NO   (c) DON'T KNOW

(20) Do the pupils receive special English lessons while they are abroad? YES NO

(20a) IF YES, how many hours instruction (total) ____________ hours

(20b) IF YES, how closely does lesson content integrate with your regular English program?
___ (a) very closely - lesson review or build directly upon our own lessons
___ (b) somewhat - lesson materials are designed to match the level and content of our classes
___ (c) not at all - there is no particular consideration to our syllabus and lesson content

(21) Do you offer your pupils personal or linguistic support during their sojourn? YES NO

(21a) In what ways do you support them? (check all that apply)
___ (a) you contact the host families on a regular basis
___ (b) the pupils can contact you if they have difficulties
___ (c) the pupils meet their accompanying teachers on a regular basis
___ (d) pupils are given a telephone number to use in an emergency
___ (e) you have regular problem discussion / solving sessions with the pupils
___ (f) other ____________________________

PART V: RESPONDENT’S OPINION ABOUT THE OSP AND PARTICIPATING PUPILS

(22) Our pupils are mature enough to benefit from the cultural experience
    (a) agree strongly (b) agree somewhat (c) neutral
    (d) disagree somewhat (e) disagree strongly

(23) Our pupils are mature enough to benefit from homestay (omit this if you don’t use homestay)
    (a) agree strongly (b) agree somewhat (c) neutral
    (d) disagree somewhat (e) disagree strongly

(24) The sojourn is long enough to achieve a noticeable effect upon their ability to speak and understand English.
(a) agree strongly  (b) agree somewhat  (c) neutral  
(d) disagree somewhat  (e) disagree strongly

(25) The sojourn is long enough to achieve a significant effect upon their attitudes toward host country / culture.

(a) agree strongly  (b) agree somewhat  (c) neutral  
(d) disagree somewhat  (e) disagree strongly

(26) These study trips are an effective way of improving pupils' attitudes towards the target culture / language.

(a) agree strongly  (b) agree somewhat  (c) neutral  
(d) disagree somewhat  (e) disagree strongly

(27) Homestay is an effective way of improving pupils' listening and speaking skills.

(a) agree strongly  (b) agree somewhat  (c) neutral  
(d) disagree somewhat  (e) disagree strongly

(28) ESL lessons in the host country (as part of our program) are effective in improving listening and speaking skills.

(a) agree strongly  (b) agree somewhat  (c) neutral  
(d) disagree somewhat  (e) disagree strongly

(29) Attending regular school classes in host country as observers effectively improves listening and speaking skills.

(a) agree strongly  (b) agree somewhat  (c) neutral  
(d) disagree somewhat  (e) disagree strongly

PART VI: FOLLOW-UP AND EVALUATION OF THE OSP

(30) Do the pupils follow up their study abroad experience?

YES       NO

(30a) IF YES, how do they follow it up? (check all that apply)

___  (a) pupils write reports about their experiences

___  (b) pupils correspond (email or post) with friends in the host country

___  (c) trip is reviewed in English classes

___  (d) trip is a highlight of the annual Culture Festival
__ (e) experiences from the trip help create ideas for lesson plans or teaching points

(31) Have you attempted to evaluate your program? YES NO

(31a) IF YES, which instruments have you used to perform the evaluation? (check all that apply)

__ (a) student feedback or evaluations
__ (b) parent feedback or evaluations
__ (c) host family feedback or evaluations
__ (d) host institution feedback or evaluations
__ (e) student before and after proficiency tests
__ (f) student before and after attitude surveys

(32) Have there been any major changes in your program recently?
(e.g. number of students, period of sojourn, number of students electing solo homestay) Please explain.

(33) Are there any particularly successful features or special characteristics of your program? Please explain.

(34) Are there any particular problems you have encountered? Please explain.

Thank you very much for your time and effort.

The original Japanese-language version of this questionnaire is available from the author at heb2278@yahoo.com
Reviews


Reviewed by
Debra L. Simms

The Handbook of Bilingualism is one of the latest offerings in Blackwell Publishing’s well-received series in linguistics. At 884 pages, it represents a significant investment of time. The editors, Tej K. Bhatia and William C. Ritchie of Syracuse University, have compiled what is likely the most comprehensive treatment of bilingualism available in a single volume. There are 31 well-written and thoroughly referenced chapters, each one authored by an internationally recognized researcher. The chapters are organized under four broad subject areas: Overview and Foundations, Neurological and Psychological Aspects, Societal Bilingualism and its Effects, and Global Perspectives and Challenges.

The shortest of the four sections, “Overview and Foundations,” introduces basic concepts in the study of bilingualism. Distinctions are made, for example, between additive and subtractive bilingualism and between primary and secondary bilingualism. Readers are also introduced to the idea of monolingual bias, wherein monolingualism is seen as the norm while bilingualism is seen as aberrant. Newcomers to the study of bilingualism will find these introductory chapters engaging and informative.

Whereas the “Overview and Foundations” section is appropriate for the general reader, the same cannot be said for “Neurological and Psychological Aspects.” This section deals with bilingual brain function, the idea of innate grammatical knowledge, the interaction between two grammatical systems, the roles of memory and emotion, and the complexities of code-switching. Brain science aficionados and die-hard grammar lovers will thoroughly enjoy getting up to speed on the most recent thinking in what many consider to be “hardcore” linguistics. Casual readers, on the other hand, may possibly be intimidated by or uninterested in this subject matter.
The remaining two sections, “Societal Bilingualism and its Effects,” and “Global Perspectives and Challenges,” take the study of bilingualism away from mathematical formulas and medical models and place it in the context of the social world. The social and political dimensions of bilingualism are explored here from a wide variety of perspectives including gender, race, class, and culture. And while bilingualism which includes English does predominate, there are also enlightening articles that deal with languages other than English. Ad Backus, for example, describes what has been happening to Turkish as it makes contact with German, French, and Danish via immigration, while Bhatia and Ritchie skillfully demystify the Hindi-Hundustani-Urdu-Punjabi core/axis in multilingual India.

Choosing which articles should go into a project like this is no easy task. Given that *The Handbook of Bilingualism* already weighs over 1 kilogram, it is perhaps unfair to suggest that something has been left out. Nevertheless, one might hope to find at least one chapter entirely devoted to the arguments around the global spread of English and the controversial idea of linguistic imperialism. Linguistic imperialism does get minimal attention from Bhatia and Ritchie in “Bilingualism in the Global Media and Advertising” and more substantial treatment in “Bi-/Multiculturalism in Southern Africa” by Nkonko M. Kamwangamalu. Nancy Dorian provides an excellent long-range view of language change in “Minority and Endangered Languages.” Still, an article that deals directly and specifically with the social and political ramifications of the unprecedented spread of English from the perspective of an expert on bilingualism would have been most welcome.

Overall, *The Handbook of Bilingualism* is an excellent compilation by able scholars and a fine addition to Blackwell’s series.

Reviewed by
Justin Falkus
Kanda University of International Studies

Teaching Listening in the Language Classroom is one of the latest additions to the RELC Portfolio Series, designed as “practical resource booklets” (p. i) for language teachers. As the term “booklet” implies, this is a small, thin publication running to only 55 pages. Its stated aim is to “provide background and theoretical information about listening, and offer suggestions for planning listening tasks and lessons” (p. iii). The author, Christine Goh, certainly achieves this, though at a cost.

The layout of the book is simple and clear. There are five chapters, each with a short overview in which the main points of the chapter are highlighted, and each with a single-paragraph conclusion. In between, the main points are treated simply and systematically. Each chapter finds space for several “Tasks” which encourage the reader to reflect on issues raised in the book and relate them to the classroom. The book concludes with a page of references and three useful appendices.

Chapter 1 is a six-page theoretical introduction to the topic. Necessarily brief, Goh nevertheless manages to give readers an informative and fairly wide-ranging survey of such factors as purposeful listening, cognitive models of listening comprehension, bottom-up and top-down processing, and listening strategies. The references provided are useful gateways to exploring these areas in detail greater than this book permits.

Chapters 2 and 3 are the heart of the book. We are given short descriptions of about fifteen language tasks as well as sample procedures for these tasks, summaries of listening responses, ideas for pre- and postlistening activities and, most usefully, four lesson outlines. These outlines help make sense of the ideas raised in the first three chapters and are excellent templates for readers to use when lesson planning.

Chapter 4 encourages teachers to raise their students’ metacognitive awareness and suggests several ways of doing this, including keeping a listening diary. In this reviewer’s opinion, Chapter 5, on authentic listening materials, is the weakest. The ideas presented—such as “you
should] screen songs for inappropriate language and unsuitable content” (p.44)—are not particularly interesting nor new. The book would have been better served with Chapter 5 removed and extra space given over to earlier chapters.

Throughout *Teaching Listening* Goh writes well and efficiently, but some serious weaknesses must be mentioned. First, much of the information is presented as a series of lists (five categories of..., ten types of...), and the descriptions of some points are so short as to be bare summaries. The result is an absence of clarifying examples. Thus we are told that “the concepts of bottom-up and top-down listening and the interactive nature of these processes are crucial to an understanding of how language learners listen” (p. 7), but no example is given of bottom-up and top-down processing at work. Then, in Task 1.3 (p. 7), we are asked to consider the difference between bottom-up and top-down processing and think of cases where one would take precedence over the other. This is very difficult for anyone for whom these abstract terms have no concrete application.

This point leads to my second criticism. Generally speaking, the tasks are appropriate, stimulating, and of great practical benefit. Task 2.4, for example, asks teachers to adapt one-way listening tasks so that they become two-way tasks—that is, those where students are interacting in pairs or small groups rather than listening to a recording or a teacher (p. 24). This is a very instructive activity, but the problem is that Goh does not provide any suggestions of her own to confirm, guide, or redirect our ideas. This can be quite unmotivating, and given that *Teaching Listening* is not a textbook (readers, that is, will probably be left to their own devices) and that it is designed to be “practical and technique oriented” (p. i), this seems a serious flaw.

Finally, the lack of a glossary means that terms such as “co-text” (p. 6) and “micro-markers” (p. 8) remain unexplained. Readers are unable to guess the meaning of new terms from context for the reasons mentioned above.

*Teaching Listening* is a stimulating book, and I applaud Goh for achieving so much in so few pages. Had the limited space been more thoughtfully allocated, thereby adding to the book’s accessibility and usefulness, my recommendation would have been unreserved.

Reviewed by
Andy Maggs
Tokyo Woman’s Christian University

This title is one in a series of booklets designed for ESL/EFL teachers as short practical resources to aid teaching. The booklets have these main characteristics: “They are practical and technique-oriented; they are written in an accessible, nonacademic style; they focus on both principles and procedures” (p. i). The rationale behind Managing Vocabulary Learning is based on current research in second language pedagogy, and with this in mind, the author aims to cover those vocabulary topics that best provide teachers with “the basis for an understanding of the major principles lying behind an effective vocabulary program” (p. iii).

This booklet is divided into seven chapters, each with a set of tasks for the teacher to complete. The tasks are included to increase understanding of the main elements of each chapter. There are also four appendices: the first appendix is an answer key to the chapter tasks; the second appendix contains a Vocabulary Levels Test, one way for teachers to gauge the current vocabulary knowledge of students; the third appendix is a corpus-derived Academic Word List; and the fourth appendix contains the General Service List of the first 2,000 most frequently used words in English, alphabetically arranged.

Chapter 1 deals with how to balance a language course from a vocabulary perspective. Nation identifies four strands for an effective language course: meaning-focused input, language-focused learning, meaning-focused output, and fluency development. He outlines the following areas: the conditions that need to occur for each strand to work effectively; the vocabulary coverage, skills, and strategies learners need; and the activities and techniques teachers need to implement in order to facilitate effective vocabulary learning. Chapter 2 covers how to select the words to teach according to frequency and range. There is an explanation of the principles and procedures behind the teaching of high- and low-frequency words, academic words, and technical words.

Chapter 3 focuses on measuring learners’ current vocabulary knowledge. Nation explains the Vocabulary Levels Test (Appendix 2 in the booklet), how to interpret it, and how to apply the results. Chapter 4
deals with teaching unfamiliar words in intensive reading texts. The author explains briefly the key principles behind vocabulary learning, lists teaching techniques, and, in the form of a task, asks the teacher to justify the reasons for using a particular technique to explain an unfamiliar word. Chapter 5 focuses on how to plan, in a principled fashion, an extensive reading program using graded readers at the beginner and intermediate levels. It includes key headings for the curriculum designer to consider, such as content and sequencing, format, monitoring, and assessment.

The final chapters are rather more specialized: Chapter 6 considers ways to train learners in guessing strategies for unknown words in context. Chapter 7 focuses on teaching specialized vocabulary to learners of English for special purposes, for example, those taking academic or vocational courses. It introduces a corpus-derived Academic Word List of 570 word families that are not in the first 2,000 most frequently used words in English (this list is claimed to cover 8.5 to 10% of the running words in academic text). The author first explains how the list was constructed and then advises that it be covered in a language course in all four strands, in addition to being learned by rote.

This booklet is concise, clearly laid-out, and very simply written. It can be recommended for new teachers, both native speakers of English and non-native, and for teacher-trainers as an excellent basic principles and procedures guide to approaching vocabulary learning. The appendices are all very practical and useful for teachers. However, there are some notable content omissions: there is no mention of extensive reading for advanced learners and no mention either of the importance of teaching vocabulary in chunks, such as collocations and set phrases. In addition, most of the suggested techniques will be familiar to experienced teachers.

Reviewed by
Paul Lyddon
University of Arizona

This three-part volume brings together 14 papers covering a wide range of perspectives from among those presented at the 37th SEAMEO RELC International Seminar of April 2002.

Section I (Materials Design and Evaluation in Language Teaching) comprises five articles, beginning with Richard R. Day’s “Authenticity in the Design and Development of Materials,” a thought-provoking challenge to the superiority of authenticity and a plea for a reconsideration of appropriateness. Next, Brian Tomlinson’s “Humanising the Coursebook” poignantly underscores the primacy of accommodating learners’ affective needs and suggests several practical ways of adapting to individual situational constraints. Then, in “Materials for New Technologies: Learning from Research and Practice,” Denise E. Murray discusses considerations for designing and implementing computer-assisted language instruction, for which Beverly Derewianka subsequently provides a practical example in “Designing an On-Line Reference Grammar for Primary English Teachers,” a description of the TeleNex project in Hong Kong. The section concludes with A. Mehdi Riazi’s “What do Textbook Evaluation Schemes Tell Us? A Study of the Textbook Evaluation Schemes of Three Decades,” which describes eight rubrics that have been proposed since the 1970s, situates them within the prevailing methodological frameworks of their time, and suggests implications for textbook appraisals today.

Section II (Methodology and Text) contains three articles, starting with Erwin Tschirner’s “Skill, Text, and Register: Rethinking Grammar in the IT Age,” an incisive reconceptualization of grammar from an information processing perspective. Next, in “Developing Academic Texts to Enhance Inference Use,” Ronald L. Brown argues for the need to foster learners’ inferencing skills and suggests numerous helpful strategies. The section ends with Amos Paran’s “Helping Learners to Become Critical: How Coursebooks Can Help,” a guide to creating activities that utilize
texts not just as linguistic objects or as vehicles for information transmission or even strategy instruction, but as conduits to thinking itself.

Finally, Section III (Materials in Use in Southeast Asia) includes six articles, the first of which is Andrew Gonzalez’s “ESL Materials for Philippine Use in Primary and Secondary Schools: Across Four Paradigmatic Generations,” a historical account of the development and spread of ELT materials in the Philippine context and an appeal for more serious research on their effectiveness. Next, in “Mandated English Teaching Materials and their Implications to Teaching and Learning: The Case of Indonesia,” Iwan Jazadi critically evaluates his country’s government-sponsored secondary-level English textbooks and signals the need for increased localization. Then, in “Where are the ELT Textbooks?” Shanti Chandran explains the reasons behind Malaysian teachers’ overwhelming preference for workbooks as the instructional materials of choice. The following article, Bao Dat’s “Localising ELT Materials in Vietnam: A Case Study,” echoes Jazadi’s earlier concerns and offers concrete examples of adaptation strategies under similar constraints. Next, in “Developing an Oral Communication Skills Training Package: Process and Product; Problems and Solutions,” Gloria Poedjosoedarmo describes the evolution of a pan-ASEAN program to improve local teachers’ speaking and listening abilities. The volume concludes with Carmel Heah and Li Shu Yun’s “Collaborative Materials Design for Communication Skills Training in an Engineering Curriculum,” an informative account of cooperation between language specialists and content experts in developing English for specific purposes (ESP) materials at a Singaporean university.

While the book provides an adequate overview of many current issues in this part of the field, it also has several significant shortcomings that need pointing out. First of all, numerous proofreading oversights exacerbate already stark differences in the quality of a few contributions. Next, not only is there no cross-referencing of recurring threads between articles, but there is not even a comprehensive subject index at the end. Finally, it is ironic that a collection in which the theme of localization plays such a prominent role includes neither biographical information about its contributors nor background on the publishing organization to assist readers in evaluating the relevance of its contents to their own individual contexts. Nevertheless, the Day, Tomlinson, and Tschirner articles alone are worth the price of the volume, which thus ultimately merits recommendation, however tentatively.

Reviewed by
Michael Kindler
Nagoya University of Commerce and Business

This slim volume of 270 pages packs a lot of punch. If readers liked Marianne Celce-Murcia’s work, Teaching English as a Second or Foreign Language (2001), then they will like this volume as well because it supports and expands on the first with plenty of praxis and theory.

The purpose of this book is to inform the reader about the relationship between language and context, as is suggested by the title. There are some 30 contributing authors, most of whom are from west coast American universities. While most of these authors are English L1 speakers, there are a few EFL contributors. However, these contributors also are working in the United States. The reason I emphasize this is that the value of this book, although it is presumably intended for a global audience of language educators, ultimately pertains to the very dimension of context that its authors are at pains to inform us about. So much of the data is related to teaching English in America, which one could safely say is different from, for example, teaching English in Asia, that even if there are some Asian contributors to the volume, the focus of the book is definitely on an American context. Although the concepts espoused in this book are valuable, by and large they do not apply to my working situation, as my context is culturally entirely different from those under discussion.

The editors and several authors acknowledge being influenced by the Marianne Celce-Murcia’s volume cited above which looks at language and cultural context as being interdependent, and not separable. The acknowledgement of Celce-Murcia is significant, as the volume under review can arguably be seen as a source book, which by described practice and theory supports the central task of teaching English as a second or foreign language in the light of Celce-Murcia’s sound and respected approach.

More specifically, the balance of the Power of Context emphasizes teaching English as a second language in an American context, rather than teaching English as a foreign language in countries which are not
English-dominant. Many cultural contexts cited in the book as data do not realistically apply in places where English is taught as a foreign language. That said, the particular orientation of this book is interesting in terms of how it is written and constructed, and this is appealing to any language educator. The praxis, culture, and context is defined and limited, and in that respect its content cannot be universally applied. I am confident that the editors would accept this description as valid, as indeed it is a reflection of their very argument.

So how is this volume organized? There are four sections. The first one is about discovering what context reveals about grammatical structures. In this part, the contributing authors discuss how, by adopting contextual analysis, they have experienced, identified, and backgrounded different linguistic behaviors, such as the difficulty of explaining native speaker intuition, nouns without articles, and the distribution and meaning of any, some, and every in discourse. This section also includes some useful references for researchers, as well as information on how using expanded computer-based corpora can aid the investigator. Action-based research, or the teacher-as-researcher in his/her own classroom are also shown as valid methods of documenting contextual analysis.

The second part of the book broadens out into a discussion of varying methods of oral and written discourse analysis. Discourse pragmatic issues dominate these chapters, ranging from a discussion of requests, complaints, and apologies to an examination of the coherence of oral and written language between speakers/writers and listeners/readers. An interesting chapter for me here was the one on information structure, or how information is presented, in terms of given and new information. The concept of scaffolding vocabulary and conceptual expansion of words is an ongoing concern of language educators. I also enjoyed reading Jeanette DeCarrico’s essay on questions of form and function in lexical phrases and discourse content.

An exploration of the interface between language skills and discourse makes up the third part of this book. The focus is the intersection of language skills, grammar, and discourse contexts. Top-down and bottom-up models are discussed, as is the role of grammar and context, and the relationship of one to the other. This section has a pragmatic ring to it. In addition, Tetsuo Harada from the University of Oregon has a nice discussion of the effects of CBI (content-based instruction) on learners of English as a second language.

The final part of this book deals with creating contexts for effective language teaching and is arguably the most appealing part of the book as
it deals with practical “how-to questions.” In particular, this section supports the assertion that reflective classroom researchers, who evaluate and rethink their approach as a result of the impact and feedback which they receive from the students, are likely to be more successful than teachers who automatically and robotically reach for the textbook and methodically work through it irrespective of the relevance or appropriateness of the material to the learning needs of students. This means that the thinking teacher, sensitive to the students’ ability to receive and comprehend, is ahead of the practitioner who follows a particular method because of his/her belief in its absolute superiority. It also means that the language educator remains the focus not so much as supplier, but rather facilitator of language learning, as the context and content are molded together with the place, needs, and aspirations of all participants.

In conclusion, for educators wishing to ground themselves in some of the governing concepts current in teaching English in America, this volume provides an attractive overview and gives sufficient references. The book is even indexed for this purpose. As a multilingual person, I regard many of the notions discussed somewhat self-evident and redundant, but that is only because I have empirically absorbed them through some 50-odd years of an examined life working mostly with people learning English as a first or second language outside the United States. What I especially liked about the book is that it is relatively ideologically unencumbered, in the sense that while it has succeeded in showing where current language teaching trends stand, it has not done so polemically, but with a freshness and originality that is to be commended.

Reference

Information for Contributors

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

Editorial Policy

JALT Journal, the refereed research journal of the Japan Association for Language Teaching (Zenkoku Gogaku Kyoiku Gakkai), invites practical and theoretical articles and research reports on second and foreign language teaching and learning in Japanese and Asian contexts. Submissions from other international contexts are accepted if applicable to language teaching in Japan. Areas of particular interest are:

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

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

Guidelines

Style

The JALT Journal follows the Publication Manual of the American Psychological Association, 5th edition (available from APA Order Department, P.O. Box 2710, Hyattsville, MD 20784, USA; by e-mail: <order@apa.org>; from the website: <http://www.apa.org/books.ordering.html>). Consult recent copies of JALT Journal or TESOL Quarterly for examples of documentation and references. A downloadable copy of the JALT Journal style sheet is also available on our website at <http://www.jalt-publications.org/jj/>.

Format

Full-length articles must not be more than 20 pages in length (6,000 words), including references, notes, tables, and figures. Research Forum submissions should be not more than 10 pages in length. Perspectives submissions should 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 750 words. All submissions must be typed and double-spaced on A4 or 8.5”x11” paper with line spacing set at 1.5 lines. The author’s name and identifying references should appear only on the cover sheet. Authors are responsible for the accuracy of references and reference citations.

Submission Procedure

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

Materials to be submitted

1. Cover sheet with the title and the author name(s).
2. One (1) copy of the manuscript, with no reference to the author. Do not use running heads.
3. Contact information, including one author’s full address and, where available, a fax number and e-mail address.
4. Abstract (no more than 150 words).
5. Japanese translation of the title and abstract, if possible (less than 400 ji).
6. Biographical sketch(es) (no more than 25 words each).
7. Authors of accepted manuscripts must supply camera-ready copies of any diagrams or figures and a disk copy of the manuscript (RTF or ASCII).

Evaluation Procedures

Submissions will be acknowledged within one month of their receipt. All manuscripts are first reviewed by the editorial board to insure they comply with JALT Journal Guidelines. Those considered for publication are subject to blind review by at least two readers, with special attention given to: (1) compliance with JALT Journal Editorial Policy, (2) the significance and originality of the submission, and (3) the use of appropriate research design and methodology. Evaluation is usually completed within three months. The first author of a published article will receive one copy of the Journal and 10 off-prints with the option to order more at a set rate (contact JALT Central Office for price details). Review authors receive one copy of the Journal.
Restrictions

Papers submitted to JALT Journal must not have been previously published, nor should they be under consideration for publication elsewhere. JALT Journal has First World Publication Rights, as defined by International Copyright Conventions, for all manuscripts published. We regret that manuscripts or computer disks cannot be returned. In the interest of facilitating clarity, the editors reserve the right to make editorial changes to accepted manuscripts.

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

Please send submissions in these categories or general inquiries to:

jj-editor@jalt-publications.org
Steve Cornwell, JALT Journal Editor
Osaka Jogakuin College
2-26-54 Tamatsukuri, Chuo-ku, Osaka 540-0004, Japan

Perspectives

jj-editor2@jalt-publications.org
Deryn Verity, JALT Journal Associate Editor
Osaka Jogakuin College
2-26-54 Tamatsukuri, Chuo-ku, Osaka 540-0004, Japan

Reviews

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

jj-reviews@jalt-publications.org
Yuriko Kite, JALT Journal Reviews Editor
c/o Canadian Academy
4-1 Koyo-cho Naka, Higashinada-ku, Kobe 658-0032, Japan

Japanese-Language Manuscripts

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

jj-editorj@jalt-publications.org
Yoshinori J. Watanabe, JALT Journal Japanese Language Editor
Faculty of Education and Human Studies
Akita University,
1-1 Tegata Gakuenmachi, Akita 010-8502, Japan

Address for Inquiries about Subscriptions, Author Reprints, or Advertising

JALT Central Office
Urban Edge Building 5F
1-37-9 Taito, Taito-ku, Tokyo 110-0016, Japan
Tel.: 03-3837-1630; Fax: 03-3837-1631
(From overseas: Tel.: 81-3-3837-1630; Fax: 81-3-3837-1631)
website: http://www.jalt.org
日本語論文投稿要領

JALT Journalでは日本語で執筆された論文、研究報告、実践報告、書評等を募集しています。

文体: 一般的な学術論文のスタイルを用い、章立ての仕方や参考文献のデータの書き方などは、Publication Manual of the American Psychological Association (5th ed.)の定める方式に合わせて下さい。不明の場合は、JALT Journalの英語論文を参考にするか、日本語編集者までお問い合わせ下さい。なお、JALT Journalの読者は現場の教師が主なので、特殊な専門用語や統計的手法は、わかりやすく定義するか説明を加えるなどして下さい。

原稿: 長さは、参考文献リストも含め18,000字（書評の場合は1,500字）以内です。A4の用紙に横書きで、1行40字、1ページ30行で印刷して下さい。手書きの原稿は受け付けません。図表をいれる場合は、JALT Journalのページのサイズに合わせて作成し（縮小コピー可）、本文とは別に作成して下さい。執筆者の名前や所属機関名は、原稿には書かないで下さい。原稿締切日は定めていません。いつでも受け付けます。

提出するもの:
- 原稿3部（執筆者名は書かないこと）（尚、eメールで送付の場合は1部のみ添付ファイルで送付のこと）
- 論文タイトル、執筆者の名前と所属機関名および連絡先（住所、電話番号、ファックス、e-mailアドレス）を一枚に収めた表紙
- 400字以内の和文要旨
- 英文のタイトルと、500〜750語の英文要旨（書評の場合は100語程度の英文要旨）
- 100字以内の執筆者略歴
- MS-Wordで保存したファイル（マッキントッシュ使用の場合はrtfファイルで保存したもの）

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。査読は通常三か月以内に終了します。

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:

〒010-8502 秋田市手形学園町1-1
秋田大学 教育文化学部
JALT Journal日本語編集者 渡部良典
電話・Fax: 018-889-2639
jj-editorj@jalt-publications.org

JALT Journal 第27巻 第1号
2004年4月20日 印刷
2004年5月1日 発行
編集人 スティーブ・コーンウィル
発行人 スティーブ・ブラウン
発行所 全国語学教育学会事務局
〒110-0016 東京都台東区台東1-37-9 アーバンエッジビル5F
TEL (03)3837-1630; FAX (03)3837-1631
印刷所 コーシンシャ株式会社
〒530-0043 大阪市北区天満1-18-4天満ファーストビル301 TEL（06）6351-8795