JALT Publications • Online Journals

JALT Journal

JALT Journal is the research journal of the Japan Association for Language Teaching (JALT). It is published semiannually, in May and November. As a nonprofit organization dedicated to promoting excellence in language learning, teaching, and research, JALT has a rich tradition of publishing relevant material in its many publications.

Links

jalt journal

JALT Journal

Volume 36 • No. 1

May 2014

Contents

Articles

Reviews

JALT Journal Information

Japan Association for Language Teaching

A Nonprofit Organization

The Japan Association for Language Teaching (JALT) is a nonprofit professional organization dedicated to the improvement of language teaching and learning in Japan. It provides a forum for the exchange of new ideas and techniques and a means of keeping informed about developments in the rapidly changing field of second and foreign language education. Established in 1976, JALT serves an international membership of approximately 3,000 language teachers. There are 33 JALT chapters and one forming chapter, all in Japan, along with 23 special interest groups (SIGs) and three forming SIGs. JALT is one of the founders of PAC (Pan-Asian Consortium), which is an association of language teacher organizations in Pacific Asia. PAC holds regional conferences and exchanges information among its member organizations. JALT is the Japan affiliate of International TESOL (Teachers of English to Speakers of Other Languages) and is a branch of IATEFL (International Association of Teachers of English as a Foreign Language).

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

The JALT International Conference on Language Teaching and Learning and Educational Materials Exposition attracts some 2,000 participants annually and offers over 600 papers, workshops, colloquia, and poster sessions. Each JALT chapter holds local meetings and JALT's SIGs provide information and newsletters on specific areas of interest. JALT also sponsors special events such as workshops and conferences on specific themes, and awards annual grants for research projects related to language teaching and learning. Membership is open to those interested in language education and includes automatic assignment to the nearest chapter or the chapter you prefer to join, copies of JALT publications, and reduced admission to JALT-sponsored events. JALT members can join as many SIGs as they wish for an annual fee of ¥1,500 per SIG. For information, contact the JALT Central Office or visit the JALT website at <www.jalt.org>.

JALT National Officers, 2014

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

Chapters

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

Special Interest Groups

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

JALT Central Office

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

In this Issue

Articles

In this spring issue, we are pleased to present five full-length research articles. The first is by **Eric Hauser**, who, drawing on conversational analysis, examines university students' embodied uses of electronic bilingual dictionaries. In the second, **Rie Koizumi** and **Yo In'nami** use structural equation modeling to examine the complexity, accuracy, and fluency of English speaking proficiency among junior and senior high school students. The third is by **Kazunari Shimada** who also looks at Japanese junior and senior high school students' spoken English, exploring their use of discourse markers by way of contrastive interlanguage analysis. Our fourth article was written by **Judith Runnels**, who used Mokken scaling to determine the reliability of can-do statements from the five skills of the Common European Framework of Reference-Japan. In the fifth article, a Japanese-language contribution, **Hiroaki Tanaka** did quantitative and qualitative analyses of a 15-week intervention to increase intrinsic language-learning motivation in university students.

Reviews

In this issue of *JALT Journal* we provide seven book reviews on a range of titles of interest to both teachers and researchers. In the opening review, **Keith Adams** covers an edited volume on replication research. The second review, by **David Beglar**, looks into the future direction of SLA research presented by the editors of *Language Learning* in their latest biennial supplement. In the third review, **Howard Brown** addresses an academic text on content and language integrated learning (CLIL) of interest to program directors and teachers who are dealing with language *of*, *for*, and *through* learning. The next review, written by **Tyler Burden**, is on the third edition of a title on trends in teaching practice, methods, and materials. **Robby Caughey** then examines a book from the Routledge ESL & Applied Linguistics Professional Series on teaching and learning listening with a metacognitive focus. In the sixth review, **Gilbert Dizon Jr.** explores an edited collection on computer-assisted language learning (CALL) drawn from both research and practical points of view. In the final review, **Nick Doran** highlights Simon Borg's title on teacher cognition research and practice.

From the Editor

I would like to dedicate this issue of *JALT Journal* to the memory of Kevin Cleary who, while President of JALT, was taken from us suddenly in January of this year. To tell the truth, I didn't really know Kevin very well, but he made a very strong and lasting impression on me. He emanated, among other things, strength, solidity, kindness, humility, and fairness. His leadership skills were exemplary. He surrounded himself with qualified people and allowed them to do their work. And, because he put such faith and trust in us, we wanted and tried our utmost to live up to his expectations.

In my mind's eye, I see Kevin, tall and dapper, in the middle of an Educational Materials Exhibition hall at any JALT conference. He's never alone, approached by a constant stream of frantic conference workers with questions, friends with comments and compliments, plenary speakers needing to know where to go and when, and so on and so on, and so on. Always busy, always in demand, Kevin always had time for a word, a smile, or an encouraging hand on a shoulder.

We are all of course bereft, but Kevin touched and had such a positive impact on so many of our lives, that we must believe that he did the work he was put here to do. Kevin taught us to stand tall, to be strong, to help others, to be kind, to have humility, to trust others to do their work well, and that JALT is something worth devoting oneself to. He also taught us that it doesn't hurt to dress up and that there's always time for a beer with friends. He has left us his legacy; it's up to us to carry it on.

Melodie Cook

Articles

Embodied Uses of Electronic Bilingual Dictionaries

Eric Hauser
University of Electro-Communications and University of Hawai'i at Mānoa

Electronic bilingual dictionaries are widely used among university students in East Asia. There is a small body of research, based on questionnaires or experiments or both, on their use and effectiveness, but with one exception, research has not been focussed on the details of actual dictionary use. Drawing on conversation analysis, the current study presents analyses of students' embodied use of electronic dictionaries during second language English discussions. It is shown that (a) the layout of items on the screen is a resource for recognition, (b) there is an orientation to dictionary ownership, (c) the configuration of objects and bodies is consequential for how dictionaries are used, (d) manipulation of a dictionary can be interactionally significant, and (e) there is not a strong normative element to how dictionaries should be consulted. It is argued that dictionaries are used to accomplish a variety of objectives unlikely to be revealed through questionnaire or experimental research.

電子辞書は、東アジアの大学生の間に広く普及している。これまで、アンケートや実験の結果に基づいた電子辞書の効果と使用状況についての報告はいくつかあるが、実際の辞書使用を詳細に分析した研究はほとんど見られない。本論文は、第二言語としての英語ディスカッションに見られる学生の電子辞書の具現化された使い方を、会話分析を使って詳細に示す。分析では、以下の点について提示する：（a）電子辞書画面上に表示される画像・文字が認識のリソースになること、（b）辞書の所有権に対する意識が見られること、（c）辞書使用に関して、物と身体の配置が重要であること、（d）辞書の操作が相互行為的な意味を持つ可能性があること、（e）辞書の使用について、それほど強い規範的な志向性がみられないこと。これらの分析に基づいて、アンケートや実験に基づく従来の研究では観察されなかった多様な目的の辞書使用の実態を論じる。

S ince the 1990s, the pocket electronic bilingual dictionary (ED)[1] has become popular among East Asian university students,[2] who even if they do not major in English, are often required to take EFL classes. This may be specific to East Asia, with such dictionaries being less popular outside East Asia (Chen, 2010; Jian, Sandnes, Law, Huang, & Huang, 2009). Perhaps for this reason, East Asia has also been the location for most research into the use of EDs. As pointed out by Kobayashi (2008), this research can be divided into two types—research investigating how students use EDs and research investigating its effectiveness—though reports of research often contain both. Such research typically involves some sort of comparison between EDs and paper dictionaries.

Research investigating how students use EDs has relied heavily on questionnaires (e.g., Bower & McMillan, 2007; Chen, 2010; Jian, et al., 2009; Kobayashi, 2007, 2008; Weschler & Pitts, 2000), though Kobayashi (2007) also drew on data from retrospective reports to investigate how participants used either a paper dictionary or an ED during an L2 reading task, and Kobayashi (2008) drew on data from interviews with a subset of questionnaire respondents. Research investigating the effectiveness of dictionaries has tended to be experimental (e.g., Chen, 2010; Kobayashi, 2007; Loucky, 2002, 2003; Weschler & Pitts, 2000). With regard to such things as retention of vocabulary looked up during a task, experimental research has not found any significant differences between the effectiveness of EDs and paper dictionaries, though there may be a speed advantage for EDs. Weschler and Pitts (2000) and Loucky (2002, 2003) reported that it took students slightly less time to find words in an ED, but none of these reported any tests of statistical significance.

With what appears to be only one exception (Barrow, 2009, see also 2010), ED research has not been based on careful observation and analysis of what students actually do when using such a dictionary. Such observation and analysis have the potential to provide useful information for English teachers, who may be helped to better understand how their students use EDs. As is shown by the substantial body of ethnomethodological and conversation analytic (CA) work on the use of technology in work places (e.g., Goodwin, 1995; Heath & Luff, 2000; Suchman, 1987, 2007; Whalen, 1995), detailed analysis of what people do with technology can reveal unnoticed, taken-for-granted features of how people actually use it. When it comes to how students use EDs, apparently the only study with a detailed analysis of dictionary use is Barrow (2009), in which were described three ways that Japanese university students consult EDs during L2 English discussions: consultations that occur during a turn-at-talk, consultations that

are abandoned, and consultations in order to find a word for later use. Barrow's study focused on the first of these, which he found to be somewhat more common than the other two. Barrow suggested on the basis of its more common occurrence that the way participants organize consultations that occur during a turn-at-talk is a normative organization. This is problematic, though, as he did not provide any evidence that participants orient to this organization as normative. What can be said is that Barrow's analysis demonstrated that practices of self-initiation of repair (e.g., cutting off a word, sound stretches, changes of gaze direction) cluster prior to participants' dictionary consultations. These repair initiation practices, and other practices involved in consulting a dictionary, demonstrate how the participants commonly use their EDs to solve problems with finding L2 vocabulary during a turn-at-talk. They also more or less strongly project what the L2 word is, sometimes by the participant who is consulting a dictionary articulating a Japanese translation equivalent during the consultation. This allows for certain forms of collaboration, such as the other participant proposing an L2 word or also consulting his or her own dictionary. Finally, through the use of video data, Barrow demonstrated that practices of consulting a dictionary are embodied practices.

The current study differs from Barrow (2009, 2010) in that, although it presents CA-type analyses of the sequential organization of interaction that involves the use of EDs, rather than focusing on recurrent features of sequences involving dictionary consultation, it focuses on such things as how the affordances of dictionary design, the location of a dictionary relative to the participants, dictionary ownership, and the placement and orientation of bodies and material objects contingently influence participants' organization of their dictionary use. To do this, I first describe what is happening in a particular instance and then make analytic observations about ED use during the episode on the basis of this description. Each instance analyzed below is treated as a unique occurrence. General points that can be learned from these unique occurrences will be discussed in the conclusion.

Data

The data are drawn from video recordings of L2 English-language discussions among students at two different universities in Tokyo. The students are not majoring in English but are taking either required or elective EFL classes. They are participating in these discussions as part of a class assignment. Some of the discussions are conducted during class, others outside class. The corpus consists of slightly over 4 hours of recorded discussions,

which have been transcribed based on CA transcription conventions (see Jefferson, 2004). The data were collected for the purpose of investigating how participants in L2 discussions interact and the linguistic and nonlinguistic resources they draw on to do so. They were not collected for the specific purpose of investigating dictionary use. All participants gave oral consent for these recordings to be used for research purposes. Participants are referred to in the text by pseudonymous Japanese surnames. Data from this corpus have been used in other publications (e.g., Hauser, 2013). It should be noted that some participants did not use dictionaries, others made limited use of them, and a few participants relied on them heavily. No participants used a paper dictionary during their discussion, which may reflect the ubiquity of EDs among East Asian, or at least Japanese, university students.

The data are presented as a mixture of transcripts and video frames, with the frames available in Online Appendix A, Transcripts With Frames. A list of transcription symbols, based on Jefferson (2004), is in Appendix B. Where Japanese words appear in the transcripts, a morpheme-by-morpheme gloss is provided in the following line. Each frame is numbered, with the number of the frame placed in the transcript beneath the talk that was being produced at that time. There are no frames for Excerpt 1a. The original video data can be viewed through the URL provided with each excerpt.

Embodied ED Use

In Excerpt 1a, the students are talking about a vocabulary item from the reading that is the topic of their discussion.

Excerpt 1a http://youtu.be/nprgXV1a-lM

```
01    K: eh .h (kore) (0.4) <cluster bombs>
                    this
02          °↓tte nani.°
               QT    what
03    W: ↑cluster bombs uh (0.2) ↓uh-
04    K: °jirai:°
          land mine
05    (0.2)
06    W: ↑no. ↓chigau.
                different
```

Note. QT = quotative

Kobayashi (K), asks what a *cluster bomb* is. In line 04, she states a candidate understanding, the Japanese word *jirai*.[3] Watanabe (W), sitting in the middle, rejects this in line 06 and goes on, though this is not shown in the transcript, to explain what a cluster bomb is, after which he then explains what carpet bombing is.

The interaction in Excerpt 1b occurs a couple of minutes later. In lines 01 to 05, Watanabe, still seated in the middle, is finishing his explanation of carpet bombing. Kobayashi is seated on the left and a third participant, Chiba (C), is on the right.

Excerpt 1b http://youtu.be/6AI6DepT1RI

```
01    K: °ah° [ah .h ↑ah:
02    W:      [if-
03    W: if large area bombed, (0.9) it is kuh- carpet.
04    K: n:: [:: n n n.
05    W:      [car- carpet.
06    W: uh: (2.4) ((C turns toward dictionary, body
               1   2
07       mostly oriented toward group, RH on dictionary))
08    K: ↑ee::: (0.2) so: .h (0.3) °↓m::° (0.9) uh
            IT
09        ↑if [if
10    C:      [land mine_ ((still facing dictionary))
                    3
11    (1.0) ((W gazes and leans toward dictionary))
         4 5
12    K: (whach)=
13    C: =jirai is ↑land mine. ((brings dictionary
              6              7
          land mine
14       closer to W; K leans forward))
15    W: la- ↑land mine. ((K leans back))
               8
16    C: land mine.
```

```
17     (0.4) ((W gazes away from dictionary; C turns
          9
       dictionary toward self))
18     W: land mine [ah
19     K:           [°m eh? nani. ↓land mine. °((leans
                                    10
                         what
20       forward))
21     C: land °mine_° ((shows/holds out dictionary to K))
                  11
22     (1.2) ((K takes dictionary))
         12
23     K: hee:: ↑lando mine_ ((returns dictionary))
                   13
           IT
24     (2.3) ((C places dictionary in original location))
        14
25     W: uh (1.1) ↑to use:
          15
26     K: n:
27     W: ↓uh ↑land mine and, (0.2) >cluster bombs
28       and,< (0.2) uh- (0.2) huge bomb
29     K: m m
30     W: = is: (0.2) not (0.5) ↑not (0.3) uh:: (0.9)
31       righ- right thing. ↑things:.
```

Note. IT = interactional token

During the silence in line 06 (frames 1 and 2), Chiba turns her head toward her dictionary, located on the table slightly behind her, while her torso remains oriented to the group. She operates the dictionary with her right hand. Even though Kobayashi starts to talk in lines 08 and 09, in line 10, Chiba says "land mine," while her head is still turned toward her dictionary (frame 3). In response to this, Watanabe turns his head toward Chiba (frame 4) and then leans toward her dictionary (frame 5). These changes in posture

occur during the gap in line 11. In line 12, Kobayashi says something that sounds like a cross between *what* and *which.* In line 13, as Chiba says "*jirai* is land mine," Watanabe leans closer to the dictionary (frame 6) and Chiba then moves her dictionary closer to him (frame 7). Meanwhile, Kobayashi also leans forward. In line 15, Watanabe says "land mine." As can be seen in frame 8, he is still leaning toward the dictionary, while Kobayashi leans back in her chair, apparently giving up on being able to see Chiba's dictionary. Chiba then says "land mine" again, while Watanabe gazes away from the dictionary, which Chiba reorients away from Watanabe's line of sight (frame 9). In line 18, Watanabe repeats "land mine" again, after which Kobayashi asks the meaning of land mine (line 19). She leans forward as she says "land mine" and Chiba moves her dictionary toward Watanabe (frame 10). As Chiba once again says "land mine" in line 21, she passes the dictionary to Kobayashi (frame 11). During the gap in line 22, Chiba withdraws her hand, so that Kobayashi is now holding the dictionary (frame 12). During this change in who is holding the dictionary, Kobayashi remains leaning forward. In line 23, Kobayashi notes the newness of the information with "*hee*" (Mori, 2006), repeats "land mine," and returns the dictionary to Chiba (frame 13). During the following gap in line 24, Chiba places the dictionary back in its original location (frame 14). She then turns her head back toward the group as Watanabe starts to talk in line 25 (frame 15). He then goes on to use land mine as part of a three-part list (Jefferson, 1990) of "land mine, cluster bomb, and huge bomb."

There are several observations that can be made about the interaction presented in this excerpt on the basis of this description. First, the Japanese word that Chiba translates using her dictionary is a word that was first introduced by another participant, in Excerpt 1a. However, initially, neither the word *jirai* nor its English translation was topicalized. By announcing the result of her dictionary search, Chiba topicalizes the translation, *jirai*. Second, all three participants orient to Chiba's dictionary as the source of the translation. Chiba does this by facing her dictionary as she says "land mine" and "*jirai* is land mine" and by moving the dictionary so that the screen can be seen first by Watanabe and then by Kobayashi. This also involves handing her dictionary to Kobayashi. Watanabe and Kobayashi do this by leaning toward the dictionary and, in the case of Kobayashi, briefly taking it from Chiba. Third, Watanabe and Kobayashi treat the dictionary as belonging to Chiba. Watanabe does not try to take the dictionary and, in fact, never brings his hands to his left side while gazing at it. Kobayashi takes hold of the dictionary, but remains leaning forward, holding it with one hand and not

bringing it much closer to her body. Fourth, the location of the dictionary at the start of the excerpt—on a desk to Chiba's right—and the fact that there is nothing (e.g., a table or desk) between the participants on which to place the dictionary influence Chiba's actions. She turns her head away from the other participants in order to consult the dictionary, holds it out for Watanabe to see, and passes it to Watanabe, all of which would have been done differently if, for example, the participants had been seated around a table and the dictionary had been placed on this table. Fifth, the size of the dictionary screen and the configuration in which the participants are sitting appear to constrain the number of people who can look at the dictionary at one time. As a result, Kobayashi abandons her first attempt to look at the dictionary. Her second, successful attempt comes after Watanabe has withdrawn his gaze. And finally, both Watanabe and Kobayashi apparently know where to look on the dictionary screen to find the translation. They do not need Chiba to point out where the translation can be found. This can be understood as an affordance of how the Japanese word and its primary translation appear on the screen. That is, they appear at the top of the screen and the other participants, knowing this, know where to look.

Excerpts 2 and 3 both involve the same two participants. More than any other participants, these two rely very heavily on their dictionaries. In Excerpt 2, one participant touches and slightly moves the other's dictionary.

Excerpt 2 http://youtu.be/pKhCHVOhs5g

```
01    T: I (5.5) ((dictionary use)) I- (8.2)
         1  2 3
02       I expect (0.4) you. °h h°
         4           5          6
03    (4.4) ((H looks at T's dictionary))
          7
04    H: hh
           8
05    (2.3)
          9
06    H: I expect (.) you too.
         10          11
```

In line 01, Tanabe (T), says "I" and then pauses. At the start of the pause, he is not consulting or moving toward his dictionary. Rather, as can be seen in frames 1 and 2, as he pauses after saying "I," he crosses his arms and gazes to middle distance. However, in frame 3 he then unfolds his arms and moves his hands and gaze toward his dictionary. Most of the pause following this "I" and the next pause following the second "I" are filled with dictionary use. At the end of the pause, Tanabe removes his hands from the dictionary and from the table. With his gaze still on the dictionary (frame 4), he says "I expect." He then pauses briefly and shifts his gaze to Hamada (H), as can be seen in frame 5. Tanabe then adds "you" and produces a slight laugh. By the end of the laugh (frame 6), Hamada has shifted his gaze to Tanabe. Hamada has trouble understanding what Tanabe has said and responds by leaning forward and turning Tanabe's dictionary toward himself (frame 7). He then smiles (frame 8) and turns the dictionary back toward Tanabe while laughing slightly. The laugh, audible as a response to what he has seen on the dictionary screen, indexes that he now understands. Still smiling, Hamada leans back in his chair and gazes to middle distance, with Tanabe gazing at him during the silence in line 05 (frame 9). At the start of line 06, Tanabe shifts his gaze off Hamada and Hamada shifts his gaze towards Tanabe, as can be seen in frame 10. Hamada says "I expect" and pauses briefly. During the pause (frame 11), Tanabe returns his gaze to Hamada. Hamada then says "you too."

Again, several observations can be made on the basis of this description. First, what Tanabe says to Hamada following his consultation of the dictionary, "I expect you," is visibly based on what Tanabe has found in the dictionary. This allows Hamada to assume that he can solve his problem understanding what Tanabe has said to him by looking at the on-screen product of Tanabe's dictionary work. Second, Hamada knows where to look on the dictionary screen for the information he needs to solve his understanding problem. Third, although Tanabe's actions in line 01 indicate that he is engaging in a word search, or what comes to be self-initiated self-repair accomplished through the use of the dictionary as a tool, Hamada's actions can be understood as other-initiated other-repair (Schegloff, Jefferson, & Sacks, 1977). The repairable is "I expect you" and Hamada uses the dictionary to both initiate and accomplish the repair. He initiates the repair by leaning forward and turning the dictionary and then accomplishes it by silently reading what is on the screen. This excerpt thus contains a type of repair found to be common by Barrow (2009, 2010), but this is followed by additional, and differently organized, repair work. Fourth, though he reaches out and turns Tanabe's dictionary toward himself, Hamada treats the dictionary as belong-

ing to Tanabe, as he moves it only slightly and turns it back toward Tanabe after he has solved his understanding problem. Finally, the configuration of the furniture and the orientation of the participants' bodies allow Hamada to gain visual access to the dictionary screen with only minimal adjustment of Tanabe's dictionary. Hamada only needs to lean forward and turn the dictionary slightly. The resulting change in the direction the dictionary is facing is small enough that Tanabe can still see the screen, which he continues to gaze at.

Excerpt 3 involves the same two participants.

Excerpt 3 http://youtu.be/uPILc5JgLIU

```
01    H: global warming (2.0) °kah° (1.6) ((starts to
02       check dictionary)) ah causee(0.5)zu (2.3)
                                                1
03       cauzu? (13.0) ((checks dictionary, shows T))
              2     3
04    T: cause.
05    H: cause (.) ↑cauzu (4.2) the ice (0.8) melted.
            4            5                    6   7
```

In line 01, Hamada has problems finding a word that he wants, but he does not immediately use his dictionary. During the first pause in line 01, he does not make any move toward his dictionary. During the second pause, he moves his hands and shifts his gaze to his dictionary. However, he abandons this and says "ah causes," apparently completing the word search. However, he then pauses again, shown in frame 1, and says "cause" with rising intonation while moving his hands and gaze back to his dictionary (frame 2). Most of the long pause in line 03 is taken up with Hamada using his dictionary. Near the end of the pause (frame 3) he turns the dictionary toward Tanabe and points to something on the screen. Tanabe responds by shifting his gaze to the dictionary and saying "cause." As shown in frame 4, both participants keep their gaze on the dictionary as Hamada repeats "cause" in line 04. Hamada then turns his dictionary away from Tanabe and leans back (frame 5) as he again says "cause." He keeps his gaze on his dictionary through the following long pause, as he says "the ice," pauses again, and articulates the first syllable of "melted" (frame 6). Finally, as he articulates the second syllable of "melted," he shifts his gaze to meet Tanabe's (frame 7).

Again, it is possible to make a few observations. First, Hamada points to something on the dictionary screen, so the layout of items on the screen does not necessarily make it clear which item is the relevant one. Second, although this excerpt also involves the sort of repair practices discussed by Barrow (2009, 2010), the participants' actions in lines 03 and 04 can be further understood as self-initiated other-repair (Schegloff, et al., 1977) related to how to articulate a word, with Hamada initiating repair by turning his dictionary to Tanabe and pointing, and Tanabe doing the repair by saying "cause." Third, the participants orient to Hamada's ownership of his dictionary. Tanabe, while he keeps one or both hands on his own dictionary, leans towards Hamada's dictionary as he gazes at it, but does not attempt to manipulate it himself. Hamada, after he has elicited Tanabe's assistance, reorients his dictionary back toward himself. Fourth, the configuration of furniture and participants' bodies creates a shared space in which Hamada can turn the dictionary so that both of them can simultaneously see the screen. Finally, by turning the dictionary toward Tanabe and pointing at the screen, Hamada is able to attract Tanabe's attention to the dictionary screen.

In Excerpt 4, the analysis will be focused on Abe (A), seated in the middle, facing the camera.

Excerpt 4 http://youtu.be/NP2uAR0DVL0
```
01   A: ah:: (1.2) ↑so: (0.4) ↑sometimes the m-
02      media (.) uh like a newspaper or eigh-
03      (.) a tee vee ((TV)) program .h uh: have
04      uh (1.2) uh (2.1) ↑provide a (0.3) uh: (2.5)
05      larger larger meaning.
06   (2.8)
07   A: eh heh heh
        1
08   D: heh
09   A: .hh (0.2) can I look at the dictionary?=hh
                                                  2
10   (0.8)
11   B: no.
12   A: no:?
13   B: cunning.
```

```
14    (0.2)
15    A: cunning?
16    (0.2)
17    B: uh cheating.
18    A: ch(h)eat(h)ing (.) reall(h)y=hh
                                      3
19    (0.2)
20    A: uh:: (0.4) .t I forgot the (0.8) word uh::
                        4
21    (0.5)
22    A: .h ((sniff))
23    (1.8)
24    A: yes. ah(h). exaggeration.
          5
25    (0.6)
26    B: n= [n
27    A:    [exagger- exaggeration
                    6              7
28    A: m. you know there's eigh exaggeration.
                          8
29    (0.4)
30    A: Japanese say kochoo?
                        exaggeration
31    (0.7)
32    A: ii sugi?
            overstatement
33    (0.4)
34    A: uh: (0.3) in the some tee vee ((TV))
35       program:. (.) or eh uh newspaper.
```

In lines 01 to 05, with several disfluencies, Abe says "so sometimes the media, like a newspaper or a TV program, provide a larger meaning." However, this gets no response from any of the other participants, resulting in a

long silence in line 06. In line 07, Abe laughs and reaches for his bag on the floor next to him, as shown in frame 1. In response to Abe's laughter, Doi (D), on the left, also laughs briefly in line 08. In line 09, Abe indicates why he is reaching for his bag by asking "can I look at the dictionary?" In frame 2 he is looking in his bag as he says this. There is then an exchange between Baba (B), on the right, and Abe about whether looking in the dictionary is allowed, but Abe does not treat this seriously and continues preparing to use his dictionary. As he finishes the word *really* in line 18, he opens his dictionary (frame 3). While looking in the dictionary, he gives an account for why he needs to do this in line 20 (frame 4). In line 24, Abe indicates that he has found what he wants by saying "yes." As can be seen in frame 5, he raises his head slightly as he says this, beginning disengagement from the dictionary. He then says "ah" and "exaggeration," presumably the word he was looking for. Baba makes a minimal response in line 26, after which Abe again says, with some disfluency and repetition, "exaggeration." As he says this (frames 6 and 7), he closes the dictionary, looks to his right, and starts to place the closed dictionary on the table. However, as he begins to use the word he has found in a larger turn, in line 28, he instead returns the dictionary to his bag (frame 8). He next offers two translations of the word, in lines 30 and 32. What he then says in lines 34 and 35 is built syntactically as a continuation of what he has said in line 28, so that from line 28 to line 35, he says the sentence, "you know there's an exaggeration in some TV program or newspaper," with the two translations inserted after the word *exaggeration*.

Once more, it is possible to make several observations based on this description. First, Abe neither holds his dictionary so that others can see the screen nor does anything to invite them to look at the screen. Nor do the other participants do anything to be able to see the screen. Abe's observable ability to remove the dictionary from his bag and use it within his personal and private space before returning it to his bag shows the participants' orientation to the dictionary as belonging to Abe.[4] Second, Abe's dictionary use can be understood as self-initiated self-repair in third position (Schegloff, 1992). In line 05, he has reached a completion point, but does not receive any response, which can be taken as indicating a lack of understanding. He initiates repair by retrieving his dictionary and does the repair by reformulating what he has said with the word that he has found in the dictionary. While Abe engages in word search in lines 03 and 04, as can be seen and heard in the sound stretches, the inbreath, the pauses, and the nonlexical *uh*s, he completes the search and reaches a completion point in line 05. It is only when this completed turn gets no response that Abe initiates repair

as subsequent action. The organization of repair work and dictionary consultation in this excerpt is thus quite different from the same turn repair initiation described by Barrow (2009, 2010). Third, Abe's actions involved in using the dictionary take time. He treats this as accountable and uses talk to indicate what he is doing (line 09) and why (line 20). His accounting work in line 09 provides Baba with the opportunity to engage Abe in some nonserious interaction unrelated to the topic that they are discussing, but still in English. Fourth, Abe states the word that he has found and provides two translations. The others thus share in the benefits of Abe's dictionary use. Finally, though he does not treat seriously Baba's claim that using the dictionary is cheating, Abe switches from placing the dictionary on the table next to him, where it would be more accessible, to returning it to his bag. He treats the dictionary as something that is not properly out of his bag during the discussion and, perhaps, tacitly agrees that using the dictionary is cheating.

Conclusions

As did Barrow (2009, 2010), I have analyzed in detail some of what participants in L2 discussions do when they use their EDs. However, unlike Barrow's research, the focus of this study has been on the contingent and unique features of embodied use of EDs during interaction, rather than a collection of a particular practice. Nevertheless, based on the observations related to each excerpt above, it is possible to make some more general points about the use of EDs and to consider implications of these points. First, the layout of items on the screen may—but does not necessarily—provide information for other participants about what is relevant. This is an affordance of how EDs are designed and would seem to be a major difference from paper dictionaries, a difference, though, that is unlikely to be found through either experimental or questionnaire research. In paper dictionaries, the location of any particular word with relation to the other words does not change. The location of a word on the page is unrelated to the fact that it is the word that is being looked up. With an ED, in contrast, the location on the screen of the word that is being looked up is predictable. In addition, the design of EDs can constrain how participants use them in their interaction with others. Such affordances and constraints may have implications for teachers who wish to encourage students to use a particular type of dictionary.

Second, though they do this in different ways in different episodes of dictionary use, participants orient to dictionary ownership. Even when touching or moving another's dictionary, they do not bring it into their own

space. Some researchers who have conducted research related to EDs have considered the expense of these dictionaries as possibly preventing some students from owning one (Kobayashi, 2008; Weschler & Pitts, 2000) or limiting them to dictionaries of limited quality (Chen, 2010). How the participants use their dictionaries in the excerpts analyzed above does not reveal whether they think of them as expensive, but they do treat the dictionaries as valuable objects, in the sense that their ownership is publicly recognized. This has implications for student group work, in that a student without an ED may have reason to refrain from freely using another's. Third, the configuration of material objects and participants' bodies has consequences for how shared use of a dictionary is accomplished. How the participants arrange, for example, their chairs (or have the chairs arranged for them) can have consequences for what happens during the discussion. This also has implications for group work, in that the arrangement of chairs and desks may influence how students are able to work together. Fourth, physical manipulation of an ED, such as turning it one way or another, can have interactional significance (e.g., attracting another's gaze to the dictionary). Like other material artifacts (Cekaite, 2009), EDs can be a resource with which participants organize their interaction. Another implication for group work, then, is that the usefulness of these dictionaries for participating in L2 discussions is not limited to the provision of L2 vocabulary. Fifth, contrary to what is suggested by Barrow (2009), the participants do not seem to orient to any normative organization of dictionary use. The closest to a normative orientation among participants appears in Excerpt 4, in which one participant accounts for his dictionary use and another states that dictionary use is cheating. Though this is treated as nonserious, the participant who has consulted his dictionary chooses to return it to his bag, rather than place it somewhere that would make it more accessible. An implication for group work is that students engaged in classroom tasks may or may not view dictionary use as illegitimate.

Finally, EDs can be used to accomplish a variety of local interactional objectives. In Excerpt 1b, an electronic dictionary was used to topicalize a Japanese word and its English translation that had been introduced a few minutes earlier. In other excerpts, EDs were used to accomplish repair, but in each case, a different repair organization was involved. In Excerpt 2, in addition to repair practices associated with word search, the repair was organized as other-initiated, other-repair. In Excerpt 3, also in addition to repair practices associated with word search, repair was organized as self-initiated other-repair. And in Excerpt 4, it was organized as self-initiated self-repair in

third position. As mentioned in the introduction, research on how university students use EDs has relied heavily on questionnaires. However, although such research may be able to reveal the extent to which students use EDs to, for example, read a newspaper, write a report, or participate in a discussion, careful observation and analysis is necessary to reveal the variety of tasks that EDs can be used to accomplish while reading a newspaper, writing a report, or participating in discussions. This has implications for the direction of future research on EDs.

Notes

1. The terms *electronic dictionary*, *e-dictionary*, *portable electronic dictionary*, and *pocket electronic dictionary* have all been used to describe EDs. *Electronic dictionary* has also been used to describe CD-ROM-based and online dictionaries. In this paper, I do not include these latter types.
2. The popularity of such dictionaries may not be limited to university students, but it is this population's use of EDs that has been the object of research.
3. In order to facilitate ease of reading, talk from the transcript that is quoted in the text has had the details of the talk's production removed. For example, as shown in Excerpt 1a, "*jirai*" is produced quietly and with elongation of the final vowel, but the conventions used to show this (degree signs, colon) are not reproduced in the quotation in the text.
4. At one point during this recorded discussion, Abe loans his dictionary to Baba, who opens it, consults it, closes it, and then returns it. This way of using and returning it also shows his orientation to Abe's ownership of the dictionary.

Eric Hauser received his PhD in second language acquisition from the Department of Second Language Studies, University of Hawaiʻi at Mānoa. His research is focused on interaction involving second language speakers of

English. He has recent publications in *Language Learning* and *Pragmatics and Society*.

References

Barrow, J. (2009). Electronic dictionary use in novice L2 learner interaction (Unpublished doctoral dissertation). Temple University, Japan.

Barrow, J. (2010). Electronic dictionary look-up practices of novice English learners. In T. Greer (Ed.), *Observing talk: Conversation analytic studies of second language interaction* (pp. 55-72). Tokyo: Pragmatics Special Interest Group of JALT.

Bower, J., & McMillan, B. (2007). Learner use and views of portable electronic dictionaries. In K. Bradford-Watts (Ed.), *JALT2006 Conference Proceedings* (pp. 697-708). Tokyo: JALT.

Cekaite, A. (2009). Soliciting teacher attention in an L2 classroom: Affect displays, classroom artefacts, and embodied action. *Applied Linguistics*, *30*, 26-48. http://dx.doi.org/10.1093/applin/amm057

Chen, Y. (2010). Dictionary use and EFL learning. A contrastive study of pocket electronic dictionaries and paper dictionaries. *International Journal of Lexicography, 23*, 275-306. http://dx.doi.org/10.1093/ijl/ecq013

Goodwin, M. H. (1995). Assembling a response: Setting and collaboratively constructed work talk. In P. ten Have & G. Psathas (Eds.), *Situated order: Studies in the social organization of talk and embodied activities* (pp. 173-186). Washington, DC: International Institute for Ethnomethodology and Conversation Analysis and University Press of America.

Hauser, E. (2013). Beyond intersubjectivity: Task orientation and first language use in foreign language discussions. *Pragmatics and Society, 4*, 285-316. http://dx.doi.org/10.1075/ps.4.3.02hau

Heath, C., & Luff, P. (2000). *Technology in action*. Cambridge: Cambridge University Press.

Jefferson, G. (1990). List-construction as a task and resource. In G. Psathas (Ed.), *Interaction competence* (pp. 63-92). Washington, DC: International Institute for Ethnomethodology and Conversation Analysis and University Press of America.

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 13-31). Amsterdam: John Benjamins.

Jian, H.-L., Sandnes, F. E., Law, K. M. Y., Huang, Y.-P., & Huang, Y.-M. (2009). The role of electronic pocket dictionaries as an English learning tool among Chinese students. *Journal of Computer Assisted Learning, 25*, 503-514. http://dx.doi.org/10.1111/j.1365-2729.2009.00325.x

Kobayashi, C. (2007). Comparing electronic and printed dictionaries: Their effects on lexical processing strategy use, word retention, and reading comprehension. In K. Bradford-Watts (Ed.), *JALT2006 Conference Proceedings* (pp. 657-671). Tokyo: JALT.

Kobayashi, C. (2008). The use of pocket electronic and printed dictionaries: A mixed-method study. In B. Watts, T. Muller, & M. Swanson (Eds.), *JALT2007 Conference Proceedings* (pp. 769-783). Tokyo: JALT.

Loucky, J. P. (2002). Improving access to target vocabulary using computerized bilingual dictionaries. *ReCALL, 14*, 295-314. http://dx.doi.org/10.1017/S0958344002000721

Loucky, J. P. (2003). Using computerized bilingual dictionaries to help maximize English vocabulary learning at Japanese colleges. *CALICO Journal, 21*, 105-129.

Mori, J. (2006). The workings of the Japanese token *hee* in informing sequences: An analysis of sequential context, turn shape, and prosody. *Journal of Pragmatics, 38*, 1175-1205. http://dx.doi.org/10.1016/j.pragma.2005.05.004

Schegloff, E. A. (1992). Repair after next turn: The last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology, 97*, 1295-1345.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language, 53*, 361-382.

Suchman, L. A. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge: Cambridge University Press.

Suchman, L. A. (2007). *Human-machine reconfigurations: Plans and situated actions*. Cambridge: Cambridge University Press.

Weschler, R., & Pitts, C. (2000). An experiment using electronic dictionaries with EFL students. *The Internet TESL Journal, 6*. Retrieved from http://iteslj.org/Articles/Weschler-ElectroDict.html

Whalen, J. (1995). A technology of order production: Computer-aided dispatch in public safety communications. In P. ten Have & G. Psathas (Eds.), *Situated order: Studies in the social organization of talk and embodied activities* (pp. 187-230). Washington, DC: International Institute for Ethnomethodology and Conversation Analysis and University Press of America.

Online Appendix A

Transcripts With Frames

This appendix can be downloaded from <http://jalt-publications.org/downloads/jj/jj36.1_art1.pdf>

Appendix B

Transcription Conventions

Based on Jefferson (2004)

[	start of overlap
]	end of overlap (not always marked in transcript)
=	latching (i.e., no beat of silence), or continuation of a turn across noncontiguous lines of transcript
(0.2)	silence, measured to tenths of a second
(.)	silence of less than two tenths of a second
:	elongation of sound, more colons indicate longer elongation
↑↓	shift in pitch up or down
→	line of transcript in which object of interest occurs
,	continuing intonation
.	falling intonation
?	rising intonation
up_	final flat intonation marked by underlining after last word
into	stress marked by underlining
°°	start and end of quiet talk
> <	start and end of faster talk
< >	start and end of slower talk
h	outbreath; more h-letters indicate longer outbreath
.h	inbreath; more h-letters indicate longer inbreath
(h)	laugh particle within a word
(x)	unintelligible talk; number of x-letters indicates best guess at number of syllables
(word)	best guess at a word; words in parentheses separated by slash indicate alternative hearings
(( ))	transcriber's comments in double parentheses

Modeling Complexity, Accuracy, and Fluency of Japanese Learners of English: A Structural Equation Modeling Approach

Rie Koizumi
Juntendo University

Yo In'nami
Shibaura Institute of Technology

With this study, we aimed to obtain a better understanding of the factor structure of the complexity, accuracy, and fluency (CAF) of English speaking proficiency. For this purpose, 224 Japanese junior and senior high school students with an English level of elementary to lower intermediate took an English speaking test. We transcribed what they said, computed measures to assess CAF, and used structural equation modeling (SEM) to examine whether the model in which the CAF factors are related fit the data. We found that syntactic complexity (SC), accuracy, speed fluency, and repair fluency represent distinct factors and that there are weak, moderate, or strong correlations among these factors. This generally suggests that those who speak fluently by using more words per minute tend to repair their speech more, but they also produce more accurate utterances with more clauses. We suggest pedagogical implications of considering CAF separately in teaching and assessment and benefits of using SEM for analyzing CAF.

本研究では、スピーキング熟達度における複雑さ、正確さ、流暢さ（complexity, accuracy, and fluency: CAF）の因子構造を調べる。中学生・高校生（初級から中級下レベル）の日本人学習者224名に、スピーキングテストを受けてもらった。発話をCAFの指標で数値化し、CAF因子が関連しあうモデルを共分散構造分析を用いて分析した。その結

果、統語的複雑さ、正確さ、スピードに関する流暢さ、修正に関する流暢さの4因子の相関モデルがデータに適合し、4因子は関連しあいながらも別個に捉えられることが分かった。因子間の関連は弱いものから強いものがあったが、全体的には、1分間あたりにより多くの語を使って話す学習者は、修正をより多く行うが、より正確な発話と、より多くの節を産出する傾向が見られた。指導や評価の際にCAFを別々に考慮することの重要性や、共分散構造分析でCAFを分析する利点が示された。

S econd language speaking proficiency and performance has garnered increasing attention from researchers into L2 learning and assessment. One way to conceptualize L2 proficiency and performance is to use the components of complexity, accuracy, and fluency (Housen & Kuiken, 2009; Housen, Kuiken, & Vedder, 2012). These three factors (or constructs), hereinafter abbreviated as CAF, have been extensively measured in numerous studies (e.g., Foster & Tavakoli, 2009; Robinson, 2001). Despite the wide use of CAF, some issues remain unresolved (Housen & Kuiken, 2009), such as how CAF can be measured and the extent to which CAF are interrelated. To deal with these two issues, we attempt to model CAF using data from L2 Japanese learners of English with English proficiencies of elementary to lower intermediate level by employing structural equation modeling (SEM). Explicit modeling employing SEM helps in understanding the nature of CAF and their measures.

Background

Although CAF are now often grouped together, it was only in the 1990s that pedagogical and research considerations of fluency and accuracy began to be combined with the concept of complexity (Housen & Kuiken, 2009). CAF are often measured using discourse analytic measures derived from quantifying target aspects in utterances and computing values that reflect a certain dimension of language use (see Ellis & Barkhuizen, 2005; Housen et al., 2012).

Complexity is commonly defined as "the ability to use a wide and varied range of sophisticated structures and vocabulary in the L2" (Housen et al., 2012, p. 2). According to Bulté and Housen (2012), complexity is subdivided into three types: propositional, discourse-interactional, and linguistic. The former two refer to the number of idea units produced and "the number and type of turn changes that learners initiate and the interactional moves and participation roles that they engage in" (Bulté & Housen, 2012, pp. 24-25). Linguistic complexity encompasses a wide range of linguistic features; it is further classified into four dimensions: lexical (words and collocations),

morphological (inflectional/derivational levels), syntactic (sentential, clausal, and phrasal levels), and phonological (segmental/suprasegmental levels). Among these, the most discussed and researched dimension is syntactic complexity (SC). The most examined (sub)factors underlying SC are overall and sentential subordination, and these are typically indexed by mean length of unit and clauses per unit (Bulté & Housen, 2012).

Accuracy refers to "the ability to produce target-like and error-free language" (Housen et al., 2012, p. 2) and is measured using global measures (e.g., the percentage of error-free clauses) or specific measures (e.g., the percentage of correct pronouns).

Fluency is defined as "the ability to produce the L2 with native-like rapidity, pausing, hesitation, or reformulation" (Housen et al., 2012, p. 2). Tavakoli and Skehan (2005) subcategorized fluency into speed fluency (measured by, for example, speech rate and mean length of run), repair fluency (assessed with measures of reformulation, repetition, false starts, and replacements), and breakdown fluency (operationalized as pause-related indices). Bosker, Pinget, Quené, Sanders, and de Jong (2013) showed that speed fluency and breakdown fluency contribute more to raters' fluency ratings than does repair fluency.

CAF Factors

Although CAF factors are assumed to be reflected in the measures used in previous CAF studies, empirical validation studies of CAF measures—that is, investigations into whether the measures indeed assess CAF factors—have been rare (Norris & Ortega, 2009; Sheppard, 2004; Wolfe-Quintero, Inagaki, & Kim, 1998). In order to explore the relationships among CAF factors and measures, L2 researchers have used exploratory factor analysis (e.g., Mehnert, 1998; Ortega, 1995) as well as simple correlations (e.g., Kormos & Dénes, 2004; Mora & Valls-Ferrer, 2012). In this study, we focus on previous studies using exploratory factor analysis because this method can explicitly handle both latent factors (i.e., underlying or unobserved) and observed variables. This is more appropriate for examining factor structures because with correlations only observed variables can be examined.

Table 1 summarizes nine factor-analytic studies that aimed to either identify redundancy among measures and select representative measures for further analysis (Nitta, 2007; Ortega, 1995) or explore relationships among measures and identify underlying structures (the remainder of the studies in Table 1). Unfortunately, except for the study done by Sakuragi (2011),

all studies had rather small sample sizes (ranging from 17 to 80), which is known to cause instability in factor structures. Consequently, it would be safer to reinterpret previous findings than to take them at face value. On the basis of their simulation studies, Guadagnoli and Velicer (1988) argued that a factor loading pattern of .60 is likely to be stable when the sample size is 150 or greater and that a factor loading of .80 tends to be stable even when the sample size is 50. Accordingly, only measures with loadings of .80 or above were used for our reinterpretation, although the method used may produce rather conservative interpretations.

Table 1. Previous Studies Analyzing the CAF Factor Structure of Speaking Proficiency

Study [N]	L2; Proficiency level	No. of measures included	Extraction; Rotation method	Reinterpreted factors (measures with loadings of .80 or above)
Ortega (1995) [32]	Spanish; Upper intermediate	10 (2LC, 3A, 5F)	PCA; Oblique	1. Speed fluency (words per utterance[a]; propositions per utterance[a]; unpruned syllables per second; pruned syllables per second) 2. Accuracy (percentage of correct noun-modifier agreement; percentage of correct articles)
Skehan & Foster (1997) [72]	English; Pre-intermediate	9 (1SC, 1A, 1F for 3 tasks)	PCA; Varimax	1. SC (clauses per c-unit for the first and the third tasks)
Mehnert (1998) [31]	German; Intermediate	13 (3SC, 1LC, 4A, 5F)	Not reported	1. Speed and breakdown fluency (unpruned syllables per second; pruned syllables per second; mean length of run; total pausing time) 2. Accuracy (errors per 100 words; error-free clauses per clause; number of lexical errors) 3. SC (words per c-unit; subordinate clauses per T-unit; S-nodes per T-unit)

Study [*N*]	L2; Proficiency level	No. of measures included	Extraction; Rotation method	Reinterpreted factors (measures with loadings of .80 or above)
Taki-guchi (2003) [17]	English; Elementary	31 (3SC, 6A, 11F plus 11 non-CAF measures)	PCA; Varimax	1. SC (subordinate per AS-unit; clauses per AS-unit; pruned tokens per AS-unit; pruned tokens per turn; turns per minute) 2. Speed fluency (pruned tokens per minute; unpruned tokens per minute) 3. Accuracy (errors per minute; AS-units with errors per AS-unit; errors per AS-unit; errors per pruned token; error-free clauses per clause; error-free AS-unit per AS-unit) 4. Repair fluency (AS units with disfluency markers; disfluency markers per minute; AS units with disfluency markers per AS-unit) 5. Breakdown fluency (pauses per minute; AS-units with pauses per minute)
Shep-pard (2004) [82]	English Elementary	27 (4SC, 3LC, 4A, 16F)	PCA; Varimax	1. Speed fluency (unpruned tokens per minute; pruned tokens per minute; unpruned syllables per minute; pruned syllables per minute) 2. SC (clauses per T-unit; verbs per T-unit; phrases per T-unit; pruned tokens per T-unit) 3. Breakdown fluency (3 types of percentages of pause time, with the cut-off point of 250 milliseconds, 600 milliseconds, and 1 second) 4. Accuracy (error-free clauses per clause; error-free clauses per T-unit; error-free phrase per phrase)
Skehan & Foster (2005) [61]	English; Intermedi-ate	9 (1SC, 2A, 6F)	PCA; Varimax	1. Accuracy (error-free clauses per clause; accuracy for clauses of five words or more)

Study [*N*]	L2; Proficiency level	No. of measures included	Extraction; Rotation method	Reinterpreted factors (measures with loadings of .80 or above)
Tavakoli & Skehan (2005) [80]	English; Elementary and intermediate	12 (1SC, 1A, 10F)	Not reported	1. Speed and breakdown fluency (syllables or words per minute; total amount of silence; time spent speaking; number of pauses; mean length of pause) 2. Repair fluency (number of reformulations; number of false starts)
Nitta (2007) [27]	English; Elementary and advanced	13 (3SC, 3A, 7F)	PCA; Varimax	1. Speed and breakdown fluency (total length of pauses; mean length of run; pruned tokens or syllables per minute; number of mid-clause pauses) 2. Accuracy (error-free clauses per clause; percentage of correct verb forms) 3. Others[b] (number of chaining integration devices; number of filled pauses)
Sakuragi (2011) [113]	Japanese; Intermediate and advanced	10 (2SC, 2LC, 3A, 3F)	Principal factor analysis; Promax	1. SC (clauses per AS-unit; subordinate clauses per AS unit; pruned tokens per AS-unit) 2. Accuracy (errors per clause; error-free AS-units per AS-unit; errors per AS-unit)

Note. Only studies analyzing speaking proficiency were included. Factors with two or more measures with loadings of .80 or above (rounded off) were presented. SC = syntactic complexity; LC = lexical complexity; A = accuracy; F = fluency; PCA = principal components analysis.

[a]Although Ortega originally considered the number of words per utterance and that of propositions per utterance as SC measures, she doubted the validity of such interpretation in the discussion; Norris and Ortega (2009) further interpreted the number of words per utterance as reflecting fluency in the same manner as the mean length of run. [b]Only appeared under the online planning condition.

A reinterpretation of the results of previous studies yields several findings. First, three of the studies obtained at least three factors each for SC, accuracy, and fluency (Mehnert, 1998; Sheppard, 2004; Takiguchi, 2003); two studies obtained an accuracy factor and a fluency factor (Nitta, 2007; Ortega, 1995); and one study obtained an SC factor and an accuracy factor

(Sakuragi, 2011). The others all had one factor reflecting either SC, accuracy, or fluency (e.g., Skehan & Foster, 1997, 2005; Tavakoli & Skehan, 2005). These results indicate that accuracy appeared as a factor in most cases, followed by fluency and SC, but that all three were not always present. This suggests insufficient empirical evidence about whether we can derive distinct CAF factors. Research Question 1 was designed to examine this aspect.

Second, an SC factor appeared as a single dimension in five studies (e.g., Mehnert, 1998; Sheppard, 2004). The derivation of one SC factor consistently across studies may suggest that SC dimensions—for example, overall SC and sentential-subordination SC—can be conceptually distinguished but not empirically discriminated (Pallotti, 2009). In addition, in all five studies that derived an SC factor, one measure—the number of tokens (i.e., words) divided by the number of units (e.g., T-units)—loaded on the SC factor. Only one study, Skehan and Foster (1997), did not use this measure. The interpretation of this measure, called *mean length of unit* or *unit length*, has been controversial. Although the mean length of run, or the number of syllables/tokens per unit primarily related to pause or repair, is interpreted as fluency (e.g., Segalowitz & Freed, 2004; Tavakoli & Skehan, 2005), the number of tokens per primarily syntactic unit (e.g., T-unit) has two distinct accounts: fluency (e.g., Ishikawa, 2007; Robinson, 2001; Wolfe-Quintero et al., 1998) and SC (e.g., Bulté & Housen, 2012; Koizumi, 2005b; Norris & Ortega, 2009). Results of CAF factor-analytic studies support the latter interpretation (e.g., Mehnert, 1998; Sakuragi, 2011) because the number of tokens per syntactic unit loaded on an SC factor. In addition, it was found that no lexical complexity factor emerged as distinct. This may be attributable to the limited number and types of measures of lexical complexity employed in factor-analytic studies.

The accuracy factor, if any, consistently appeared as a single dimension (e.g., Nitta, 2007; Ortega, 1995). This supports Pallotti's (2009) observation that the accuracy factor is stable in nature. Further, fluency has been found to comprise up to three factors. It appears that more factors are extracted when more relevant fluency measures are involved (e.g., 3 fluency factors using 11 fluency measures in Takiguchi, 2003). The interpretations of factors suggest that the speed dimension is often linked to the breakdown dimension. Moreover, the speed dimension tends to be a primary component of fluency, whereas the repair dimension is found as a separate factor and a secondary dimension of fluency.

Although previous factor-analytic studies have provided an invaluable foundation for clarifying CAF factors and measures, three methodological

issues must be addressed to derive stronger evidence. First, it is unclear if the data satisfied the statistical assumptions for using exploratory factor analysis. Although some studies (e.g., Ortega, 1995; Sakuragi, 2011) conducted Bartlett's test of sphericity to determine if a correlation matrix was adequate for factor analysis, no studies reported multivariate normality—another essential assumption for factor analysis: "all variables, and all linear combinations of variables, are normally distributed" (Tabachnick & Fidell, 2007, p. 613). Second—as Plonsky and Gass (2011) argued about L2 studies in general—some CAF studies (e.g., Mehnert, 1998; Tavakoli & Skehan, 2005) did not report how factor analysis was conducted, such as what extraction and rotation methods were used or how the number of factors was determined. This is troubling because results would change according to these specifications (e.g., Tabachnick & Fidell, 2007). Further, except for studies done by Ortega (1995) and Sakuragi (2011), varimax rotation was the rotation method of choice, which assumes no correlations among extracted factors. However, this is often too strong an assumption to hold because interrelationships between factors can usually be hypothesized; thus, oblique (e.g., Promax) rotation is recommended.

Third, exploratory factor analysis is of limited value due to its data-driven nature. Given a growing number of previous studies on CAF that permit the construction of a theory-based model, SEM is a more suitable method. However, thus far, no studies have employed this method for CAF analyses. The following are the main advantages of SEM (Byrne, 2006). First, SEM uses a confirmatory, hypothesis-testing method. It can model not only observed variables but also latent variables (i.e., factors) and can flexibly model complex relationships on the basis of previous findings. Second, it can separate measurement errors from observed and latent variables and estimate relationships among the variables that are being investigated, thereby statistically controlling for such errors. One source of errors is the variability caused by task differences—earlier studies suggested that task variations (e.g., cognitive demand of tasks) led to different speaking performances (e.g., Tavakoli & Skehan, 2005; Robinson, 2001). As SEM is a large-sample technique (usually requiring a sample size of at least 100), its application to the investigation of the CAF structure is rather difficult in studies in which sample size is considerably smaller. In order to take full advantage of SEM, we collected a large sample, tested statistical assumptions, and examined the CAF factor structure.

Relationships Among CAF

A second aspect investigated in this study is the question of how CAF are interrelated. Norris and Ortega (2009) indicated the need for research into revealing the interdependence and dynamism of CAF using multivariate modeling such as SEM. Generally, positive and relatively strong relationships are predicted because CAF are expected to improve gradually as learners' proficiency increases, although not necessarily simultaneously. However, previous studies have reported divergent degrees of correlations, even among learners with a wide range of proficiency.

For example, a weak correlation was reported in Sakuragi (2011), which documented the relationship between SC and accuracy (r = .19) among 113 intermediate and advanced learners of Japanese. Ortega (1995) also presented low correlations between accuracy and speed fluency (r = .08 to .22) among 32 upper intermediate learners of Spanish. Further, Koizumi (2005b) reported marginal to fairly weak correlations among SC, accuracy, speed fluency, and repair fluency (r = -.21 to .47) among 74 elementary to upper elementary Japanese learners of English. Kormos and Dénes (2004) reported a moderate correlation between accuracy and speed fluency (r_s = .66) and a low correlation between speed fluency and repair fluency (r_s = -.19) among 16 low-intermediate and advanced Hungarian learners of English. These varied correlations suggest the need for further investigation and lead to Research Question 2.

The Current Study

To clarify the CAF structure, we examine factors of SC, accuracy, and fluency (fluency is further divided into speed fluency and repair fluency), as well as the relationship among these factors. Two research questions were asked with a specific focus on Japanese learners of English at the elementary to lower intermediate level.

RQ1: Do complexity, accuracy, and fluency (CAF) represent distinct factors?

RQ2: How are complexity, accuracy, and fluency (CAF) interrelated?

Method

Participants

The participants were 224 Japanese learners of English—97 males and 127 females—attending 10 junior or senior high schools, aged from 14 to

18. Their first language was Japanese. They had received EFL instruction at secondary schools in Japan for from 2 to 5 years. The overall English proficiency levels on the Eiken Test (Society for Testing English Proficiency [STEP], 2011) were reported by the participants and ranged from Grades 5 (2%) to 2 (4%), with the majority at Grades Pre-2 or 3 (61%), although 23% reported no experience of taking the Eiken Test. According to the STEP (2011), Eiken Test Grades 2 to 5 are roughly equivalent to the A1 to B1 levels of the Common European Framework of Reference for languages (Council of Europe, 2001). Thus, the participants were considered to have novice- to lower intermediate-level English proficiency. They were selected for participation in this study from a larger sample only if they took a speaking test and produced at least one clause for every speaking task.

Instrument

The students took a speaking test that contained five tasks to elicit real-time monologues without pretask planning time (Koizumi, 2005a). The test lasted for 15 minutes in a tape-mediated format. Task 1 was a self-introduction task, Tasks 3 and 4 involved describing a single picture, and Tasks 2 and 5 involved explaining the differences between two pictures (see Appendix for a sample of utterances). We used these five tasks to tap wider areas of speaking proficiency. The output from learners was limited; the mean of the number of tokens for each task ranged from 25.39 (SD = 10.98) in Task 3 to 37.14 (SD = 13.64) in Task 1.

Analyses

We created the coding scheme using Foster, Tonkyn, and Wigglesworth's (2000) definitions. Raters (native speakers and highly proficient Japanese learners of English) practiced coding and later, using the scheme, independently coded one-third (randomly sampled) of the transcribed utterances for each task (45 seconds for each task; a total of 225 seconds) for features such as the number of AS-units. The number of raters varied depending on the coded features: Four raters were used for assessing error-free clauses because of difficulty in judgment; two raters were employed for the other features. The inter-coder reliabilities were found to be high (e.g., r = .86 to 1.00 for the number of AS-units, clauses, and disfluency markers; α = .86 to .93 for the number of error-free clauses). Because the features were measured on interval scales, we used Pearson product-moment correlations for two raters and Cronbach's alpha for four raters. Further, we resolved dis-

agreement through discussion and created the final detailed coding scheme that clarified the aspects on which raters diverged and that required little judgment from raters.

The remainder of the transcripts were coded by a single rater (the first author) who had judged one-third of the transcripts for all the coded features; she coded them while examining the coding scheme carefully. This method can be justified because the inter-coder reliability among raters for one-third of the transcripts was sufficiently high and because this is a common procedure for coding data (see Révész, 2012).

For the analysis of speaking proficiency, we computed five discourse analytic measures for each task (see Table 2). Similar measures were initially computed but excluded because of high correlations with the remaining measures (e.g., number of disfluency markers per token) and inconsistent results across tasks (number of tokens per clause). We did not include pause-based measures due to poor recording conditions that hampered such in-depth analysis.

Table 2. Summary of Five Measures

Factor	Code	Measure	Source example
Syntactic complexity	SC1	Overall SC: AS-unit length: No. of tokens per AS-unit	Mehnert (1998)
(SC)	SC2	Sentential-subordination SC: No. of clauses per AS-unit	Tavakoli & Skehan (2005)
Accuracy	A	No. of error-free clauses per clause	Skehan & Foster (2005)
Fluency	F1	Speed fluency: No. of tokens per minute	Sheppard (2004)
	F2	Repair fluency: No. of disfluency markers per minute	Sheppard (2004)

Note. Tokens (i.e., words) refers to pruned tokens after disfluency markers were excluded (i.e., functionless repetitions, self-repairs, and filled pauses, such as *mm*, *ah*). The definition of clauses was based on Foster et al. (2000; for instance, the utterance "I like reading books" had two clauses: *I like* and *reading books*). Abbreviations in this table are used in text and figures.

For SEM analyses, we used EQS (Version 6.1; Bentler, 2010), but drew diagrams using Amos (Version 7.0.0; Arbuckle, 2006) for visual display. The SEM analyses enable us to examine whether a model depicting relationships between variables that is based on a theory or the literature, or both, fits the data. If it fits the data, it implies that relationships among variables specified in the model accord well with relationships among variables in the data. Then, we can empirically interpret the findings to mean that the model represents the data well and that the data has a factor structure in which factors and observed variables are related as specified in the model.

The requisites for appropriate SEM practices (e.g., Byrne 2006; In'nami & Koizumi, 2011) involve normality, parameter estimation methods, model fit indices used, missing data treatment, and sample size. Univariate and multivariate normality of measures was judged on the basis of skewness and kurtosis values and Mardia's normalized estimate and found to be violated. Thus, the robust maximum likelihood method was used. One of the factor loadings from each factor was fixed to 1.00 for scale identification. Model fit was checked by the comparative fit index (CFI) of 0.90 or above (Arbuckle & Wothke, 1995), root mean square error of approximation (RMSEA) of 0.08 or below (Browne & Cudeck, 1993), standardized root mean square residual (SRMR) of .08 or below (Hu & Bentler, 1999), and other indices. There were no missing data. The sample size exceeded 200, which is considered large according to Kline's (2005) guidelines. Further, intervariable Pearson product-moment correlations (r = –.13 to .74) were not so high as to cause problems of multicollinearity (r = .90 or above; Tabachnick & Fidell, 2007).

Results

We followed four steps. First, we constructed a model with only one latent factor assessed by five observed variables (one measure from each task). Figure 1 depicts a model (Model 1) of speed fluency in which a factor is represented by an oval (Speed fluency), observed variables are represented by rectangles (F1 from five tasks; T1F1 [Task 1 F1] to T5F1), and measurement errors (e1 to e5) are represented by circles. One-headed arrows depict the influence of the speed fluency factor on the five variables, which are also affected by five errors. Based on existing literature, this model specifies that there are five F1 variables underlying a speed fluency factor, but that there are some aspects of F1 variables that are unexplained by the factor but explained by errors. We also constructed four other models separately (for overall SC, sentential-subordination SC, accuracy, and repair fluency) but have not included them here because of lack of space.

Second, we tested whether each model fits the data. Fit statistics for five models indicate that all models fit the data (e.g., CFI = .95 to 1.00, RMSEA = 0.00 to 0.077, SRMR = .02 to .04), thereby indicating that the measures used represented each factor well.
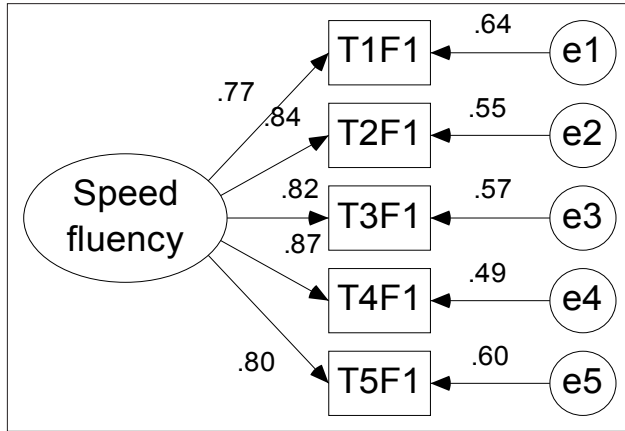
Figure 1. One-factor model for speed fluency (Model 1).
T1 = Task 1; F1 = number of tokens per minute; e = measurement error.
Standardized estimates are shown. All the testable path coefficients were
significant.

Third, we constructed a model (Model 2) with five CAF factors (overall SC, sentential-subordination SC, accuracy, speed fluency, and repair fluency), all of which were related to one another (this model has not been displayed in this paper due to space limitations). This model did not fit the data (e.g., CFI = .84; RMSEA = 0.07 [95% confidence interval: 0.06, 0.08]; SRMR = .07) mainly because the correlation between overall SC and sentential-subordination SC was too high (r = 1.03). We retained a factor of sentential-subordination SC because SC2 (number of clauses per AS-unit) is considered a more typical measure of SC than SC1 (number of tokens per AS-unit), which is occasionally used as a fluency measure. Because of the strong correlation between the two SC factors, the results derived from SC2 can be considered to be applicable to SC1, and the sentential-subordination SC factor is hereinafter interpreted as SC in general.

Finally, we tested a model with four factors (SC, accuracy, speed fluency, and repair fluency) that were correlated with one another (indicated by

two-headed arrows), as evident in Model 3 in Figure 2. Fit statistics of this model were sufficient (e.g., CFI = .93; RMSEA = 0.05 [0.04, 0.06]; SRMR = .06). Other competing models did not fit the data well, such as one with a higher order speaking proficiency factor represented by the four factors (e.g., CFI = .84; RMSEA = 0.08 [0.07, 0.09]; SRMR = .17) and another with a unitary speaking proficiency factor without any CAF factors (e.g., CFI = .73; RMSEA = 0.10 [0.08, 0.11]; SRMR = .10).

The standardized estimates range from –1.00 to 1.00 and are interpreted in the same manner as the correlation and regression coefficients, with values close to zero indicating marginal associations and those close to –1.00 or 1.00 indicating strong associations. A good fit of Model 3 to the data suggests two types of relationships: those between a factor and each observed variable and those among factors. First, all observed variables were shown to reflect each factor well, thereby indicating that the variables assessed each factor appropriately. Further, path coefficients from speed fluency and repair fluency factors to observed variables were found to be strong (β = .65 to .88), whereas those from accuracy and SC factors were moderate (β = .20 to .57). This indicates that there was less variation in the path coefficients for fluency than in those for SC and accuracy, thereby suggesting that fluency measures may be more generalizable across tasks.

Second, the model indicates positive but varied degrees of relationships among CAF factors: SC was more closely related to accuracy (r = .88) than to speed fluency and repair fluency (r = .63 and .43, respectively). Further, accuracy was more closely related to speed fluency (r = .35) than to repair fluency (r = .13), and the two fluency factors were moderately correlated (r = .66).

Discussion

Do CAF Represent Distinct Factors?

A good fit of Model 3 in Figure 2 suggests that the answer to the question of whether CAF represent distinct factors is affirmative. A structure with distinct CAF factors accords well with Mehnert (1998), Sheppard (2004), and Takiguchi (2003) but does not with others (e.g., Nitta, 2007; Ortega, 1995). Researchers in all previous studies used exploratory factor analysis and attempted to extract CAF factors reflected in different measures, whereas we used SEM with each factor reflected in the same measures from five tasks. The results indicate that our approach is useful for examining the distinctiveness of CAF factors.

The result that two fluency factors (speed and repair) were moderately positively associated (r = .66) is indicative of their distinct yet related na-
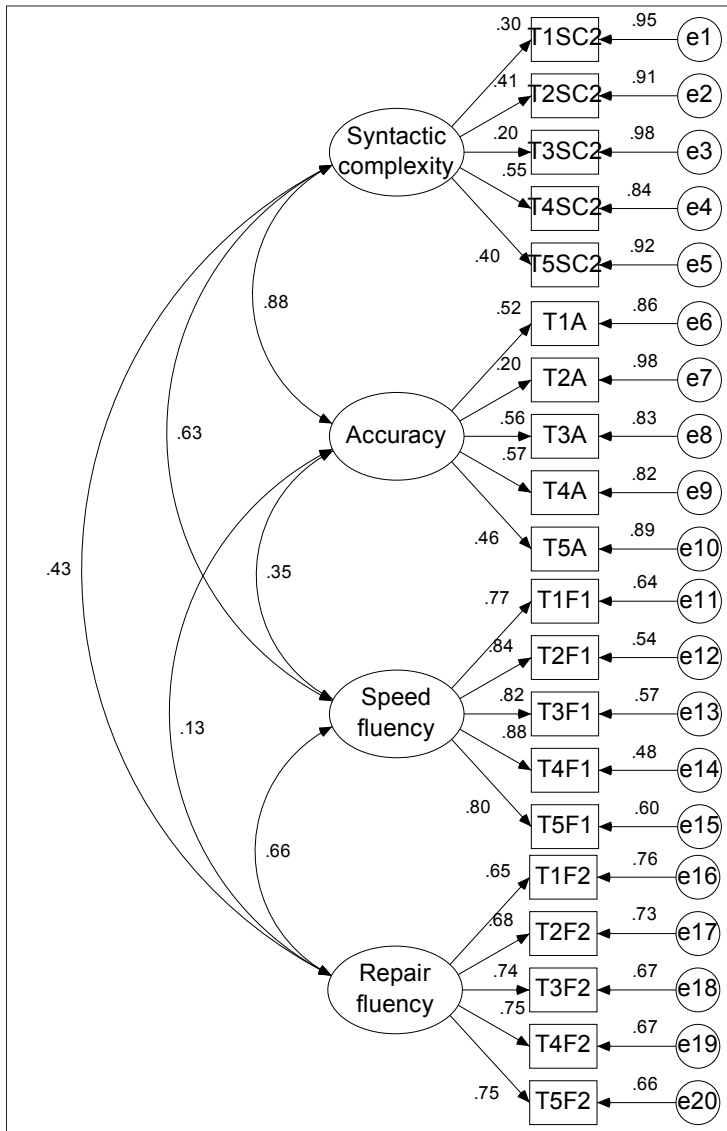
Figure 2. Four-factor correlated model for CAF (Model 3).
T1 = Task 1; F1 = number of tokens per minute; e = measurement error.
Standardized estimates are shown. All the testable path coefficients were
significant.

ture. This is in line with previous exploratory factor analyses that reveal the distinct characteristics of speed and repair fluency (Takiguchi, 2003; Tavakoli & Skehan, 2005), but other studies reported only negative correlations between speed and repair fluency (e.g., r = -.19 in Kormos and Dénes, 2004). The positive correlation between speed fluency and repair fluency factors in our study implies that—given that F2, which taps repair fluency, is calculated by the number of *disfluency markers* per minute—speakers who produce more tokens (excluding disfluency markers) tend to use *more* repetitions and self-repairs and produce *more* filled pauses in their utterances. This could be explained by the participants' lower proficiency levels and the targeting of a wide range of proficiency levels. It is possible that at this proficiency range, those who try to search for and utter words more rapidly are unable to avoid hesitation due to insufficient automatized skills, as reported in Wood (2010). Alternatively, learners with higher proficiency can monitor their utterances (Kormos, 2006); therefore, they repair their speech more while speaking faster. However, we also found a very weak relationship between accuracy and repair fluency (r = .13), in line with Koizumi (2005b; r = -.05 to .21), which suggests that more repairing does not likely lead to more accurate speech according to the proficiency range of the current study.

In addition, although a strong claim cannot be made due to the lack of a model fit, a very strong relationship (r = 1.03) between factors of overall SC and sentential-subordination SC in Model 2 indicates that the length of the AS-unit is an SC measure, which supports all previous studies (e.g., Norris & Ortega, 2009; Sakuragi, 2011). It also indicates that although they are differentiated conceptually, dimensions of overall SC and sentential-subordination SC are not empirically distinct among learners, or at least among learners at a lower proficiency level.

How Are CAF Interrelated?

As displayed in Figure 2, CAF were found to be independent but related to varying degrees (r = .13 to .88). Overall, the model suggests that those who try to speak fluently by using more words per minute tend to repair their speech more; however, they also produce more accurate utterances with a greater number of clauses and longer units (sentences). Further, the results also indicate that as learners progress from beginning to lower intermediate levels, they develop the ability to produce such speech, thereby gradually improving SC, accuracy, and speed fluency (although not necessarily synchronously).

There were moderate or strong positive correlations of SC with accuracy ($r = .88$) and speed fluency ($r = .63$), whereas there was a weak relationship between accuracy and speed fluency ($r = .35$). It is speculated that improvement in fluency may lead to enhanced SC, which may result in heightened accuracy; that is, when learners learn to speak faster, they may gradually come to use a greater number of clauses and longer units (sentences) and subsequently may produce more accurate utterances. Such correlation patterns among CAF factors were not evident in previous studies. Previous studies (e.g., Ortega, 1995; Sakuragi, 2011) generally showed similar or weaker relationships than those revealed in our results (e.g., between accuracy and speed fluency, $r = .08$ to $.22$ in Ortega, 1995 vs. $r = .35$ in our study). The exception is relationships between accuracy and speed fluency in Kormos and Dénes (2004; $r_s = .66$ vs. $r = .35$ in our study). The higher correlations in our study may be partially because of the different statistical methods used. These results also suggest that the strengths of relationships vary across contexts.

Conclusion

The current study showed that CAF represent distinct factors that are correlated to varying degrees among elementary to lower intermediate Japanese learners of English. This insight into the CAF factor structure using a rigorous statistical method makes several contributions to the field.

The key pedagogical implication derived from this study is that English language teachers and testers should consider CAF factors of speaking proficiency separately. In planning their curricula and speaking instructions, teachers must carefully consider which of the CAF factors they should aim to enhance and how. In analytically assessing speaking proficiency, test makers should contemplate whether and to what extent to include SC, accuracy, and fluency in their rating criteria because they are all essential elements of speaking proficiency. The manner in which practitioners use this information would vary depending on the context. Some may decide to focus on all three; others may alter aspects to emphasize across activities and tasks, classes, or assessments, thereby aiming to achieve the development and assessment of balanced speaking proficiency; others may exclude SC and focus on accuracy and fluency for the criteria, based on moderate and strong relationships of SC with accuracy and fluency. Additionally, teachers should know that at lower proficiency levels, repetitions, self-corrections, and filled pauses tend to increase along with an increase in words uttered. Given the importance of speed fluency over repair fluency (Bosker et al., 2013), teach-

ers should devote more attention to the development and assessment of speed fluency rather than repair fluency and encourage learners to speak more rather than discourage the use of words for repair.

Our results may be limited to the study context. We targeted Japanese elementary- to lower intermediate-level learners of English, using speaking tasks that elicited basic monologues and a limited number of speaking measures. Greater generalizability of results would need replication studies in different contexts, for example, by using more cognitively challenging tasks (e.g., discussions and debates). In contrast, the strengths of our study are that it includes a larger number of learners than other CAF studies and involves meticulous analyses using SEM. SEM enabled us to separate measurement errors from variables of interest in the model and conduct a more rigorous analysis of relationships in a confirmatory manner on the basis of previous studies. The following example underscores the benefits of using SEM. Accuracy and SC factors were found to be strongly correlated (r = .88), whereas simple (zero-order) correlations between accuracy (A) and SC (SC2) from the same task—when measurement error was not controlled for—were much lower (r = .02 to .37). This clearly illustrates the importance of controlling for measurement error by using SEM. Although SEM requires the use of large sample sizes, a confirmatory approach to analyzing the factor structure of CAF has helped deepen our understanding of these factors.

Acknowledgment

Rie Koizumi is an Associate Professor at Juntendo University. **Yo In'nami** is an Associate Professor at Shibaura Institute of Technology. Their area of interest is modeling language ability structures. The following website contains supplementary material for this article: http://www7b.biglobe. ne.jp/~koizumi/Koizumi_research.html

References

Arbuckle, J. L. (2006). Amos (Version 7.0.0) [Computer software]. Spring House, PA: Amos Development Corporation.

Arbuckle, J. L., & Wothke, W. (1995). *Amos 4.0 user's guide.* Chicago: SmallWaters Corporation.

Bentler, P. M. (2010). EQS (Version 6.1) for Windows [Computer software]. Encino, CA: Multivariate Software.

Bosker, H. R., Pinget, A.-F., Quené, H., Sanders, T., & de Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing, 30,* 159-175. http://dx.doi.org/10.1177/0265532212455394

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21-46). Amsterdam: John Benjamins.

Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming* (2nd ed.). Hillsdale, NJ: Erlbaum.

Council of Europe. (2001). *Common European Framework of Reference for languages: Learning, teaching, assessment.* Cambridge: Cambridge University Press.

Ellis, R., & Barkhuizen, G. (2005). *Analyzing learner language.* Oxford: Oxford University Press.

Foster, P., & Tavakoli, P. (2009). Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language Learning, 59,* 866-896. http://dx.doi.org/10.1111/j.1467-9922.2009.00528.x

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21,* 354-375. http://dx.doi.org/10.1093/applin/21.3.354

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin, 103,* 265-275.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30,* 461-473. http://dx.doi.org/10.1093/applin/amp048

Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy and fluency. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 1-20). Amsterdam: John Benjamins.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55. http://dx.doi.org/10.1080/10705519909540118

In'nami, Y., & Koizumi, R. (2011). Structural equation modeling in language testing and learning research: A review. *Language Assessment Quarterly, 8,* 250-276. http://dx.doi.org/10.1080/15434303.2011.565844

Ishikawa, T. (2007). The effect of manipulating task complexity along the [+/- Here-and-Now] dimension on L2 written narrative discourse. In M. P. Mayo García (Ed.), *Investigating tasks in formal language learning* (pp. 136-156). Clevedon, UK: Multilingual Matters.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.

Koizumi, R. (2005a). *Relationships between productive vocabulary knowledge and speaking performance of Japanese learners of English at the novice level* (Unpublished doctoral dissertation). University of Tsukuba, Japan. Retrieved from http://www.tulips.tsukuba.ac.jp/limedio/dlam/B25/B2599596/1.pdf

Koizumi, R. (2005b). Speaking performance measures of fluency, accuracy, syntactic complexity, and lexical complexity. *JABAET (Japan-Britain Association for English Teaching) Journal, 9,* 5-33.

Kormos, J. (2006). *Speech production and second language acquisition.* Mahwah, NJ: Erlbaum.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32,* 145-164. http://dx.doi.org/http://dx.doi.org/10.1016/j.system.2004.01.001

Mehnert, U. (1998). The effects of different length of time for planning on second language performance. *Studies in Second Language Acquisition, 20,* 83-108.

Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal Instruction and study abroad learning contexts. *TESOL Quarterly, 46,* 610-641. http://dx.doi.org/10.1002/tesq.34

Nitta, R. (2007). *The focus-on-form effects of strategic and on-line planning: An analysis of Japanese oral performance and verbal reports* (Unpublished doctoral dissertation). University of Warwick, UK.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30,* 555-578. http://dx.doi.org/10.1093/applin/amp044

Ortega, L. A.-O. (1995). *Planning and second language oral performance* (Unpublished MA thesis). University of Hawai'i.

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics, 30,* 590-601. http://dx.doi.org/10.1093/applin/amp045

Plonsky, L., & Gass, S. (2011). Quantitative research methods, study quality, and outcomes: The case of interaction research. *Language Learning, 61,* 325-366. http://dx.doi.org/10.1111/j.1467-9922.2011.00640.x

Révész, A. (2012). Coding second language data validly and reliably. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 203-221). West Sussex, UK: Wiley-Blackwell.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22,* 27-57. http://dx.doi.org/10.1093/applin/22.1.27

Sakuragi, T. (2011). Fukuzatsusa, seikakusa, ryuchosashihyo no koseigainendatosei no kensho [The construct validity of the measures of complexity, accuracy, and fluency: Analyzing the speaking performance of learners of Japanese]. *JALT Journal, 33,* 157-174.

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition. *Studies in Second Language Acquisition, 26,* 173-199. http://dx.doi.org/10.1017/S0272263104262027

Sheppard, C. (2004). The measurement of second language production: The validity of fluency, accuracy, and complexity. *ICU (International Christian University) Language Research Bulletin, 19,* 139-156.

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research, 1,* 185-211. http://dx.doi.org/10.1177/136216889700100302

Skehan, P., & Foster, P. (2005). Strategic and on-line planning: The influence of surprise information and task time on second language performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 193-216)*.* Amsterdam: John Benjamins.

Society for Testing English Proficiency (STEP). (2011). *Eiken: Test in Practical English Proficiency.* Retrieved from http://stepeiken.org/

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Needham Heights, MA: Allyn & Bacon.

Takiguchi, H. (2003). *A study of the development of speaking skills within the framework of fluency, accuracy and complexity among Japanese EFL junior high school students* (Unpublished MA thesis). Joetsu University of Education, Niigata, Japan.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 238-273). Amsterdam: John Benjamins.

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy & complexity.* Honolulu: University of Hawai'i Press.

Wood, D. (2010). *Formulaic language and second language speech fluency: Background, evidence and classroom applications.* London: Continuum.

Appendix

Sample of Utterances From Five Tasks

A male participant: 1st-year senior high school student studying English for 3.5 years—claimed to have Eiken Grade Pre-2.

Task 1: Self-introduction

My name is * *. I have a sister. Her name is *. My parents {are in} are normal. My friends are many in my school.

Task 2: Comparison of two pictures

The windows is opened. The door's color is blue. {There are} there is a cow. There is a tree is around. There are four windows at the house.

Task 3: Picture description

A girl is washing a cup in the kitchen. The woman help the girl to washing. There are many books on

Task 4: Picture description

A man and a girl is riding a bike by the lake. There are many trees by the lake. The weather is very good.

Task 5: Comparison of two pictures

I think :: the apple before is one. But after, the apple is half. And {the} the book is mine before. But after, the book is name jiro. Another, there

Note. * = The student said a name. { } = repetitions, self-corrections, and other functionless words uttered (words in { } were ignored in accuracy rating and token counting). :: = subordinate clause

Contrastive Interlanguage Analysis of Discourse Markers Used by Nonnative and Native English Speakers[1]

Kazunari Shimada
Takasaki University of Health and Welfare

In this study, the use of discourse markers (DMs) in the speech of Japanese learners of English was investigated. To explore the features of their DM use, corpora of nonnative and native English speakers' speech were analysed using the methodology called Contrastive Interlanguage Analysis. A frequency analysis of DMs revealed significant differences between Japanese learners' and native speakers' speech, supporting earlier findings. Quantitative and qualitative analyses of the learner corpus data suggest that Japanese learners may use the marker *so* more frequently than other nonnative English learners, while also using certain interpersonal or cognitive function markers such as *you know*, *I mean*, and *just* less frequently. The findings suggest the need for language instructors and materials writers to understand the characteristics of Japanese learners' interlanguage and to provide them with appropriately designed DM input.

　本研究は、日本人英語学習者の話し言葉における談話標識（discourse markers: DMs）の使用を調べたものである。日本人英語学習者のDMs使用の特徴を探るために、対照中間言語分析の手法に基づき、非英語母語話者と英語母語話者の話し言葉コーパスを分析した。まず、日本人英語学習者と英語母語話者の話し言葉におけるDMsの使用頻度を分析したところ、先行研究と同じく、大きな差が見られた。次に、非英語母語話者の話し言葉を量的・質的の両面で分析した結果、日本人英語学習者が、他の非英語母語話者に比べて*so*を多く使用し、*you know, I mean, just*などの対人関係的、認知的機能をもつDMsをあまり使用しないことが明らかになった。その結果は、教師や教材作成者が日本人英語学習者の中間言語の特徴を理解し、学習者に対して慎重にDMsをインプットしていく必要性があることを示唆している。

D iscourse markers (DMs) are lexical items whose pragmatic functions play a crucial role in speech communication: Speakers use them to create textual coherence in interaction, as well as to express their own feelings or stances (Carter & McCarthy, 2006). For example, *OK/ okay*, *really*, and *right* are used to respond to a speaker's utterance and to suggest agreement, alignment, or active listening. *But*, *first*, and *then* serve to organise discourse structure. Words like these are tools that enable speakers to convey their meanings to their listeners. Additionally, even if spoken sentences or phrases are grammatically correct, the lack of DMs may make it difficult to attract listeners' attention in a polite way (Romero-Trillo, 2002) and may create a negative impression of being uncollaborative or awkward in conversation (Svartvik, 1980). Therefore, DMs are of special importance to nonnative speakers (NNSs), who can use them to compensate for limited English language proficiency and to improve the comprehensibility of their messages (e.g., Tyler, Jefferies, & Davies, 1988; Williams, 1992).

Considerable interest has emerged in the roles and functions of individual DMs such as *because*, *oh*, and *well* (e.g., Blakemore, 2002; Fraser, 1999, 2009; Schiffrin, 1987). The development of corpus linguistics has enabled datadriven quantitative and qualitative analyses of the use of DMs by native speakers (NSs) of English (e.g., Lenk, 1998; McCarthy & Handford, 2004). However, a relatively limited amount of research has been conducted concerning DM use in terms of second language acquisition, especially in the Japanese EFL context (see Hays, 1992; Shimada, 2011).

Positioned against this contextual background, the present study was focused on DM use in the speech of Japanese English learners. The methodology followed Granger's (1996, 2002) Contrastive Interlanguage Analysis (CIA), a corpus-based approach that employs two types of comparisons: "between native language and learner language (L1 vs L2) and between different varieties of interlanguage (L2 vs L2)" (Granger, 2009, p. 18). The CIA approach has been applied in a number of corpus studies (e.g., Ädel, 2006; Granger & Tyson, 1996), and it offers insights into the nature of interlanguage as well as aids in the identification of usage trends (e.g., overuse, underuse, and misuse) in learners' speech and writing. Thus, the aim of the present study was to investigate differences in the use of DMs (a) between Japanese L2 speakers and NSs of English, and (b) between nonnative English learners with different L1 backgrounds (Japanese, Chinese, Dutch, German, French, and Spanish).

Literature Review and Research Questions

DMs in Spoken English

DMs have been defined by researchers in a number of different ways; however, there is generally a consensus that they mainly serve syntactic and pragmatic functions in discourse. Fraser (1999, 2009) addressed their syntactic functions and considered them to be linguistic items signalling a relationship between two segments of discourse. He argued that a DM must be included as an integral syntactic part of its next discourse segment. The DMs are italicized in the following examples:

1. a. Jones died last night. *But* he had been very ill for a long time.

    b. I went to Boston first *and* later on, went to Cape Cod.

    c. The water wouldn't boil, *so* we couldn't make any tea. (Fraser, 2009, p. 294)

In other words, the purpose of each marker in examples 1a, 1b, and 1c is to make coherent links between one discourse segment and another.

In spoken English, DMs often execute pragmatic functions. Schiffrin (1987) stated that they serve as contextual coordinators for establishing or maintaining a relationship between speaker and hearer.

2. Zelda:　Are you from Philadelphia?

    Sally:　*Well* I grew up uh out in the suburbs. And then I lived for about seven years up in upstate New York. And then I came back here t'go to college. (Schiffrin, 1987, p. 106)

In example 2, Sally uses *well* as a signal that she cannot give a clear answer to Zelda's yes-no question—in other words, that her pragmatic contribution is at odds with her interlocutor's expectations. Thus, as Schiffrin pointed out, the marker *well* plays the role of contextual coordinator, marking a juncture between a speaker's intention and a hearer's interpretation.

Additionally, Schiffrin examined discourse particles such as *I mean*, *you know*, *oh*, and *like*. These items do not serve essential syntactic functions; rather, they are optional devices through which speakers can shape their utterances to affect hearers' knowledge.

3. a. *I mean* I may be wrong, but I'm—*I mean* that's what I'm—that's my opinion.

    b. We have some *y'know*. (Schiffrin, 1987, pp. 34-35)

Despite the fact that their predominant function is pragmatic instead of syntactic, markers such as those in examples 3a and 3b are ubiquitous in everyday spoken English. The markers in 3a and 3b play a role in indicating the speakers' intention to keep conversation going, and help the hearers focus on the upcoming words. Schiffrin's definition of DMs, then, was broader than Fraser's (1999, 2009), and her model illustrated features of the spoken mode in more detail.

Fung and Carter (2007) also examined the spoken mode, and they incorporated Schiffrin's (1987) model while proposing a functional paradigm of DMs drawn from their analysis of spoken English data produced by NSs and

Table 1. A Functional Paradigm of DMs in Speech

Category	Discourse functions and markers
Interpersonal	Marking shared knowledge, indicating attitudes, or showing responses:
	absolutely, actually, basically, exactly, great, I see, I think, just, kind of, like, listen, obviously, oh, oh great, OK/okay, really, right/alright, see, sort of, sure, to be frank, to be honest, well, yeah, yes, you know, you see
Referential	Indicating relationship between utterances:
	and, anyway, because/'cause, but, cos, however, likewise, nevertheless, or, similarly, so, yet
Structural	Organising or managing the direction of conversations:
	and, finally, first, firstly, how about, let me conclude the discussion, let's discuss, let's start, next, now, OK/okay, right/alright, second, secondly, so, then, well, what about, yeah
Cognitive	Denoting thinking process, or reformulating utterance:
	and, I mean, I see, I think, in other words, like, sort of, that is, to put it in another way, well, what I mean is, you know

Note. Adapted from "Discourse markers and spoken English: Native and learner use in pedagogic settings," by L. Fung and R. Carter, 2007, *Applied Linguistics*, *28*, p. 418. Some DMs such as *and*, *I think*, and *well* have multiple functions in discourse.

NNSs. They identified 57 common English DMs and classified them into four categories: interpersonal, referential, structural, and cognitive (see Table 1). This taxonomy is an extensive one, useful for characterising a large number of DMs in spoken English.

Learner Corpus Analysis of DM Use

Despite the widely recognised importance of DMs in spoken discourse, there have been only a limited number of studies examining the use of DMs by language learners. Romero-Trillo (2002) and Müller (2004) conducted corpus-driven comparisons of DM use by NSs and NNSs, and their results suggested that the use of certain DMs was influenced by the L1 of NNSs. Romero-Trillo quantitatively analysed spoken English data from Spanish children and adults. He found that Spanish children overused the English word *listen* due to the influence of its high-frequency counterpart in their L1 speech. Similarly, Müller compared the use of *well* and *so* by German speakers of English with their use in the speech of American NSs and found that German speakers used *well* much more frequently, and *so* much less frequently, than American NSs did. Müller pointed out that both DMs were translated as the German adverb *also*, and that German speakers might have a preference for *well* in order to avoid confusing English *so* and German *so*. In addition, Aijmer (2004) and Fung and Carter (2007) conducted corpus-based analyses revealing significant differences in the distributions of certain DMs between NS and NNS speech. Aijmer found that Swedish learners of English overused *I don't know* in order to signal uncertainty or hesitation, and Fung and Carter showed that learners in Hong Kong underused many markers, such as *right*, *yeah*, *well*, and *you know*, compared to the frequencies found in British NS data.

Only a few researchers have empirically investigated DM use in the speech of Japanese English learners. Hays (1992) described the acquisition of DMs by Japanese college students of various English proficiency levels. His analysis of the spoken data revealed that although the markers *and*, *but*, and *so* were frequently used, *you know* and *well* were rarely uttered by Japanese students learning English. In other words, his results indicated that the Japanese learners had greater difficulties acquiring pragmatic markers such as *you know* and *well*. Likewise, Miura (2011) compared the frequency of DMs used by Japanese learners of English to those of English NSs and found that certain markers such as *well*, *I mean*, *kind of*, and *like* were underused by novice and lower level learners. Additionally, Shimada (2011) conducted a corpus-based analysis of English DM use by Japanese learners and NS chil-

dren and adults. The results revealed that as speakers' proficiency improved, they used many items more frequently, regardless of their L1. However, the quantitative analysis confirmed significant differences in the distributions of DMs between Japanese learners and NSs. One of the notable findings was that Japanese learners overused relatively simple types of DMs such as *OK/ okay*, *so*, and *yes.*[2]

Most studies on learners' use of spoken DMs have revealed that learners use certain items much more or less frequently than NSs do. However, the differences in DM frequency between NS and NNS speech are not enough to fully explain the features of DM use in interlanguage—that is, researchers have not yet determined whether the differences are due to the specific influences of individual L1 backgrounds or whether they are common to language learners in general. In order to address the issue, as Granger (2002) argued, it is necessary to construct a comparison of learner languages that incorporate speakers of different L1 backgrounds.

In addition, many comparative studies are based on disparate databases. For example, Shimada (2011) compared three spoken corpora, but there were considerable differences in the ways the data were collected. In that study, the Japanese learner corpus comprised a collection of interviews from a speaking test, but the speech data of NS children and adults were extracted from naturally occurring conversations in daily situations. These different situations may affect how speakers use DMs to facilitate communication, and different types of data collection may generate different results.

Research Questions

In the present study, features of DM use in the speech of Japanese learners of English were explored . The following research questions were addressed using the methods of CIA:

RQ1: How do levels of use of spoken English DMs by Japanese learners differ from those of NSs of English?

RQ2: How do levels of use of spoken English DMs by Japanese learners differ from those of other English language learners with different L1 backgrounds?

RQ1 is intended to replicate previous studies but using homogeneous databases. RQ2, on the other hand, is designed to explore the features of Japanese learners' DM use by comparing interlanguages of different L1 backgrounds.

Method

Databases

In order to make comparisons based on the CIA approach, the present study used two corpus databases. Data for EFL learners were from the Louvain International Database of Spoken English Interlanguage (LINDSEI; Gilquin, De Cock, & Granger, 2010), and data for native English speakers were from the NICT JLE Corpus (Izumi, Uchimoto, & Isahara, 2004).

The former database, LINDSEI, is a spoken corpus consisting of interviews produced by university undergraduates with different L1 backgrounds. All are higher intermediate and advanced learners of English. The spoken corpus consists of 11 subcorpora, classified according to learners' L1, and the data collection was performed using the same procedure for all subcorpora. Each interview lasts about 15 minutes and contains three tasks: (a) warm-up questions on a set topic (e.g., the most impressive country they have visited, their favourite film or play), (b) free and informal discussion with the interviewer, and (c) a picture description. The present study drew on six of the subcorpora, which are characterised in Table 2 below.

Table 2. Number of Interviews and Words per Subcorpus

L1 subcorpus	Language family	*n* of interviews	*n* of words
Japanese (JP)	Asian	51	37,126
Chinese (CH)	Asian	53	63,542
Dutch (DU)	Germanic	50	79,652
German (GE)	Germanic	50	85,950
French (FR)	Romance	50	91,402
Spanish (SP)	Romance	50	64,804
Totals		304	422,476

Note. Adapted from *LINDSEI: Louvain international database of spoken English interlanguage* by G. Gilquin, S. De Cock, and S. Granger (Eds.), 2010, p. 25. Louvain-la-Neuve, Belgium: Presses universitaires de Louvain.

Each subcorpus is made up of about 50 interviews, but the number of words in the Japanese subcorpus is much lower than that in the other subcorpora.[3]

NS data from the NICT JLE Corpus consisted of 20 interviews (94,845 words) produced by American speakers aged 20-24. Each interview lasts about 15 minutes. The interview tasks are also similar to those of LINDSEI, comprising warm-up questions, a single picture description task, and a role-play with the interviewer. The aim of the present study, therefore, is to address gaps in earlier work, ensuring the homogeneity of databases in order to permit an effective comparison of NS and NNS speech.

Procedure

The present study was focused on the 57 DMs listed in Fung and Carter's (2007) functional paradigm, which embraces the features of DMs in spoken English. In the first procedure, the corpus analysis software WordSmith Tools 5.0 (Scott, 2008) was used to obtain frequencies for each of the 57 items. Concordance lines were also viewed to differentiate words used as DMs from those playing other grammatical roles. Some examples are as follows:

Words used as DMs:

> They are advertising by the week, *so* I found it. (The NICT JLE Corpus, N_file00006.stt)

> . . . *well* first of all it's her expression she's got this really sour expression. (LINDSEI-GE050)

Words not used as DMs:

> . . . I . . . wouldn't be able to come back *so* early. (LINDSEI-FR006)

> . . . but now I cannot speak English very *well*. (LINDSEI-JP051)

The categorization was carried out by the author. In order to test the reliability of the coding, a post-hoc intra-coder reliability check was conducted based on Müller (2004) at an interval of about 2 years. Despite the long interval, the simple agreement rate of the coding of *like*, *so*, and *well* was 94%, 99%, and 98%, respectively. Thus, the reliability of the coding process is considered high.

Statistical analyses of the frequencies of DMs were conducted to answer RQ1 and RQ2. The raw frequency of each item was standardized as a frequency per 10,000 words, and then used to calculate the log-likelihood ratio[4] and chi-square value for comparison between corpora of different sizes. In corpus studies, although chi-square tests have often been performed to

compare word frequencies across corpora, log-likelihood tests are considered to have higher reliability than other statistical methods when comparing different-sized datasets (Rayson & Garside, 2000). When researchers compare two datasets with a single degree of freedom, significance is statistically tested by the log-likelihood ratios. If the log-likelihood ratio is ±3.84 or more, a significant difference exists between the two datasets at a 5% significance level (Rayson, Berridge, & Francis, 2004). Additionally, Mann-Whitney tests were employed to compare the frequency of DMs by each functional category, following Fung and Carter (2007).

In addition to these quantitative analyses, the study included qualitative observations about the context, situation, and discourse function of spoken DMs. These observations serve to complement the quantitative analyses, providing vital details on the functions of DM use in actual learner speech.

Results and Discussion

Comparisons of DM Use Between Japanese EFL Learners and NSs of English

In order to answer RQ1, a comparative analysis was conducted using the frequency of DMs in two subsets of speech data: the Japanese subcorpus of LINDSEI (i.e., LINDSEI-JP) and the NS subcorpus of the NICT JLE Corpus (i.e., NICT-NS). Table 3 provides the standardized frequency of each marker, the log-likelihood ratios, and chi-squared values. If the occurrence rate of DMs was 0.01% or below in either database, the items were not included in the analysis.

Chi-square tests revealed that significant differences existed between the two databases in the frequencies of 21 out of 27 DMs with an occurrence rate of more than 0.01%. Additionally, log-likelihood ratios were added to the results obtained with the chi-square tests. If the ratio applied to the two databases was +3.84 or more, the item was considered to be used more frequently in LINDSEI-JP than in NICT-NS. On the other hand, when the ratio was -3.84 or less, the item was considered to be used less frequently in the Japanese learner data. The tests revealed that Japanese learners more frequently used relatively simple markers such as *yes*, *so*, and *I think*, while they used some interpersonal or cognitive markers such as *like*, *really*, *you know*, *kind of*, and *I mean* less frequently than NSs of English. Moreover, Mann-Whitney tests showed that significant differences existed between the two databases in the frequency of DMs in the interpersonal category (U = 110, p = .040). Therefore, the results support those of previous studies (e.g., Hays,

1992; Miura, 2011; Shimada, 2011), in finding that there was a significant discrepancy between Japanese learners and NSs of English in the frequency of DMs.

Table 3. Comparisons of DM Use Between Japanese EFL Learners (LINDSEI-JP) and NSs of English (NICT-NS)

DM	Category	Frequency per 10,000 words		LLR	Chi-square value
		LINDSEI-JP	NICT-NS		
yes	IP	71.92	14.55	248.791	287.012**
so	Ref/Str	206.86	133.38	88.213	95.000**
I think	IP/Cog	88.35	51.66	54.020	58.292**
but	Ref	145.72	101.22	44.215	46.994**
now	Str	13.47	3.58	35.907	40.969**
first	Str	2.96	0.11	21.678	23.961**
finally	Str	2.96	0.74	8.470	9.684**
yeah	IP/Str	86.46	72.54	6.599	6.817**
and	Ref/Str/Cog	420.46	398.02	3.297	3.464
because/'cause	Ref	47.68	46.29	0.109	0.111
I see	IP/Cog	1.08	1.48	-0.326	0.311
or	Ref	50.10	54.09	-0.811	0.806
exactly	IP	2.15	3.48	-1.622	1.507
anyway	Ref	1.08	2.32	-2.356	2.090
basically	IP	0.27	4.32	-20.173	13.780**
oh	IP	7.54	21.30	-34.107	29.021**
then	Str	15.35	38.91	-53.065	46.000**
right/alright	IP/Str	0.27	11.07	-60.590	38.787**
OK/okay	IP/Str	22.90	59.25	-83.548	72.304**
actually	IP	4.85	27.94	-86.724	66.491**
I mean	Cog	2.15	25.73	-110.554	77.784**
well	IP/Str/Cog	5.39	37.32	-128.558	96.303**
kind of	IP	5.39	41.12	-148.569	110.000**
just	IP	10.77	77.39	-271.486	203.074**
you know	IP/Cog	4.31	64.32	-294.673	203.503**
really	IP	8.62	78.13	-304.263	221.379**
like	IP/Cog	28.82	140.65	-390.444	308.967**

Note. The occurrence rate of the markers *cos, great, next, obviously, sort of, sure,* and *what about* was 0% in either corpus. They were excluded from this analysis due to the impossibility of computing the log-likelihood ratio (LLR). Further research should be done to investigate why a certain DM occurs in one dataset but not in the other.

IP = interpersonal; Ref = referential; Str = structural; Cog = cognitive.
**$p < .01$.

Comparisons of DM Use Between Japanese EFL Learners and Other English Learners

This section addresses RC2, which was about comparing DM frequencies in NNS speech from the Japanese subcorpus with the five other subcorpora of LINDSEI (i.e., LINDSEI-OTHERS). Table 4 shows comparisons of the frequency of DMs. As in the analysis of the previous section, if the occurrence rate of a given DM was 0.01% or below in either database, the item was not included in the analysis.

The results of chi-square tests revealed that although Japanese learners often used some items such as *so* and *but*, they also used 14 out of 27 DMs less frequently than other nonnative English learners did. These findings were supported by tests of log-likelihood ratios.[5] Although Mann-Whitney tests did not show significant differences in the frequencies of DMs according to functional category, interpersonal or cognitive function markers such as *well*, *really*, *you know*, *I mean*, and *just* were used less frequently by Japanese learners than by other English learners. Thus, the significant differences in the frequencies of DMs may represent the features of Japanese learners' DM use.

On the other hand, the results given in Table 4 reveal no significant differences between the two databases in the frequency of seven items: *exactly*, *kind of*, *or*, *OK/okay*, *anyway*, *cos*, and *basically*. There were only small differences between learners' respective frequencies of three markers—*and*, *yes*, and *right/alright*—although the differences were significant at a 5% significance level. In short, it was notable that Japanese learners used some items just as frequently as other nonnative English learners. Among these items, the use of *kind of*, *OK/okay*, *basically*, *yes*, and *right/alright* may be regarded as features of DM use in NNSs' interlanguage because the frequency of the five items differed significantly between Japanese learners and NSs of English (see Table 3).

Table 4. Comparisons of DM Use Between Japanese EFL Learners (LINDSEI-JP) and Other Nonnative English Learners (LINDSEI-OTHERS)

DM	Category	Frequency per 10,000 words		LLR	Chi-square value
		LINDSEI -JP	LINDSEI -OTHERS		
so	Ref/Str	206.86	96.04	315.280	397.358**
but	Ref	145.72	119.45	18.157	19.430**
now	Str	13.47	8.15	9.638	11.130**
finally	Str	2.96	1.09	7.093	9.470**
first	Str	2.96	1.17	6.292	8.234**
and	Ref/Str/Cog	420.46	394.14	5.815	6.164*
OK/okay	IP/Str	22.90	19.05	2.456	2.591
kind of	IP	5.39	4.88	0.173	0.178
exactly	IP	2.15	2.13	0.001	0.001
or	Ref	50.10	55.35	-1.749	1.711
anyway	Ref	1.08	2.36	-3.025	2.484
cos	Ref	4.31	6.90	-3.859	3.414
basically	IP	0.27	1.32	-4.363	3.057
yes	IP	71.92	84.57	-6.790	6.553*
right/alright	IP/Str	0.27	2.15	-9.283	6.050*
I think	IP/Cog	88.35	109.15	-14.451	13.799**
yeah	IP/Str	86.46	111.48	-20.767	19.613**
like	IP/Cog	28.82	44.56	-21.778	19.507**
actually	IP	4.85	14.07	-28.085	21.731**
oh	IP	7.54	18.42	-28.653	23.000**
because/'cause	Ref	47.68	73.26	-34.943	31.434**
then	Str	15.35	33.61	-42.937	35.367**
just	IP	10.77	47.72	-145.738	104.410**
I mean	Cog	2.15	31.30	-164.463	100.366**
you know	IP/Cog	4.31	39.91	-182.483	117.121**
really	IP	8.62	57.53	-227.775	153.006**
well	IP/Str/Cog	5.39	70.01	-357.270	221.268**

Note. The occurrence rate of the markers *sort of* and *that is* was 0% in either corpus. They were excluded from this analysis due to the impossibility of computing the log-likelihood ratio (LLR). Further research should be done to investigate why a certain DM occurs in one dataset but not in the other.

IP = interpersonal; Ref = referential; Str = structural; Cog = cognitive.
*p < .05. **p < .01.

However, these data do not address differences in DM use within the category LINDSEI-OTHERS, and distributions within individual subcorpora could boost or lower the overall frequency. To provide a clear picture, the frequencies of 12 DMs mentioned in this section were also compared across the six subcorpora of NNS speech. The further comparison was made to

confirm whether the use of *so*, *but*, *well*, *really*, *you know*, *I mean*, and *just* exhibited the features of Japanese learners' speech, and whether the use of *yes*, *kind of*, *right/alright*, *basically*, and *OK/okay* reflected the features of DM use in NNSs' interlanguage.

Figure 1 shows the frequency of *so* and *but* in each subcorpus. Although *so* was used in the Japanese subcorpus substantially more frequently than in any other nonnative subcorpus, only small differences existed among subcorpora in the frequency of *but*. Thus, the results confirm that the marker *so* is used more frequently by Japanese learners, and that the lower usage levels of *but* in the Chinese and German subcorpora lower the overall frequency of LINDSEI-OTHERS.
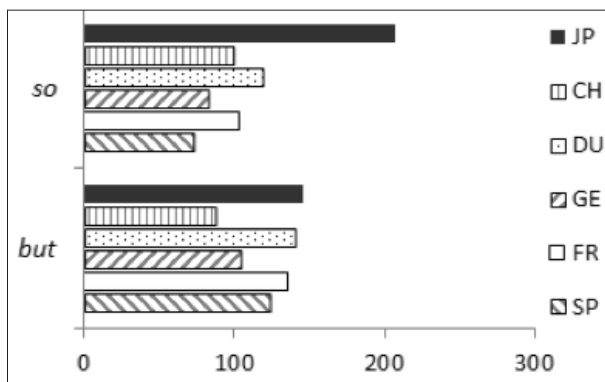
Figure 1. Frequency of *so* and *but* per 10,000 words in each subcorpus of LINDSEI.

Figure 2 shows a comparison of the frequency of *well* and *really* in each subcorpus. The analysis revealed that both Japanese and Chinese learners of English used the two markers notably less frequently than other nonnative English learners. In other words, the results suggest that English learners whose L1 belongs to an East Asian language family may be more likely to use the markers *well* and *really* much less frequently.
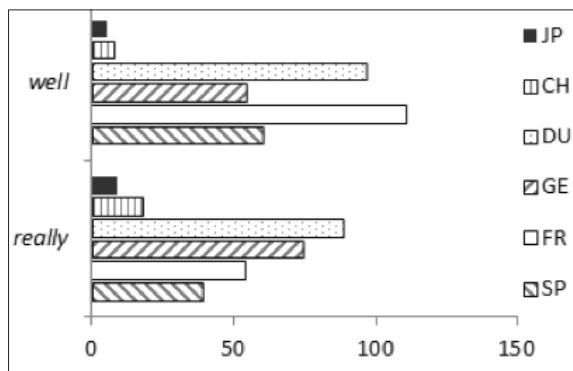
Figure 2. Frequency of *well* and *really* per 10,000 words in each subcorpus of LINDSEI.

Figure 3 shows the frequency of *you know*, *I mean*, and *just* in each subcorpus. The analysis revealed that Japanese learners used the three markers less frequently than other nonnative English learners. In other words, the results display a marked tendency for Japanese learners to use the interpersonal or cognitive function markers less often. These distinguishing features can be found only among Japanese learners of English; that is, they are not shared by nonnative English learners with different L1 backgrounds.
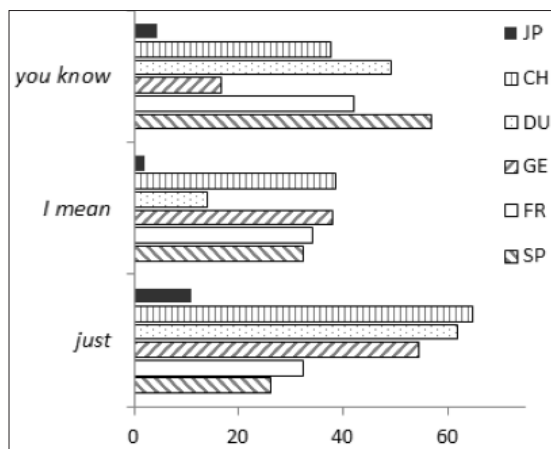

Figure 3. Frequency of *you know*, *I mean*, and *just* per 10,000 words in each subcorpus of LINDSEI.

Figure 4 shows the frequencies of *yes*, *kind of*, *right/alright*, *basically,* and *OK/okay* in each subcorpus. The marker *yes* generally displays small differences among the subcorpora except for in the French subcorpus, where it was quite frequent indeed. On the other hand, the three markers *kind of*, *right/alright*, and *basically* were infrequently used in all six subcorpora. The general frequent use of *yes* and the low frequencies of *kind of*, *right/alright*, and *basically* may be common to learners of English. With regard to the frequencies of *OK/okay*, Figure 4 shows that there is a considerable variability among the subcorpora.
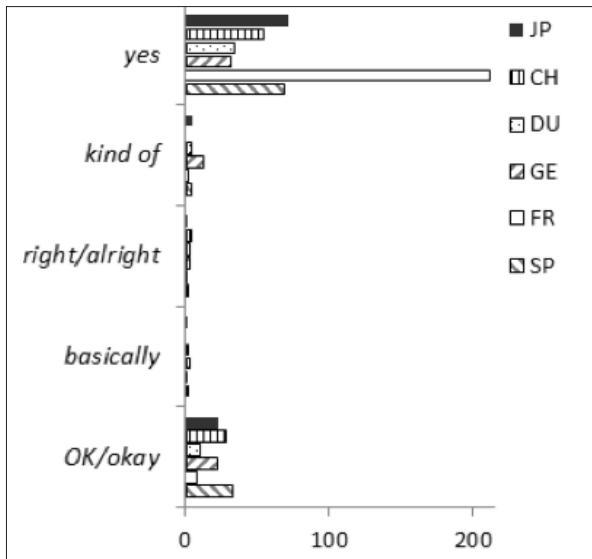
**Figure 4. Frequency of *yes*, *kind of*, *right/alright, basically,* and *OK/
okay* per 10,000 words in each subcorpus of LINDSEI.**

In short, although simple items such as *yes* may be preferred by NNSs, items such as *kind of*, *right/alright*, and *basically* may be more difficult for them to acquire.

Why Do Japanese EFL Learners Overuse the Marker So?

Previous studies such as Hays (1992), Miura (2011), and Shimada (2011) have suggested that Japanese learners may infrequently use certain pragmatic markers such as *well*, *I mean*, and *you know*, but they may frequently

use simple types of markers such as *so* and *yes*. The present study yielded similar findings and distinguished features particular to Japanese learners from those seen in the speech of other NNSs. To investigate the acquisition of DMs in Japanese learners' speech, however, it is important to explore why some items are more or less frequently used. To that end, this section is focused on the marker *so*, which is frequently used by Japanese learners.

According to Fung and Carter's (2007) framework, the marker *so* has two discourse functions, referential and structural. Although the referential marker *so* serves a syntactic function to signal a relationship between one discourse segment and another, the structural marker *so* has some pragmatic functions, such as as a signal of summarising opinions and topic shifts. In the present study, as in my earlier study (Shimada, 2012), tokens of *so* were classified by functional category: referential, structural, or other. The following are illustrative examples of *so* extracted from the speech data of LINDSEI:

4. Referential: I don't think I pronounce it very well, *so* I am a bit embarrassed . . . (LINDSEI-SP015)
5. Structural: . . . I think that's Julia Roberts. *So* that's all. (LINDSEI-CH019)
6. Structural: *So* what do you think of the city Guangzhou? (LINDSEI-CH045)
7. Other: . . . I always use bus *so* untto[6] . . . my nearest station is Ujiie Station. (LINDSEI-JP005)

In example 4, the speaker uses the referential marker *so* in order to establish a cause-and-effect link between the first clause and the second one. In example 5, the speaker tries to mark the conclusion of the topic by using the structural marker *so*. The speaker in example 6 changes the topic to the listener's impression of the city Guangzhou by using the structural marker *so*. In example 7, however, the marker *so* is neither referential nor functional; instead, it seems to be used as a filler, which can provide time for the speaker to think about what to say next.

Figure 5 shows the percentages for the three types of *so* (referential, structural, other) in the randomly sampled speech data, which comprise 10 interviews from each subcorpus. The coding of the functional categories was carried out by the author. As in the categorization of DMs described above, a post-hoc intra-coded check was conducted for the three subcorpora, LINDSEI-JP, -CH, and -DU (i.e., 30 interviews) at an interval of about

2 years. The overall agreement rate was 93%. Thus, the reliability of this analysis is considered high.
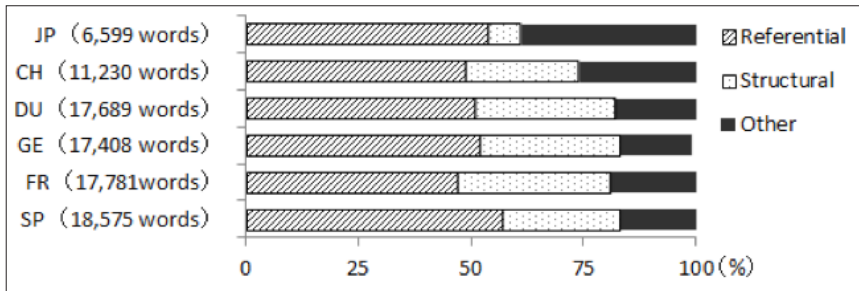
Figure 5. Percentages for the three types of *so* in randomly sampled speech data (10 interviews from each subcorpus of LINDSEI).

The results given in Figure 5 reveal that the proportion of the structural marker *so* was very low in the Japanese subcorpus. The third class of *so*, which is neither referential nor structural in function (i.e., other) was used more frequently by Japanese English speakers than by any other subcorpus group. The use of *so* as a filler may boost the frequency of the marker in Japanese English learners' speech.[7]

Conclusion

CIA was employed in this study to investigate the use of DMs in the speech data of Japanese learners of English. The results illuminate some features of these speakers' DM use.

This study's first research question was about frequencies of DMs in the speech of Japanese learners in comparison with those of NSs of English. Frequency analysis revealed significant differences between Japanese learners and NSs of English in the frequency of many DMs. Japanese learners frequently used some simple markers such as *yes*, *so*, and *I think*, yet they infrequently used certain interpersonal or cognitive function markers such as *like*, *really*, *you know*, *kind of*, and *I mean*. These findings corroborate those of previous studies, and they indicate that Japanese learners may have more difficulty acquiring particular pragmatic markers. These findings have important implications for language instructors, who may improve their students' interactional L2 skills as well as their linguistic ones through instructional focus on DMs.

The second research question was about levels of English DM use by Japanese learners in comparison with those of English learners with different L1 backgrounds. Frequency analyses revealed both similarities and differences between Japanese learners and other nonnative English learners in their use of DMs. Although Japanese learners used *so* much more frequently than other nonnative learners, they also used certain interpersonal or cognitive function markers such as *you know*, *I mean*, and *just* much less frequently. In other words, certain features of their DM use are distinguishable from those of nonnative English learners generally. This suggests the need for language instructors and materials writers to carefully provide Japanese learners with language input according to the characteristics of their interlanguage. For example, language instructors and materials writers should provide infrequent and difficult items, such as interpersonal or cognitive markers, at an intermediate or advanced proficiency level. Additionally, they should furnish Japanese learners with opportunities to use as many kinds of easy-to-use items as possible at a lower level.

This study has two basic limitations. Qualitative observations indicated that Japanese learners might use *so* as a filler, but this analysis has been far from exhaustive; more work on qualitative patterning is thus needed. As Romero-Trillo (2002) and Müller (2004) have suggested, Japanese learners' more or less frequent use of DMs may be a result of the influence of their L1. Second, some tasks to elicit speech may have an effect on learners' DM use. For example, a picture description task may not lend itself to the use of interpersonal markers such as *really* and *just*. Further research is needed to analyse learners' speech from a qualitative perspective and to investigate why Japanese learners may display different tendencies in English DM use from other nonnative English learners.

Notes

1. An earlier version of this paper was presented at the 127th Kanto Chapter Conference of the Japan Association for Language Education and Technology, Tokyo, Japan, 12 November 2011.
2. According to the online English Vocabulary Profile (http://www.englishprofile.org/), the markers *OK/okay*, *so*, and *yes* are classified into the Common European Framework (CEFR) level A1 or A2. Therefore, these markers can be regarded as easy items for English learners.
3. As Pritchard (1995) points out, Japanese learners of English may prefer slow, careful speech and take a long pause before answering a ques-

tion. If so, the interaction style may have a negative effect on fluency in speech production. However, LINSDEI does not contain audio data and does not provide the information necessary to find out why the Japanese students produced a much smaller number of words than any of the other nonnative English learners.

4. The tests of the log-likelihood ratios are also called G-tests.
5. The author combined the five subcorpora into one group and ran log-likelihood tests to compare the frequency of DMs between LINDSEI-JP and LINDSEI-OTHERS.
6. The Japanese word *untto* is approximately equivalent to the English marker *well*.
7. In the Japanese subcorpus, *so* as a filler was ubiquitous, although the frequency was not fully examined. Shimada (2012) also pointed out that the filler usage may contribute to Japanese learners' overuse of the marker. The present study confirms those earlier findings.

Acknowledgments

Kazunari Shimada is an assistant professor at Takasaki University of Health and Welfare. His research interests include second language acquisition, materials development, and corpus linguistics.

References

Ädel, A. (2006). *Metadiscourse in L1 and L2 English.* Amsterdam: John Benjamins.

Aijmer, K. (2004). Pragmatic markers in spoken interlanguage. *Nordic Journal of English Studies*, *3*(1), 173-190.

Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge: Cambridge University Press.

Carter, R., & McCarthy, M. (2006). *Cambridge grammar of English: A comprehensive guide to spoken and written English grammar and usage*. Cambridge: Cambridge University Press.

Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics*, *31*, 931-952. http://dx.doi.org/10.1016/S0378-2166(98)00101-5

Fraser, B. (2009). An account of discourse markers. *International Review of Pragmatics*, *1*, 293-320. http://dx.doi.org/10.1163/187730909X12538045489818

Fung, L., & Carter, R. (2007). Discourse markers and spoken English: Native and learner use in pedagogic settings. *Applied Linguistics*, *28*, 410–439. http://dx.doi.org/10.1093/applin/amm030

Gilquin, G., De Cock, S., & Granger, S. (Eds.). (2010). *LINDSEI: Louvain international database of spoken English interlanguage.* Louvain-la-Neuve, Belgium: Presses universitaires de Louvain.

Granger, S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), *Languages in contrast* (pp. 37-51). Lund, Sweden: Lund University Press.

Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition, and foreign language teaching* (pp. 3-33). Amsterdam: John Benjamins.

Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In K. Ajimer (Ed.), *Corpora and language teaching* (pp. 13-32). Amsterdam: John Benjamins.

Granger, S., & Tyson, S. (1996). Connector usage in the English essay writing of native and nonnative EFL speakers of English. *World Englishes*, *15*, 17–27. http://dx.doi.org/10.1111/j.1467-971X.1996.tb00089.x

Hays, P. R. (1992). Discourse markers and L2 acquisition. *Papers in Applied Linguistics-Michigan*, *7*, 24-34.

Izumi, E., Uchimoto, K., & Isahara, H. (Eds.). (2004). *Nihonjin 1200 nin no eigo supiikingu koopasu* [A spoken corpus of 1200 Japanese learners of English]. Tokyo: ALC Press.

Lenk, U. (1998). Discourse markers and global coherence in conversation. *Journal of Pragmatics*, *30*, 245-257. http://dx.doi.org/10.1016/S0378-2166(98)00027-7

McCarthy, M., & Handford, M. (2004). "Invisible to us": A preliminary corpus-based study of spoken business English. In U. Connor & T. A. Upton (Eds.), *Discourse in the professions: Perspectives from corpus linguistics* (pp. 167-201). Amsterdam: John Benjamins.

Miura, A. (2011, September). *Discourse markers in spoken corpora of Japanese EFL learners.* Paper presented at Learner Corpus Research 2011, Louvain-la-Neuve, Belgium.

Müller, S. (2004). 'Well you know that type of person': Functions of *well* in the speech of American and German students. *Journal of Pragmatics*, *36*, 1157-1182. http://dx.doi.org/10.1016/j.pragma.2004.01.008

Pritchard, R. M. O. (1995). *Amae* and the Japanese learner of English: An action research study. *Language, Culture and Curriculum*, *8*, 249-264. http://dx.doi.org/10.1080/07908319509525207

Rayson, P., Berridge, D., & Francis, B. (2004). Extending the Cochran rule for the comparison of word frequencies between corpora. In G. Purnelle, C. Fairon, & A. Dister (Eds.), *Le poids des mots: Proceeding of the 7th International Conference on Statistical Analysis of Textual Data* (pp. 926-936). Louvain-la-Neuve, Belgium: Presses universitaires de Louvain.

Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. In A. Kilgarriff & T. B. Sardinha (Eds.), *Proceedings of the Workshop on Comparing Corpora: Held in Conjunction with the 38th Annual Meeting of the Association for Computational Linguistics* (pp. 1-6). Hong Kong: Hong Kong University of Science and Technology.

Romero-Trillo, J. (2002). The pragmatic fossilization of discourse markers in nonnative speakers of English. *Journal of Pragmatics*, *34*, 769-784. http://dx.doi.org/10.1016/S0378-2166(02)00022-X

Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.

Scott, M. (2008). WordSmith Tools (Version 5) [Computer software]. Liverpool, UK: Lexical Analysis Software.

Shimada, K. (2011). Discourse marker use in learners and native speakers: A corpus-based analysis of spoken English. *Annual Review of English Language Education in Japan (ARELE)*, *22*, 377-392.

Shimada, K. (2012). Discourse markers in EFL textbooks and spoken corpora: Materials design and authenticity. *Language Education & Technology*, *49*, 215-244.

Svartvik, J. (1980). 'Well' in conversation. In S. Greenbaum, G. Leech, & J. Svartvik (Eds.), *Studies in English linguistics for Randolph Quirk* (pp. 167-177). London: Longman.

Tyler, A. E., Jefferies, A. A., & Davies, C. E. (1988). The effect of discourse structuring devices on listener perceptions of coherence in non-native university teacher's spoken discourse. *World Englishes*, *7*, 101-110. http://dx.doi.org/10.1111/j.1467-971X.1988.tb00223.x

Williams, J. (1992). Planning, discourse marking, and the comprehensibility of international teaching assistants. *TESOL Quarterly*, *26*, 693-711. http://dx.doi.org/10.2307/3586869

An Exploratory Reliability and Content Analysis of the CEFR-Japan's A-Level Can-Do Statements

Judith Runnels
Hiroshima Bunkyo Women's University

Both the Common European Framework of Reference (CEFR) and the CEFR-Japan (CEFR-J), an alternate version designed for Japanese learners of English, provide measurements of language proficiency via assessment or self-assessment on scales of descriptors of communicative competences (known as can-do statements). Although extensive empirical evidence supports these claims for the CEFR, the same cannot yet be said of the CEFR-J. Mokken scaling was thus used to measure the reliability of can-do statement scales from the five skills of the CEFR-J's five A sublevels of A1.1, A1.2, A1.3, A2.1, and A2.2. Statements that negatively affected the reliability of the scale were analysed. Lower reliability was attributed to characteristics specific to participants (homogeneity of the population, familiarity with the task, and if the material was recently studied), and content of the statement itself (whether it implied more than one language skill or none at all, whether it contained a contradiction, or was confusing or unfamiliar). Modifications to increase the reliability of can-do statement scales and limitations of using illustrative descriptor-based systems as measurement instruments are discussed.

　ヨーロッパ共通言語参照枠（CEFR）とその日本版CEFR-Jはともにコミュニケーション能力を示す指標（can-do statements）であり、評価あるいは自己評価による言語運用能力の測定を目的としている。CEFRにはその主張を裏付ける根拠が豊富にある一方で、CEFR-Jには未だ十分な裏付けがあるとは言い難い。そこで本研究は、CEFR-Jの5つのA sublevelについて5技能に関わるcan-do statementの信頼性をMokken スケールを用いて測定した。信頼性に否定的な影響を与えた指標をさらに分析したところ、信頼性の低さは学習者特有の特徴（母集団の均一性、課題に対する慣れ、最近学習された項目か否か）と指標そのもの（2つ以上の言語技能に関係している、言語技能に関係していない、矛盾がある、あいまいでわかりにくい）に起因するものであっ

た。Can-do　statementsの信頼性を高めるための修正に関する提案と、ディスクリプタを用いた
指標を使用することの限界についての考察を行った。

T heoretical work, case studies, and other evidence suggest that the Common European Framework of Reference (CEFR) provides an effective descriptive scheme for analysing the needs, goals, materials, and achievements of language learners (Alanen, Huhta, & Tarnanen, 2010; Council of Europe, 2001). It employs illustrative descriptors, known as can-do statements, of communicative competences for five skills (listening, reading, spoken interaction, spoken production, and writing). All can-do statements are divided into six proficiency levels of increasing difficulty (A1, A2, B1, B2, C1, C2). To provide an example, can-do statements 1 and 2 are from reading levels B1 and A1 respectively:

1. *I can identify the main conclusions in clearly written argumentative texts.*
2. *I can understand the general idea of simple informational texts and short simple descriptions, especially if they contain pictures which help to explain the text.* (Council of Europe, 2001)

The CEFR's descriptors were developed through qualitative and quantitative methods to ensure progressions in difficulty as a learner advances through the levels (North, 2000; 2002). This difficulty hierarchy has been continually validated in a European context since the CEFR's publication (Figueras, 2012). Although the CEFR is argued to be an "international standard for language teaching and learning" (North, Ortega, & Sheehan, 2010, p. 6), it is also frequently criticized for its theoretical underpinnings, particularly regarding how it should be used to measure proficiency. Because the hierarchy of difficulty represented by the increasing levels is largely based on difficulty judgments from language educators, Fulcher (2004; 2010) argued that it can neither be used to gauge proficiency nor provide any standardized measure of language ability. Other opponents of the CEFR have noted that it cannot and should not act as a language test for measuring ability (Weir, 2005), as ties to SLA theory have yet to be established (Hulstijn, 2007), and, as well, the progression of difficulty inherent in the levels is unsupported by empirical studies of performance samples from language learners (Westhoff, 2007).

Conversely, supporters of the CEFR praise it for how it can be used by autonomous learners to provide an estimation of proficiency or direction

for an individual's language study (Glover, 2011). Typically, such a level estimation is achieved with a self-assessment whereby learners read a set of can-do statements then decide if they are capable of performing the communicative actions entailed by each statement (Glover, 2011; Little, 2006). Level estimations are thus based on the learners' perceptions of their own achievement on the can-do statements. Future self-assessments can be compared to previous ones as a measure of progress.

Due to its success in Europe (North, et al., 2010) and other regions of the world (Figueras, 2012; Wang, Kuo, Tsai, & Liao, 2012), the CEFR has been modified into alternate versions tailored to meet local demands. One such example is the CEFR-Japan (CEFR-J), introduced to address the lack of consistently used measures for progress or proficiency among Japanese institutions.

Developing the CEFR-J

When Negishi (2012) found that over 80% of Japanese English learners fall within A1 and A2 levels, he concluded that the CEFR's can-do statements were not providing users with adequate criteria for distinguishing between the population's language abilities. He highlighted the need for a system tailored to the needs of Japanese English language learners and development of the CEFR-J thus began (Negishi, Takada, & Tono, 2013; Tono & Negishi, 2012). As part of the first stage of development, can-do statements from DIALANG (Council of Europe, 2001, p. 231-234; Huhta, Luoma, Oscarson, Sajavaara, Takala, & Teasdale, 2002) were administered to 360 Japanese university students to ascertain that the rank ordering of difficulty by Japanese students matched that of the CEFR (Negishi et al., 2013). The participants generally ordered the can-do statements accordingly and it was concluded that overall, the CEFR would be suitable for use by Japanese English learners. Nonetheless, there were some outlying can-do statements that were being rated by the Japanese population as more difficult than predicted. Negishi's (2011) analysis of an outlying A1 reading descriptor is as follows:

> *I can understand short, simple messages, e.g., on postcards* turned out to be more difficult than the A2.1 descriptor *I can understand short, simple texts containing the most common words, including some shared international words*. This might be because Japanese postcards tend to contain much more information than their European counterparts, and therefore the Japanese EFL learners considered it to be more difficult than it was originally assumed in the CEFR. (p. 108)

Negishi (2011) concluded that tasks "were judged to be more difficult than the levels they were originally assigned to [if learners had only had limited experience with them], whereas the tasks they had experienced were judged to be easier" (p. 108). Any can-do statement that was not scaling according to the CEFR was thus adjusted with real-life examples specific to a Japanese context and then retested. Following modifications, the initially outlying can-do statements ordered consistently with the CEFR's predictions, thus demonstrating that the contextualization or localization process had been successful (Negishi et al., 2013).

In addition to the contextualization of descriptors, the CEFR's A and B levels were modified in order to better distinguish between learners (Negishi, 2011; Tono & Negishi, 2012). The CEFR's four original levels (A1, A2, B1, B2) were subdivided into nine categories (A1.1, A1.2, A1.3, A2.1, A2.2, B1.1, B1.2, B2.1, B2.2) and a Pre-A1 level was also created, resulting in 12 CEFR-J levels in total. Since its publication in March 2012 (TUFS Tonolab, 2012), the CEFR-J has been promoted as a way forward for any language education program in Japan and numerous projects for developing textbooks and learner and teacher autonomy support tools are under way (Imig, 2013).

Reliability Issues

Despite interest in the implementation of the system, only a limited amount of research specific to the CEFR-J has been undertaken. For the CEFR, extensive work has demonstrated the reliability of a scale of increasing difficulty, which in turn supports any arguments regarding standardized assessment of language proficiency (Little, 2006; North, 2007; North & Schneider, 1998). For the CEFR-J however, very few studies beyond the CEFR-J's development process (Negishi, 2011; Negishi et al., 2011; Tono & Negishi, 2012) have been published. Runnels (2013a) for instance, measured the rank ordering of can-do statement difficulty by Japanese university-level English language learners for the CEFR-J's levels A1.1 through A2.2. Rasch analysis (Andrich, 1978) and analyses of variances (ANOVAs) indicated that for several adjacent levels, there were no statistically significant differences in mean difficulty ratings. This was also found to be the case when differences between levels within each individual skill were tested (Runnels, 2013b). Although Tono and Negishi (2012) had concluded that the two original CEFR A levels were not adequately distinguishing among the span of learners' abilities, Runnels (2013a, 2013b) suggested that the five sublevels the A level was divided into were perhaps too many and that the support of the language learning process that the CEFR-J is designed to provide may

be jeopardized if users cannot consistently distinguish between levels of proficiency based on the CEFR-J descriptors.

A major weakness of both of Runnels' (2013a, 2013b) studies, however, is that they were focused solely on the responses to the difficulty of can-do statements from a limited sample of users and did not provide any measure of reliability or variance to account for individual differences across participants. In general, when difficulty ratings alone are analysed, the extent to which the difficulty hierarchy may be different for every learner is not accounted for. Although an issue certainly exists if the CEFR-J's users are not able to confidently estimate language level due to negligible differences in difficulty between adjacent sublevels, there is also an issue if CEFR-J can-do statements and their scales are behaving very differently for each individual that responds to them. For instance, learners who are using an A2.2 can-do statement such as 3 to self-assess listening may conclude that they are able to perform any task entailed by 3. However, the next learner to self-assess using 3 may not come to the same conclusion, deciding instead that he or she can only perform tasks from 4 or 5, both lower order statements than A2.1:

3. *I can understand instructions about procedures (e.g., cooking, handicrafts), with visual aids, provided they are delivered in slow and clear speech involving rephrasing and repetition.*

4. *I can understand short, simple announcements (e.g., on public transport or in stations or airports), provided they are delivered slowly and clearly.*

5. *I can understand the main points of straightforward factual messages (e.g., a school assignment, a travel itinerary), provided speech is clearly articulated in a familiar accent.*

Assuming that controls for rater severity and ability are taken into account, placing the first learner at A2.2 for listening and the second at A2.1 level is supported by empirical demonstrations of common understanding of the difficulty of statements across populations of users. For the CEFR-J, though, these conclusions cannot be drawn with such confidence because no prior researcher has examined whether A2.2 can-do statements such as 3 are indeed rated as more difficult than A2.1 descriptors such as 4 and 5 by the majority of users. Empirical studies demonstrating a consistent and reliable difficulty hierarchy across the levels of the CEFR-J are lacking.

The current study was thus designed to provide preliminary evidence on the reliability of can-do statements within the CEFR-J's difficulty hierarchy and to determine the extent to which participants are behaving consistently

in their responses regarding the difficulty of can-do statements within each skill's scale. Any A-level can-do statement shown to be negatively affecting the reliability of a skill's scale (in that response patterns are found to be less consistent) is analysed, and recommendations for modification in order to potentially increase reliability are discussed.

Method

Participants

Participants consisted of 590 first- and 2nd-year students from a small private women's university in western Japan. Each participant was in one of five majors of study, one of which was English. In order to determine whether the can-do statements were interpreted consistently across a variety of users with a range of language abilities, both 1st- and 2nd-year students from all of these disciplines were included in the analysis. The can-do statement survey described below was administered at the end of the first semester of the academic year, meaning that all non-English majors (536 participants or 90.8% of the total) had completed at least 4 months or 12 months of twice-weekly 90-minute university-level English classes, depending on whether students were in their 1st or 2nd year of study. The English majors had completed one or three semesters of full-time English study depending on whether they were in their 1st or 2nd year.

All participants were unfamiliar with the CEFR-J and had no previous experience using can-do statements. They had also received no training on conducting self-assessment. Participation was voluntary and had no bearing on course grades.

Instrument

The can-do statement survey was administered on www.surveymonkey. com (SurveyMonkey, 2012) during participants' class time and in their regular classrooms. All statements are available online in both English and Japanese from the Tokyo University of Foreign Studies (TUFS Tonolab, 2012). For each of the five skills (listening, reading, spoken production, spoken interaction, and writing) there are two can-do statements for each level, for a total of 50 statements. Participants responded on a 5-category Likert scale from *strongly disagree* to *strongly agree* to all randomly ordered Japanese can-do statements from the CEFR-J's five A sublevels (A1.1, A1.2, A1.3, A2.1 and A2.2), which were selected because the institution's curriculum is targeted at these levels.

Analysis

Multivariate Statistics Inc.'s EQSIRT Version 1.0 (Bentler & Wu, 2012) was used to perform a Mokken Scale analysis to determine the reliability of the can-do statements' scales for each skill.

In testing, Guttman patterning is an ideal hypothetical pattern of item difficulties (Guttman, 1950). If the test forms a theoretically perfect Guttman pattern, all test-takers will reach a point in the question lineup (wherein all questions are lined up in order of increasing difficulty) such that all of the questions have been answered correctly up to that point, but all of the questions afterwards are too difficult and are therefore answered incorrectly. The point at which the change from correct to incorrect occurs depends on the test-takers and is often seen to represent their ability on that test. Mokken scaling is a statistical technique that assumes the order of difficulty of items is not the same across a population (van Schuur, 2003) and it provides a measure of reliability by identifying items for which Guttman patterning is occurring at higher rates (Molenaar, 1997; Sijtsma & Molenaar, 2002).

The CEFR-J's increasingly difficult levels or hierarchy theoretically forms a Guttman scale: A1.1 should be easier than A1.2 which is easier than A1.3 and so forth, such that learners will eventually reach a point beyond which the tasks are too difficult for them to perform, thus representing their CEFR-J level of proficiency. Accordingly, in this theoretically perfect system, learners should find A2.1 listening statements such as 4 and 5 easier than A2.2 statement 3. Realistically, this may not always be the case as some learners will find A2.1 statements more difficult. The response patterns from these learners would thus contradict the intended Guttman patterning of the system. For example, if a particular learner finds an A2.1 item to be extremely difficult and an A2.2 item very easy, while peers of the same ability find the A2.1 item easier, the distribution of difficulty ratings for the A2.1 can-do statement would then skew, with the mean difficulty rating increasing due to the responses from only a few learners even though the majority of respondents of the same ability were behaving similarly to one another.

Mokken scaling detects for these types of response patterns by creating a scale that reflects the difficulties of each statement according to the abilities of respondents, but also the extent to which a greater number of more able respondents found the given statement more difficult (van Schuur, 2003). Its resulting statistic, known as the coefficient of homogeneity (*H* or *H*-value), reflects response structures for each item in terms of item thresholds (Andrich, 1978; Embretson & Reise, 2000) and provides a measure of reliability for each can-do statement, reflecting the extent to which a Guttman pattern

is evident for all responses. Coefficients of homogeneity fall between 0 and 1.0, where a higher *H*-value is associated with an item that is scaling more Guttman-like (Mokken, 1971). Unacceptable *H*-values fall below .3, and anything over .6 is considered strong in terms of reliability (van Schuur, 2003).

For the current analysis, an *H*-value provides an alternate perspective to Cronbach's alpha as a measure of reliability because "the order of 'difficulty' of the items has an important theoretical interpretation that is not taken into consideration in [traditional] reliability analyses" (van Schuur, 2003, p. 141). Although classical reliability analyses assume that all items exhibit the same frequency distributions, when items are expected to form a Guttman scale such as in the CEFR-J difficulty hierarchy, the assumption is the opposite: that items exhibit differing frequency distributions (Carroll, 1945; Ferguson, 1941; van Schuur, 2003). Therefore, "if items in fact form a Guttman scale, or are expected to do so, it makes sense to analyse them with a model that takes Guttman's model assumption of cumulativity into account" (van Schuur, 2003, p. 141).

Results and Discussion

A Mokken Scale analysis, performed to examine the reliability of all A-level can-do statements, revealed that the CEFR-J's A1- and A2-level can-do statements are forming a strongly reliable scale (*H* = .624) according to commonly accepted criteria for *H* (van Schuur, 2003). Cronbach's alpha was found to be .944 across all statements, also indicating that overall, the scales were found to be strongly reliable.

The results of the Mokken Scale analysis for each statement are displayed in Tables 1-5 according to language skill. The *H*-value next to each statement represents the reliability of the scale as a whole. If a given statement is removed, there is either a positive (moving down in the table) or negative (moving from the bottom up) impact on the reliability of the scale. Therefore, statements that are closer to the top of the table are more strongly affecting the reliability of the scale in a negative way. Of particular concern are any statements from higher order CEFR-J levels that are appearing near the top of the table, because they are theoretically more difficult to perform and should therefore appear further down in the table as a result of being rated more difficult by a larger number of respondents. The two statements at the bottom of each scale exhibit the same coefficient of homogeneity because the *H*-value is incomputable for less than three items (i.e., three items are required to constitute a scale). For each skill, the least reliable statements

will be further analysed in terms of how their content may be affecting reliability. Specifically, four descriptors from Listening (L), three from Reading (R), two from Spoken Production (SP) and Writing (W), and one statement from Spoken Interaction (SI) are discussed. Of these, four are from A1.1 and A2.1, three are from A2.2, and one is from A1.3.

Table 1. Mokken Scales for the CEFR-J A-Level Can-Do Statements for Listening

Rf	Level	*H*	Listening can-do statement
(a)	A2.1	.64	I can understand short, simple announcements (e.g., on public transport or in stations or airports) provided they are delivered slowly and clearly.
(b)	A2.2	.66	I can understand and follow a series of instructions for sports, cooking, etc. provided they are delivered slowly and clearly.
(c)	A1.3	.67	I can understand instructions and explanations necessary for simple transactions (e.g., shopping and eating out), provided they are delivered slowly and clearly.
	A1.1	.68	I can understand short, simple instructions such as "Stand up." "Sit down." "Stop." etc., provided they are delivered face-to face, slowly and clearly.
	A1.1	.69	I can catch key information necessary for everyday life such as numbers, prices, dates, days of the week, provided they are delivered slowly and clearly.
	A2.1	.70	I can understand the main points of straightforward factual messages (e.g., a school assignment, a travel itinerary), provided speech is clearly articulated in a familiar accent.
	A1.2	.71	I can understand short conversations about familiar topics (e.g., hobbies, sports, club activities), provided they are delivered in slow and clear speech.
	A1.2	.72	I can catch concrete information (e.g., places and times) on familiar topics encountered in everyday life, provided it is delivered in slow and clear speech.

Rf	Level	*H*	Listening can-do statement
(d)	A2.2	.74	I can understand instructions about procedures (e.g., cooking, handicrafts), with visual aids, provided they are delivered in slow and clear speech involving re-phrasing and repetition.
	A1.3	.74	I can understand phrases and expressions related to matters of immediate relevance to me or my family, school, neighborhood etc., provided they are delivered slowly and clearly.

Note. H-values represent the reliability of the scale as a whole. Statements closer to the top more strongly affect the reliability of the scale in a negative way.

The two least reliable listening items in Table 1, references (a) and (b), are both A2-level statements. The less reliable responses to (a) may be attributable to participants' lack of experience with English-language announcements as was found in Negishi (2011), whereby familiar tasks were judged as linguistically easier to complete than nonfamiliar tasks.

Regarding the comprehension of short public announcements, however, given that many in train stations in Japan announcements are made bilingually (in Japanese and English), participants may not normally rely on the English announcement to obtain information they need as the first part of the announcement is typically in Japanese, with the English following. It may therefore be difficult for participants to conceive of their performance on this task given no real prior experience. Alternatively, the inconsistent responses to this can-do statement may have been subject to a contradiction contained within: stations or airports are typically loud and busy places, and announcements in such places are not likely to be delivered slowly and clearly.

Item (b) (from A2.2) also decreases the scale's reliability. When this statement is compared with the more reliable A2.2 statement, (d), it is evident that the latter includes greater detail regarding the circumstances surrounding performance of the task despite nearly identical content. This suggests that can-do statements may scale more reliably if the criteria of the can-do statement is more specific in that contextual and performance details of the task are provided (Green, 2012). However, it is unclear as to why these statements are at the same difficulty level when (d) appears to provide considerably more support to the listener than (b).

The third least reliable L descriptor is from level A1.3, item (c). Although the task entailed by this statement is deemed an L task, perhaps its lower reliability is due to the implication that spoken interaction is also required. Although the statement does not explicitly require a response in navigating the transaction, participants may not have considered this to be solely a listening task. Respondents may also have been confused about what kind of instructions or explanations are involved when shopping or eating out. If this can-do statement refers to listening to how products are made or how food is prepared, the difficulty of language required for that level of comprehension is likely much higher than A1.3.

Table 2. Mokken Scales for the CEFR-J A-Level Can-Do Statements for Reading

Rf	Level	*H*	Reading can-do statement
(e)	A1.1	.63	I can understand a fast-food restaurant menu that has pictures or photos, and choose the food and drink in the menu.
(f)	A1.1	.64	I can read and understand very short, simple, directions used in everyday life such as "No parking", "No food or drink", etc.
	A2.2	.66	I can understand short narratives and biographies written in simple words.
	A1.3	.67	I can understand short narratives with illustrations and pictures written in simple words.
	A2.1	.68	I can find the information I need, from practical, concrete, predictable texts (e.g., travel guidebooks, recipes), provided they are written in simple English.
	A1.2	.69	I can understand very short reports of recent events such as text messages from friends' or relatives', describing travel memories, etc.
	A1.2	.7	I can understand very short, simple, everyday texts, such as simple posters and invitation cards.
	A2.2	.72	I can understand the main points of texts dealing with everyday topics (e.g., life, hobbies, sports) and obtain the information I need.

Rf	Level	*H*	Reading can-do statement
	A2.1	.74	I can understand explanatory texts describing people, places, everyday life, and culture, etc., written in simple words.
	A1.3	.74	I can understand texts of personal interest (e.g., articles about sports, music, travel, etc.) written with simple words supported by illustrations and pictures.

Note. H-values represent the reliability of the scale as a whole.

In terms of the reliability of the reading scales (Table 2), tasks entailed by the least reliable statement, (e), do not seem to involve reading. Completion of this task could simply involve looking at photographs then pointing and nodding—behavior that is independent of language ability. This may also be the case for the second-least reliable statement, (f), also from A1.1. The examples in (f) are often presented graphically or concurrently with graphics and the directions may be comprehensible without reading. The reliability of this statement may increase if other short, simple directions that are not associated with images were included, thus rendering it a strictly reading task rather than picture-identification. Nonetheless, these statements do appear at the bottom of the Mokken Scale, in accordance with the CEFR-J's difficulty hierarchy, and in this sense, the responses to these statements were as expected.

Table 3. Mokken Scales for the CEFR-J A-Level Can-Do Statements for Spoken Interaction

Rf	Level	*H*	Spoken Interaction can-do statement
(g)	A2.1	.62	I can give simple directions from place to place, using basic expressions such as "turn right" and "go straight" along with sequencers such as first, then, and next.
	A1.1	.64	I can ask and answer questions about times, dates, and places, using familiar, formulaic expressions.
	A1.3	.65	I can make, accept and decline offers, using simple words and a limited range of expressions.

Rf	Level	*H*	Spoken Interaction can-do statement
	A1.1	.66	I can ask and answer about personal topics (e.g., family, daily routines, hobby), using mostly familiar expressions and some basic sentences (although these are not necessarily accurate).
	A1.2	.66	I can respond simply in basic, everyday interactions such as talking about what I can/cannot do or describing colour, using a limited repertoire of expressions.
	A2.2	.67	I can interact in predictable everyday situations (e.g., a post office, a station, a shop), using a wide range of words and expressions.
	A2.1	.67	I can get across basic information and exchange simple opinions, using pictures or objects to help me.
	A2.2	.68	I can exchange opinions and feelings, express agreement and disagreement, and compare things and people using simple English.
	A1.2	.69	I can exchange simple opinions about very familiar topics such as likes and dislikes for sports, foods, etc., using a limited repertoire of expressions, provided that people speak clearly.
	A1.3	.69	I can ask and answer simple questions about familiar topics such as hobbies, club activities, provided people speak clearly.

Note. H-values represent the reliability of the scale as a whole.

For spoken interaction, the lower reliability of the A2.1-level (g) in Table 3 could be accounted for by considering recently studied course content. Many participants rated this statement as easier than its predicted difficulty—as an A2.1-level statement, it should appear much lower in the table. One possibility is that this statement may have been considered more of a speaking skill by some participants, as giving directions could potentially entail responding to the initial request for directions rather than interacting in the traditional sense. However, half of the participants (the 2nd-year students) had recently become familiarized with completing this task whereas the other half (the 1st-year students) had little or no experience with it. For the 2nd-year student participants, three out of 30 lessons or 10% of the se-

mester's materials were focused on giving and following directions—essentially a task derived directly from this statement. In fact, this is also the case for tasks entailed by the spoken production statements from levels A2.2 and A2.1, (h) and (i) in Table 4, as the 1st-year students had recent experience with this task, having completed four out of 30 lessons (or just over 13% of the semester) on this topic. This suggests that differences inherent in participant demographics may significantly influence scaling and that both the homogeneity of the sample and recent experiences of participants should be kept in mind when determining difficulty. These findings also reiterate the importance of performing a reliability analysis rather than a difficulty analysis alone.

Table 4. Mokken Scales for the CEFR-J A-Level Can-Do Statements for Spoken Production

Rf	Level	*H*	Spoken Production can-do statement
(h)	A2.2	.69	I can make a short speech on topics directly related to my everyday life (e.g., myself, my school, my neighborhood) with the use of visual aids such as photos, pictures, and maps, using a series of simple words and phrases and sentences.
(i)	A2.1	.70	I can introduce myself including my hobbies and abilities, using a series of simple phrases and sentences.
	A1.3	.70	I can describe simple facts related to everyday life with a series of sentences, using simple words and basic phrases in a restricted range of sentence structures, provided I can prepare my speech in advance.
	A1.3	.70	I can express simple opinions about a limited range of familiar topics in a series of sentences, using simple words and basic phrases in a restricted range of sentence structures, provided I can prepare my speech in advance.
	A1.2	.71	I can give simple descriptions (e.g., of everyday objects) using simple words and basic phrases in a restricted range of sentence structures, provided I can prepare my speech in advance.

Rf	Level	*H*	Spoken Production can-do statement
	A1.1	.71	I can convey simple information (e.g., times, dates, places), using basic phrases and formulaic expressions.
	A1.2	.72	I can express simple opinions related to limited, familiar topics, using simple words and basic phrases in a restricted range of sentence structures, provided I can prepare my speech in advance.
	A1.1	.73	I can convey personal information (e.g., about my family and hobbies), using basic phrases and formulaic expressions.
	A2.2	.76	I can give an opinion, or explain a plan of action concisely giving some reasons, using a series of simple words and phrases and sentences.
	A2.1	.76	I can give a brief talk about familiar topics (e.g., my school and my neighborhood) supported by visual aids such as photos, pictures, and maps, using a series of simple phrases and sentences.

Note. H-values represent the reliability of the scale as a whole.

For writing, the third statement in Table 5 from A2.2, (j), negatively affects the reliability, possibly because it implicates use of a varied range of communicative competencies from W, R, and L. In this case, the reliability analysis might be highlighting the importance of unidimensionality in a can-do statement such that descriptors that implicate more than a single skill may behave less reliably.

Table 5. Mokken Scales for the CEFR-J A-Level Can-Do Statements for Writing

Rf	Level	*H*	Writing can-do statement
	A1.1	.62	I can fill in forms with such items as name, address, and occupation.
	A1.1	.64	I can write short phrases and sentences giving basic information about myself (e.g., name, address, family) with the use of a dictionary.

Rf	Level	*H*	Writing can-do statement
(j)	A2.2	.66	I can write my impressions and opinions briefly about what I have listened to and read (e.g., explanations about lifestyles and culture, stories), using basic every-day vocabulary and expressions.
	A2.1	.66	I can write texts of some length (e.g., diary entries, explanations of events) in simple English, using basic, concrete vocabulary and simple phrases and sentences, linking sentences with simple connectives like and, but, and because.
	A1.3	.67	I can write short texts about my experiences with the use of a dictionary.
	A2.2	.69	I can write a simple description about events of my immediate environment, hobby, places, and work, provided they are in the field of my personal experience and of my immediate need.
	A1.2	.70	I can write short texts about matters of personal relevance (e.g., likes and dislikes, family, and school life), using simple words and basic expressions.
	A1.3	.71	I can write a series of sentences about my hobbies and likes and dislikes, using simple words and basic expressions.
	A2.1	.75	I can write invitations, personal letters, memos, and messages, in simple English, provided they are about routine, personal matters.
	A1.2	.75	I can write message cards (e.g., birthday cards) and short memos about events of personal relevance, using simple words and basic expressions.

Note. H-values represent the reliability of the scale as a whole.

Conclusions

The reliability analysis (Tables 1-5) provided an alternate view of can-do statement scales by taking differing frequency distributions into consideration and revealing response patterns otherwise not evident if difficulty information alone is used to create a hierarchy. It was found that the can-do statements for each of the CEFR-J's A1 and A2 skills formed strongly reli-

able scales according to both Cronbach's alpha and Mokken Scaling's Coefficient of Homogeneity. Nonetheless, some statements negatively affected the reliability of the scale. Of particular concern are the higher level CEFR-J statements that were found close to the tops of Tables 1-5, as this reflects inconsistent difficulty ratings from a larger number of able participants.

Overall, the results indicate that the reliability of difficulty judgements on can-do statements may be affected by two main factors: the content of the can-do statement itself and specific characteristics of the population of respondents. In terms of the former, the results suggest reliability scores may be impacted by the specificity of criteria information (Green, 2012), whether the statement appeared to contain confusing or unfamiliar content, contradict itself, or imply either more than one skill or no language use whatsoever. Regarding the population of participants, reliability may be influenced by either familiarity or lack of experience with the task, whether participants had recently studied any material relevant to task performance, and the homogeneity of the population of participants.

This study provides some preliminary albeit limited findings on the reliability of the CEFR-J's A-level can-do statements and scales, suggesting that both could benefit from further empirical evidence to ensure that the system as a whole is functioning as intended. The analysis also highlights some considerations for future study. In this study, individual differences in a population of learners were shown to affect difficulty ratings and in turn, reliability scores on both the can-do statements and skill scales. Furthermore, examination of statements that were negatively affecting the reliability of the CEFR-J's skill scales suggested that content modification or adjustment in level may improve future versions of the system by increasing common understanding of the statements and their intended difficulties.

These findings have implications for future use of the CEFR-J and iterate issues associated with using can-do scales as measuring instruments for language proficiency. The importance of including checks for reliability is also emphasized, as individual learner characteristics are overlooked when difficulty ratings alone are used as the basis for creation of a scale. CEFR-J users should thus be mindful that unlike for the CEFR, which boasts significantly more supporting empirical evidence, sets of CEFR-J can-do statements may not behave identically or even similarly across and within different populations of learners. They should also be aware that estimations of levels derived from can-do instruments—via self-assessment or otherwise—may not be comparable within or across those same populations. Naturally, if task performance instead of self-assessment had been measured, differ-

ing reliability scores or response patterns might have been found. In fact, little research on the relationship between ability, self-assessment, and CEFR-aligned task performance for Japanese learners has been carried out. Further studies on this, the CEFR-J's target users' responses to can-do statements, and content analyses of the can-do statements should be performed to ensure a consistent, common interpretation of the system.

Acknowledgments

Judith Runnels was a lecturer and head of assessment at Hiroshima Bunkyo Women's University. Her research interests are in testing and evaluation and the CEFR and its usage in learner self-assessment. She is now a postgraduate research student in the UK.

References

Alanen, R., Huhta, A., Tarnanen, M. (2010). Designing and assessing L2 writing tasks across CEFR proficiency levels. In I. Bartning, M. Martin, & I. Vedder (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 21-56). Colchester, UK: Eurosla.

Andrich, D. (1978). Scaling attitude items constructed and scored in the Likert tradition. *Educational and Psychological Measurement, 38*, 665-680. http://dx.doi.org/10.1177/001316447803800308

Bentler, P. M., & Wu, E. J. C. (2012). EQSIRT 1.0 for Windows [Computer software]. Encino, CA: Multivariate Software.

Carroll, J. B. (1945). The effect of difficulty and chance success on correlations between items or between tests. *Psychometrika, 10*, 1-19. http://dx.doi.org/10.1007/BF02289789

Council of Europe. (2001). *The Common European Framework of Reference for languages: Learning, teaching, assessment.* Cambridge: Cambridge University Press.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

Ferguson, G. A. (1941). The factorial interpretation of test difficulty. *Psychometrika, 6*, 323-330. http://dx.doi.org/ 10.1007/BF02288588

Figueras, N. (2012). The impact of the CEFR. *ELT Journal, 66*(4), 477-485. http://dx.doi.org/ 10.1093/elt/ccs037

Fulcher G. (2004). Deluded by artifices? The Common European Framework and harmonization. *Language Assessment Quarterly, 1*, 253-266. http://dx.doi.org/10.1207/s15434311laq0104_4

Fulcher, G. (2010). The reification of the Common European Framework of Reference (CEFR) and effect-driven testing. In A. Psyaltou-Joycey & M. Matthaioudakis (Eds.), *Advances in research on language acquisition and teaching* (pp. 15-26). Thessaloniki, Greece: GALA.

Glover, P. (2011). Using CEFR level descriptors to raise university students' awareness of their speaking skills. *Language Awareness, 20*, 121-133. http://dx.doi.org/10.1080/09658416.2011.555556

Green, T. (2012). *Language functions revisited: Theoretical and empirical bases for language construct definition across the ability range*. Cambridge: Cambridge University Press.

Guttman, L. (1950). The basis for scalogram analysis. In S. A. Stouffer et al. *(*Eds.), *Measurement and prediction. Studies in social psychology in World War II* (pp. 60-90). Princeton, NJ: Princeton University Press.

Huhta, A., Luoma, S., Oscarson, M., Sajavaara, K., Takala, S., & Teasdale, A. (2002). DIALANG: A diagnostic language assessment system for learners. In J. C. Alderson (Ed.), *Common European Framework of Reference for languages: Learning, teaching, assessment. Case studies* (pp. 130-145). Strasbourg, France: Council of Europe.

Hulstijn, J. A. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *The Modern Language Journal, 91*, 663-667.

Imig, A. (2013). Development of EAP Textbooks based on the CEFR and learner/teacher autonomy support tools. *Framework and Language Portfolio SIG Newsletter*, *9*, 2-3.

Little, D. (2006). The Common European Framework of Reference for Languages: Content, purpose, origin, reception and impact. *Language Teaching*, *39*, 167-190. http://dx.doi.org/ http://dx.doi.org/10.1017/S0261444806003557

Mokken, R. J. (1971). *A theory and procedure of scale analysis with applications in political research.* New York: De Gruyter.

Molenaar, I. W. (1997). Nonparametric models for polytomous responses. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 367-380). New York: Springer-Verlag.

Negishi, M. (2011). CEFR-J kaihatsu no keii [The development process of the CEFR-J]. *ARCLE Review, 5*(3), 37-52.

Negishi, M. (2012). The development of the CEFR-J: Where we are, where we are going. Grant-in-Aid for Scientific Research Research Project Report (pp. 105-116). Retrieved from http://www.tufs.ac.jp/common/fs/ilr/EU_kaken/_userdata//negishi2.pdf

Negishi, M., Takada, T., & Tono, Y. (2013). A progress report on the development of the CEFR-J. In E. D. Galaczi & C. J. Weir (Eds.), *Exploring language frameworks: Proceedings of the ALTE Kraków Conference* (pp. 135-163). Cambridge: Cambridge University Press.

North, B. (2000). *The development of a common framework scale of language proficiency.* New York: Peter Lang.

North, B. (2002). Developing descriptor scales of language proficiency for the CEF common reference levels. In J. C. Alderson (Ed.), *Common European Framework of Reference for languages: Learning, teaching, assessment. Case studies* (pp. 87-105)*.* Strasbourg, France: Council of Europe.

North, B. (2007). The CEFR common reference Levels: Validated reference points and local strategies. In G. Francis (Ed.), *Report of the Intergovernmental Language Policy Forum: "The Common European Framework of Reference for languages (CEFR) and the development of language policies: Challenges and responsibilities*" (pp. 19-29). Strasbourg, France: Council of Europe.

North, B., Ortega, A., & Sheehan, S. (2010). *A core inventory for general English, British Council/EAQUALS.* Retrieved from: https://www.teachingenglish.org.uk/article/british-council-eaquals-core-inventory-general-english-0

North, B., & Schneider, G. (1998). Scaling descriptors for language proficiency scales. *Language Testing, 15*, 217-262. http://dx.doi.org/10.1177/026553229801500204

Runnels, J. (2013a). Preliminary validation of A1 and A2 sub-levels of the CEFR-J. *Shiken Research Bulletin*, *17*(1), 3-10.

Runnels, J. (2013b). Examining the difficulty pathways of can-do statements from a localized version of the CEFR. *Journal of Applied Research on the English Language, 2*(1)*,* 25-32.

Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory.* Thousand Oaks, CA: Sage.

SurveyMonkey. (2012). Surveymonkey.com©, LLC. Palo Alto, CA: www.surveymonkey.com.

Tono, Y., & Negishi, M. (2012). The CEFR-J: Adapting the CEFR for English language teaching in Japan. *Framework & Language Portfolio Newsletter, 8,* 5-12.

TUFS Tonolab. (2012). *CEFR-based framework for ELT in Japan*. Available from: http://www.cefr-j.org/english/index-e.html

van Schuur, W. H. (2003). Mokken scale analysis: Between the Guttman Scale and parametric item response theory. *Political Analysis*, *11*, 139-163.

Wang, H., Kuo, B., Tsai, Y., & Liao, C. (2012). A CEFR-Based computerized adaptive testing system for Chinese proficiency. *The Turkish Online Journal of Educational Technology*, *11*(4). Retrieved from: http://www.tojet.net/articles/v11i4/1141.pdf

Weir, C. J. (2005). Limitations of the Common European Framework for developing comparable examinations and tests. *Language Testing, 22*, 281-300. http://dx.doi.org/10.1191/0265532205lt309oa

Westhoff, G. (2007). Challenges and opportunities of the CEFR for reimagining foreign language pedagogy. *The Modern Language Journal, 91*, 676-679. http://dx.doi.org/10.1111/j.1540-4781.2007.00627_9.x

特性レベルの内発的動機づけを高める授業と有能性の欲求

Motivational Intervention and Satisfying Learners' Need for Competence

田中博晃（たなかひろあき）
Hiroaki Tanaka
広島国際大学
Hiroshima International University

Referring to motivation as a unitary concept is insufficient to explain and fully understand its dynamics in the classroom because individual motivation exists at different levels of generality. A previous study (Tanaka, 2009b) addressed the multiplicity of ways to represent intrinsic motivation. Intrinsic motivation was represented within the individual at three hierarchical levels of generality: intrinsic trait motivation, intrinsic classroom motivation, and intrinsic motivation to classroom activities. For the study, a motivational strategy was created for Japanese university students on the basis of self-determination theory (Deci & Ryan, 2002). In this study was examined the facilitating effect on the three motivational subconstructs and the three basic psychological needs (i.e., the needs for competence, autonomy, and relatedness). The correlational relationships between intrinsic trait motivation and the three psychological needs were not examined. The results showed that the strategy did facilitate intrinsic classroom motivation and intrinsic motivation to classroom activities; intrinsic trait motivation was not significantly enhanced. The results also showed

that the need for competence was not significantly increased as the study could not address the correlational relationship between competence and any of the three motivational subconstructs. Additionally, several motivational studies have shown that competence is an important facilitating factor for Japanese English learners, though the dynamics of competence in the motivational classroom have not been focused on so far.

Therefore, in the present study was examined the effect of a revised version of the motivational strategy created by Tanaka (2009b) on these three motivational subconstructs and the basic psychological needs. The purposes of this study are as follows: (a) to satisfy learners' basic psychological needs (especially the need for competence); (b) to enhance students' intrinsic motivation (especially intrinsic trait motivation); (c) to examine the relationship between intrinsic motivation (especially intrinsic trait motivation) and basic psychological needs (especially the need for competence); and (d) to describe how the need for competence is satisfied in the motivational classroom.

Fifty-eight university students who were enrolled in a 1st-year English language course participated in this study. The students met once a week in a 90-minute class. The motivational strategy used by Tanaka (2009b) was revised for this study. The motivational strategy was given to the students for 15 weeks. Prior to this intervention, questionnaires about intrinsic motivation and the three psychological needs were distributed. Items on intrinsic motivation consisted of three subconstructs: intrinsic trait motivation, intrinsic classroom motivation, and intrinsic motivation to classroom activities. Items on basic psychological needs consisted of three subconstructs: the needs for autonomy, competence, and relatedness. The same questionnaires were administered in the middle and at the end of the intervention. An open-ended questionnaire was also administered to students at the post-measurement stage. Descriptive statistics, correlation coefficients, one-way repeated ANOVA, and effect size were calculated to see the effect of the motivational strategy. The SCQRM M-GTA was also adopted to interpret qualitative data.

The results of quantitative analysis showed that (a) the intervention had a significant positive effect on all subconstructs of intrinsic motivation and basic psychological needs; (b) intrinsic trait motivation and intrinsic classroom motivation were strongly correlated with the need for competence; (c) intrinsic motivation for listening activities was strongly correlated with the need for autonomy and competence; and (d) intrinsic motivation for speaking activities was strongly correlated with the need for autonomy and relatedness. The results of qualitative analysis showed how the need for competence was satisfied in the motivational classroom. Once students valued academic activities and internalized them, they made an effort to use what they had learned in class. The overall result of this study was that the motivational-strategy intervention facilitated language acquisition and student competence.

本論では先行研究で用いられた動機づけを高める方略の改良版を使って、3つのレベルの内発的動機づけを高める教育的介入を行った。特に本論では、先行研究では効果が十分に発揮されなかった特性レベルの動機づけの上昇と有能性の欲求の充足に焦点を当てた。

また、内発的動機づけが高まる時の、学習者の英語授業に対する有能感を質的に把握した。日本人大学生58名に対して、15週間の教育的介入を行い、プレ測定、中間測定、ポスト測定の3時点で動機づけと3欲求の変動を検討した。その結果、改良版の方略によって、3つのレベルの内発的動機づけと3欲求のすべてを高められた。そこで3欲求と内発的動機づけの関連性の検討を行い、動機づけの上昇に貢献した要因を検証した。また自由記述データを「SCQRMを使ったM-GTA」によって分析し、動機づけが高まるときの有能感を把握することで、有能性の欲求が満たされるプロセスを探索した。その結果、学習内容の価値の内在化によって、学習者は学習事項を積極的に授業で使おうと努力し、それが学習事項の習得へとつながり、その結果、学習者は英語力の向上を自覚するという過程が示された。本論ではこれらの研究結果に基づく教育的示唆についても論じる。

近年の動機づけ研究では、動機づけを高める研究が活発に行われている。この種の研究は動機づけの構成要素の探索を行う1990年代の理論中心の研究（Dörnyei, 1990; Ehrman & Oxford, 1989; Kimura, Nakata, & Okumura, 2001）からシフトし、動機づけ理論の研究成果を教室現場に応用する研究（例えば; Dörnyei, 2001; Williams & Burden, 1997）の延長線上にある。特に近年の動機づけを高める研究では「動機づけを高める方略」（motivational strategies; Dörnyei, 1998）[1]の提案だけでなく、その方略をより具体化した上で、日本の英語学習環境で効果が検証されている。本論では検証された方略の1つである「外国ドラマ・映画を用いたコミュニケーション活動」（田中、2009b）を取り上げる。この方略は先行研究で動機づけを高める効果がある程度まで確認されていたが（田中、2009b、2010）、いくつかの限界点もあった。そこで本論ではそのような限界点を克服し、先行研究では検討が不十分だった動機づけを高める授業での有能性の欲求の働きに焦点を当てる。

背景

自己決定理論

　学習者の動機づけの中でも内発的動機づけを高める理論として有用なのが、「自己決定理論」（self-determination theory; Deci & Ryan、1985、2002）以降SDTと略記）である。SDTは内発的動機づけを高める要因として3つの心理欲求（以降、3欲求と略記）を提示している。3欲求とは具体的には、「自律性の欲求」（the need for autonomy）、「有能性の欲求」（the need for competence）、「関係性の欲求」（the need for relatedness）である。

　まず自律性の欲求とは、自身の行動がより自己決定的であり、自己責任を持ちたいという欲求である。教員による「報酬」（rewards）の付与や過剰な「統制」（control）を伴う学習は学習者の自律性の欲求の低下を導く一方、教員による「選択肢」（choices）の付与や「支援的フィードバック」（supportive feedback）を伴う学習は自律性を満たすのに効果的である。有能性の欲求とは、行動をやり遂げる自信や自己の能力を顕示する機会を持ちたいという欲求で、他者による「肯定的フィードバック」（positive feedback）の付与や学習者が適切な難易度の課題を達成することで有能性の欲求が満たされる。関係性の欲求とは、周りの人々や社会と密接な関係を持ち、他者と友好的な連帯感を持ちたいという欲求で、クラス内に協力的な学びの雰囲気を作り出すことで関係性の欲求が満たされる。

　内発的動機づけを高めるには、これらの3欲求を充足する必要がある。つまり学習者の3欲求を満たす仕掛けを授業活動に組み込むことで、SDTに基づいた動機づけを高める実践が可能になる。SDTは動機づけを高める要因を明確に提示し、且つそれらは英語授業に取り込みやすい（田中、2010）。これは他の動機づけ理論には見られない特徴である。またSDTの他の利点として、実証的な手法によって理論の妥当性を検証できる点（Dörnyei、1998）、日本人英語学習者の動機づけ研究に適用した研究例が多い点（田中、2010）も挙げられる。

自己決定理論を用いた動機づけを高める方略

　近年は、SDTを基にした動機づけを高める方略が考案され、その効果検証が行われている。まずSDTを背景にした動機づけを高める方略として「グループでのプレゼンテーション活動」（田中、2005、以降、GP活動と略記）がある。GP活動とは、学習者自らがリーディングの教科書内容と関連したテーマを見つけ出し、資料を集め、英文原稿を作成し、発表するというプロジェクト型の授業実践である。この活動の中に3欲求を満たす仕掛けを組み込むことで、学習者の内発的動機づけを高める方略である。日本人高校生（田中、2005）や日本人大学生（廣森・田中、2006；田中・廣森、2007）を対象にした研究で、GP活動の内発的動機づけを高める効果が確認されている。

　同様のプレゼンテーションベースの授業による動機づけを高める研究として、Maekawa and Yashima（2012）が挙げられる。機械工学専攻の大学2年生と3年生を対象にした調査の結果、最も内発的動機づけが低い群において、学習者の内発的動機づけを高める効果が確認された。他の方略としては"creative writing activities with student self-monitoring techniques"（Hiromori, 2006）があり、日本人大学生を対象にした調査で、この方略が外的調整状態の学習者と無動機状態の学習者の内発的動機づけを高めることが示された。以上の3種類の方略を扱った研究は、理論の実践を志向した研究としてSDTに基づいた方略が動機づけを高めることを実証的に示したと言えよう。

　これらの研究は調査協力者の特性的な動機づけ状態を高めた点で有益な知見を提供していると言えよう。その一方、これらの研究で検証された動機づけは、英語学習一般に対する動機づけであり、日々の英語授業に直結した動機づけを高める試みを行なっていない。理論の実践応用を志向するならば、英語学習一般に対する動機づけだけではなく、学習者の英語授業や授業活動など日々の授業に直結する動機づけを高めることも重要であろう。

外国語学習における動機づけの階層モデル

　英語授業に直結した動機づけを研究するための有用なモデルがVallerand and Ratelle（2002）の「階層モデル」（*hierarchical model*）である。上述の先行研究において、動機づけは一次元的な概念として捉えられていた。一方、Vallerand and Ratelle（2002）の階層モデルでは、対象の一般性に着目し、動機づけを複数の階層にレベル分けをして捉えている。つまり、最も一般性が低い対象を扱う動機づけを「状況レベル」（*situational level*）、最も一般性が高い対象を扱う動機づけを「包括レベル」（*global level*）、そして、その中間が「コンテクスト・レベル」（*contextual level*）とレベル分けをしている。

　ただ、Vallerand and Ratelle（2002）が想定しているのは、人間の社会生活全般での動機づけモデルであり、英語学習の動機づけのためのモデルではない。そこで、このモデルを日本の英語学習の文脈で動機づけを高める研究に応用できるように再構築したのが、田中（2009a）の「外国語学習における動機づけの階層モデル」である（図1参照）。このモデルは、学校教育での英語学習という観点からVallerand and Ratelle（2002）の階層モデルを再構築し、外国語学習における内発的動機づけを3つのレベルに下位分類している。このモデルの大きな特徴は、学校教育での動機づけを把握できるだけではなく、動機づけを高める方略の効果も把握できるように設計されている点である。

　1つ目のレベルがリスニング活動やスピーキング活動といった授業活動レベルに特化した最もミクロな内発的動機づけである「授業活動レベルの動機づけ」である。これは日々の授業内で行われる学習活動に直結した動機づけで、研究目的に合わせて「スピーキング活動への動機づけ」や「リスニング活動への動機づけ」という形で操作的定義が可能である。授業内の個々の活動に特化した点で、最も状況依存性の高い動機づけである。

　それに対して、最も一般性の高いマクロレベルの内発的動機づけが「特性レベルの動機づけ」である。これは動機づけを安定した特性として捉え、英語学習全般への動機づけを扱うレベルである。この中には、授業活動レベルの動機づけや英語授業レベルの動機づけも内包され、最も状況依存性の低い安定的な動機づけである。一般的には、質問紙調査で「あなたが英語を学習する理由や動機はどのようなものですか」と問うことで測定される。

　そして3つ目がその中間レベルの「英語授業レベルの動機づけ」で、日々の英語授業に直結した動機づけである。学校での英語学習というコンテクストでは、複数の英語授業が同時展開される。高校1年生であれば英語ⅠとオーラルコミュニケーションⅠが同時展開される。そのため、英語Ⅰの授業には意欲的に取り組むが、オーラルコミュニケーションⅠに対してはそれほど意欲的でない学習者も想定される。このように英語授業に特化した動機づけが英語授業レベルの動機づけである。このレベルの動機づけは、「英語Ⅰの授業への動機づけ」や「オーラルコミュニケーションⅠへの動機づけ」という形で操作的定義が可能である。
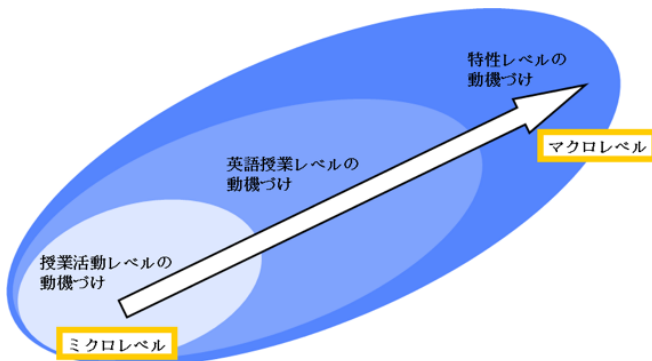

図1. 外国語学習における動機づけの階層モデル（田中、2009aを一部改変）. [2]

　以上の3つのレベルの内発的動機づけの内、日本人英語学習者を対象にした先行研究では2つ目の特性レベルを扱った研究が多い。例えば、プレゼンテーションベースの授業が工学専攻の大学生の動機づけに与える影響を検討したMaekawa　and　Yashima（2012）では、内発的動機づけ、外発的動機づけ（外的調整、取り入れ調整、同一視調整）、そして無動機を、すべて特性レベルとして扱っている。同様にGP活動の効果検証を行った研究（廣森・田中、2006；田中、2005；田中・廣森、2007）や"creative writing activities with student self-monitoring techniques"（Hiromori, 2006）の研究でも、調査対象の動機づけはすべて特性レベルである。このような介入型研究に加えて、調査型研究でも特性レベルの動機づけが研究対象となることが多い。例えば、小学生の動機づけを縦断的に捉えたNishida（2012）、日本人英語学習者の動機づけの構成要素を探索したKimura, Nakata, and Okumura（2001）などが挙げられる。

　特性レベルの動機づけは最も一般性が高く安定的であるため、これを測定したり、高める試みは極めて重要である。この特性レベルに加えて、外国語学習における動機づけの階層モデルを使うことで、英語授業レベルや授業活動レベルといった、より状況に特化した動機づけの変動を捉えられる。例えば、「英語の勉強は自体嫌いだけど、A先生の授業は好きだ」という学習者や、「読解の授業は好きではないけれど、本文を音読したりシャドーイングをするのは好きだ」という学習者の動機づけ状態は、特性レベルの動機づけだけでは十分に把握できない。このモデルを使うことで、前者の学習者の場合、特性レベルの動機づけは低いが、英語授業レベルの動機づけは高い状態と考えられる。また後者は音読やシャドーイング活動への動機づけが高い学習者だと言えよう。以上のように、今までの動機づけ理論は特性レベルの動機づけのみを把握するものであったが、このモデルを使うことで学習者の教室での動機づけ状態がより詳細に把握でき、より教育現場に直結した動機づけ研究が可能になる。

外国語学習における動機づけの階層モデルを用いた研究

　この外国語学習における動機づけの階層モデルに基づいた動機づけを高める研究として、加藤（2012）や岩中（2011）が挙げられる。加藤（2012）は外国語学習における動機づけの階層モデルの3レベルの内、英語授業レベルと授業活動レベルの動機づけを対象とした。この研究では、授業活動レベルの動機づけを研究目的に合わせて3つに細分化することで操作定義をしている。それらは「リプロダクション活動への動機づけ」、「シャドーイング活動への動機づけ」、「グループワーク活動への動機づけ」である。加藤はSDTの3欲求を満たす授業活動を日本人大学生42名に4ヶ月間行い、動機づけの変動を3時点で検討した。その結果、英語授業レベルと授業活動レベルの動機づけの両方において第1時点から第2時点での上昇が見られ、第3時点では維持期に入ることが示された。

　岩中（2011）は外国語学習における動機づけの階層モデルの中の、3レベルすべての動機づけの変動を2時点で検討した。この研究では特性レベルの動機づけを「英語学習への動機づけ」、英語授業レベルの動機づけを「英語授業への動機づけ」、活動レベルの動機づけを「英語授業活動への動機づけ」と呼んでいるが、これらは外国語学習における動機づけの階層モデルの3レベルの動機づけと同義である。この研究では、授業活動レベルの動機づけは研究目的に応じて、「リスニング活動への動機づけ」と「音読への動機づけ」に操作定義されている。日本人英語学習者19名に対して3

欲求を満たす教育的介入を行った結果、音読への動機づけと英語授業への動機づけの上昇が示された。

　上述の2つの研究だけではなく、「外国ドラマ・映画を用いたコミュニケーション活動」という方略を扱った研究がある（田中、2009a，2009b，2010）。これらの一連の研究では、多様な学習者を対象に3つのレベルの内発的動機づけを高める効果を詳細に検討している。

　外国ドラマ・映画を用いたコミュニケーション活動は、リスニングとスピーキングの教材に外国ドラマ・映画を用いて、各学習活動に学習者の3欲求を満たす仕掛けを組み込むことで内発的動機づけを高めることを目指す。この方略はもともと動機づけの低い学習者を対象にした方略で、動機づけの低い学習者から得られた質的データを基に考案された。田中（2009a）では英語授業への動機づけの低い学習者9名を対象にケース・スタディを行い、方略の効果を検証した。外国ドラマ・映画を用いたコミュニケーション活動を英語授業の中に取り入れ、3つのレベルの内発的動機づけの変動を検討した。その結果、自律性と関係性を高め、動機づけが低い学習者の英語授業レベルと授業活動レベルの動機づけを高める効果が確認された。

　この方略は動機づけの低い学習者以外にも効果がある。例えば田中（2009b）では、日本人大学生52名を対象に調査を行った。この研究における調査協力者は特性レベルの動機づけはやや高いものの、英語授業レベルの動機づけは中程度であった。[3] 外国ドラマ・映画を用いたコミュニケーション活動を取り入れた授業を15週間行い、教育的介入の前後（第1週目と第15週目）と中間時点（第7週目）の3回の測定で動機づけの変動を検討した。その結果、調査協力者の自律性、関係性、英語授業レベルの動機づけと授業活動レベルの動機づけが高められたことが示された。特に介入開始から中間測定までの間で動機づけの上昇が顕著であった。

　また田中（2010）では習熟度別に外国ドラマ・映画を用いたコミュニケーション活動の効果を比較する介入研究を行った。先行研究と同様に15週間の教育的介入の結果、外国ドラマ・映画を用いたコミュニケーション活動はTOEIC550点群と690点群の学習者の英語授業レベルと授業活動レベルの動機づけを効果的に高めることが示された。また先行研究と同様に、動機づけは前半の第7週までの間に大きく上昇していた。

　このように、外国ドラマ・映画を用いたコミュニケーション活動の効果は先行研究で扱われたどの方略よりも、より詳細な効果検証がされている。

問題の所在

　しかし外国ドラマ・映画を用いたコミュニケーション活動の効果を検証した先行研究に関して、いくつかの問題点が指摘できる。

　まずこの方略は特性レベルの動機づけを高める効果が弱かったため、3欲求の中のどの要素が最も効果的に特性レベルの動機づけを高めるのか不明である。例えば、田中（2009b）では外国ドラマ・映画を用いたコミュニケーション活動を取り入れた介入研究を行った結果、授業活動レベルと英語授業レベルの動機づけを高められたが、特性レベルの動機づけはほとんど高まらなかった。そのため、3欲求の変化と特性レベルの動機づけの変化の関係を検討できなかった。特性レベルの動機づけは3つ

のレベルの内発的動機づけの中で最も安定的な動機づけのため、変化させるのが最も難しい動機づけである。それを高めるには3欲求のどの要素を高めればよいか分かれば、教育現場に有用な知見を提供できよう。

また外国ドラマ・映画を用いたコミュニケーション活動での有能性の欲求の変化と3つのレベルの内発的動機づけの変化の関係も不明である。先行研究（田中、2009b）の調査では、この方略によって調査協力者の自律性と関係性は高められたものの、有能性の欲求における量的変化は小さく、7件法の質問紙で変化値はMdiff = 0.35だった。そのため、有能性の欲求の充足と3つのレベルの内発的動機づけの関係が十分に把握できていない。

一方、質問紙調査で得られたデータを構造方程式モデリングで分析した研究では、有能性の欲求と内発的動機づけとの間に比較的強い関係が見られる。例えば、特性レベルの動機づけと3欲求の関係を検討した廣森（2003）では、有能性の欲求から動機づけへのパス係数が最も大きかった。また英語授業レベルの動機づけを扱ったTanaka（2008）でも、廣森の結果と同様に、英語授業レベルの動機づけと有能性の欲求の影響関係が最も強かった。このように、動機づけを高めるには有能性の欲求が重要な働きをする可能性が高いが、外国ドラマ・映画を用いたコミュニケーション活動では有能性の欲求が十分に満たされていないため、英語授業に対する有能性の欲求の働きが十分に把握されていない。

以上、外国ドラマ・映画を用いたコミュニケーション活動において3つの問題点が指摘できる。

（1）特性レベルの動機づけの上昇と3欲求の変化の関係が不明である。

（2）有能性の欲求と3つのレベルの内発的動機づけの関係が不明である。

（3）動機づけが高まる時の、英語授業に対する有能性の欲求の役割が不明である。

そこで本論では、外国ドラマ・映画を用いたコミュニケーション活動をより強化した改良版の方略を使用する。この方略を用いて、学習者の3つのレベルの内発的動機づけと3欲求のすべてを高めたうえで、動機づけと3欲求の関係性を量的に検討する。また動機づけを高める授業の中での有能性の欲求の働きを把握するために、質的データの分析を行い、有能性の欲求の働きについて探索を行う。[4]

目的

本論では改良版の方略を用いて学習者の内発的動機づけを高めるための介入を行い、学習者の3つのレベルの内発的動機づけ（特に特性レベルの動機づけ）と3欲求（特に有能性の欲求）を高めた上で、3欲求と内発的動機づけの関連性の検討を行う。また、その介入の中で学習者の英語授業に対する有能性の欲求の質的側面を詳細に把握する。具体的には、本論の目的は以下の4点である。

（1）動機づけを高める方略によって、学習者の3欲求の中でも、特に有能性の欲求を満たす。

（2）動機づけを高める方略によって、学習者の3つのレベルの内発的動機づけの中でも、特に特性レベルの動機づけを高める。

（3）学習者の3欲求（特に有能性）及び3つのレベルの内発的動機づけ（特に特性

レベルの動機づけ)の関連性を検討する。
(4)動機づけが高まる時の、英語授業での有能性の欲求の働きを質的に捉える。
なお目的(1)と(2)は、目的(3)と(4)を達成するための前提条件となる。

調査

調査協力者と手続き

調査協力者は日本人大学生1年生58名(英語科目を専攻する学習者を含まず)である。3回の測定すべてに参加し、質問項目にすべて回答し、同じ項目にすべてマークするなど不誠実と判断される回答が含まれていないこと、を条件にデータをスクリーニングした結果、56名(男子23名、女子33名)が調査の対象となった。調査協力者のTOEIC IPのスコアの平均は507点であり、田中(2009b)と同じ学力水準にある学習者である。

調査は1年次の必修英語科目のクラス内で行われた。当該授業はコミュニケーション活動を中心とする授業で、全15回(週に1回90分)である。授業内に動機づけを高める方略として、外国ドラマ・映画を用いたコミュニケーション活動の改良版を実施した。

全15週の授業期間で3回の質問紙調査を行った。調査は第1回目、第8回目、第15回目の授業時に実施し、データ収集及び、教育的介入を含む授業は当該授業を担当する著者が行った。ただし研究者と授業者が同一のため、調査協力者が質問項目に率直に回答しにくい可能性がある。そこで、質問紙調査は学術目的および授業改善目的であり、単位認定や成績には一切関係ない点、個人が特定できる形での公開は行わない旨を調査前に説明した。また調査は授業の中で行われるため、調査協力に対する謝礼は発生しなかった。

第1回の授業は質問紙調査に加えて、授業の進め方の説明、CALL教室の機器の使用の説明、課題の提出方法などの説明を行い、具体的な教育的介入は第2回の授業から行った。質問紙調査の所要時間は10〜15分程度で、回答はコンピューター・ベースとした。

外国ドラマ・映画を用いたコミュニケーション活動の指導内容

ここでは教育的介入として用いられた外国ドラマ・映画を用いたコミュニケーション活動が、どのように指導され、それが3欲求にどう関わるかを概説する。まず田中(2009b)の方略を概説した後、改良版で追加・変更された点を説明する。

まず授業では教科書を使用せずに、教材プリントを毎週1枚ずつ配布し、それに沿って授業を行った。教材プリントではコミュニケーション活動が行えるように言語材料を機能別に配置した。主な内容は紹介する表現、人の性格を描写する表現、人の見た目を描写する表現、誘う表現、待ち合わせ場所や時間を決める表現、買い物をする表現、レストランを選ぶ表現、レストランで注文する表現、などであった。

授業の大まかな流れは、重要表現の学習、スピーキング活動、リスニング活動、という順番だった。スピーキング活動では外国ドラマ・映画のシーンを基に作成したモデル・ダイアログを提示し、それに沿って会話の練習を行った。会話の練習はペア活

動が中心だった。リスニング活動は、学習した言語材料が用いられている外国ドラマ・映画のシーンを使った。字幕を消した状態で大意把握を行った後、ディクテーション活動を行った。リスニング活動での聞き取りはコンピューターベースの個人活動で、答えあわせからは一斉授業のスタイルだった。

　改良版では、上述の活動に加えて学期末にペアでの発表活動を取り入れた。ペアは普段の授業と同じ組み合わせとし、発表はペアごとに英語のコミュニケーション活動をクラスメイトの前で行う形式とした。発表内容は授業で取り上げた言語材料を自由に組み合わせて、自分たちで設定した状況で会話する形式とした。発表時間はペアで3分程度とし、発表原稿は事前に作成させたが、発表時は原稿を見ないよう指導した。

外国ドラマ・映画を用いたコミュニケーション活動と3欲求の関連性

　次にこれらの活動と3欲求の関連について概説する（表1参照）。

表1. 外国ドラマ・映画を用いたコミュニケーション活動における3欲求への働きかけ

3欲求	働きかけ
自律性	• ペア活動を中心にした自律的タスク活動 • タスクが早く終わったペアに応用タスクを提供することで、個人のペースに合わせた学習 • 自宅での自主学習用の音声ファイルの提供 • 学習者個人のリスニング力に合わせた聞き取り時間の提供 • 7週目以降から目標志向学習の導入
有能性	• 難易度の高い課題の達成による成功体験の提供 • 知的好奇心を喚起する指導 • リスニングのつまずきに応じたヒントの提供 • 学習事項を応用練習する発表活動の導入 • 発表評価シートによるクラスメイトからの肯定的フィードバックの付与
関係性	• 学び合う雰囲気の育成 • ペアでの協力によるタスク活動 （※田中（2009b）よりも本研究の方が割合が多い） • 学習者個別に学習アドバイスの提供による教員との信頼関係づくり

Note. 波線部が改良版で新たに加えられた事項。

　第1に授業中に学習者のペースで学習できる時間を多く設けることで、学習者の自律性の欲求を満たすようにした。まずスピーキング活動ではペア活動を中心にタスク活動を行った。ただペアによってタスクをこなすスピードに差があるため、早く終わったペアには応用タスクも準備し、各ペアの学習速度にあわせた学習活動を取り入れ

た。またリスニングは教室に設置されている学習者用パソコンを使った。ネットワーク上から音声ファイルをダウンロードさせ、各自のペースで聞き取りを行わせた。これにより、学習者は聞き取りにくい箇所を何度も繰り返して聞くことで、自分のリスニング力に合わせた学習が可能になった。また音声ファイルはUSBメモリに保存させ、自宅での自主的学習もできるようにした。学期末の発表活動を事前に告知し、これからの学習内容を最後の発表に活かすように指導することで、授業での学習を目標志向にした。これにより授業に目的意識が生まれ、学習者はより自主的に学習に取り組むことで自律性の欲求の充足につながると考えられる。

　次に外国ドラマ・映画のリスニングという難易度の高い課題の達成は、達成感や有能感の獲得につながると考えられる。外国ドラマや映画は自然な速度で英語が発話されるため、速度に慣れるまでは難易度の高い活動である。そこで新規学習事項をリスニング前に指導した。また中学や高校で学習者が既習済みの語や表現を多く含むシーンを教材として選定することで、リスニング場面には未習事項が含まれないようにした。これによって、学習者が発話の速度にさえ慣れれば聞き取りができるようになり、学習者が成功体験を味わえるようにした。またリスニング活動中に学習者のつまずきに応じて学習者全員にヒントを与えた。リスニング活動は大意把握を行った後にディクテーションを行うが、ヒントはディクテーション時に与えた。10分程度のディクテーション後、活動を中断し、机間巡視中に把握した学習者のつまずきにヒントを提示し、その後、再びディクテーションをさせた。このように一回目の聞き取りで聞き取れなかった箇所でもヒントを参照することで聞き取りやすくなり、学習者の有能感が刺激されると考えられる。

　またスピーキング活動では、重要表現の提示だけでなく、会話の流れの中でどのように表現すると英語らしくなるかや、日常の何気ない言葉を英語にするとどうなるか、と問うなど、学習者の知的好奇心を喚起するように指導した。さらに、学期末の発表では今までの学習事項をできるだけ多く取り入れるように指示をした。これにより今までの学習事項を実際のコミュニケーションでどのように使うかを学習者に考える機会を与えられ、それが学習事項への深い理解につながると考えられる。このような学習事項の理解が有能感の獲得を導くと考えられる。また発表活動ではクラスメイトの発表に対して評価を行うように指示した。発表評価シートを用意し、発表に評価を記入させた。その際に学習者が肯定的フィードバックを得られるように、自由コメント欄で発表のよい点を取り上げるように指導した。発表評価シートは全員の調査協力者が発表を終えた後に、ペアごとに評価シートを返却した。このようなクラスメイトによる肯定的フィードバックを用いて、学習者の有能感の刺激を目指した。

　最後に、ペアでのスピーキング活動を主にすることで学び合う雰囲気を作り、関係性の欲求を満たすようにした。座席は第1回の授業で仲の良い友人同士で座らせ、その後はその席に固定し、座席の隣同士でペアを組ませた。これによりスピーキング活動時に英語を話しやすい雰囲気になるように工夫した。また席を固定することで、毎回同じペアで活動できるようにし、初回のオリエンテーション時には英語による自己紹介と日本語での簡単な雑談の時間を取ることで、知らないもの同士のペアであっても気軽に会話ができるように配慮した。ペア活動中において、授業者は机間巡視を行い、適宜、学習者にアドバイスを行った。リスニング活動中は、学習者がパソコンを使っての個人学習となるため、クラスメイトとの関わりはなくなる。しかし授業者が机間指導を行って、学習者の誤りやつまずきに対してアドバイスを与えることで、授

業者に質問しやすい雰囲気を作り、関係性の欲求に配慮した。また最後の発表はペアで行われるため、この2回の授業での発表準備はすべてペアワークであった。学習者は発表活動をうまく行うためにペアで相談しながら共同作業で学習を行う。それによって、互いに協力し合いながら学習活動が進められ、関係性の欲求が満たされると考えられる。

測定

質問紙はAcademic Motivation Scale（Vallerand, Pelletier, Blais, Briere, Senecal, & Vallieres, 1992）を参考にした田中（2009b）の質問紙を、学習者の実態を反映するように改良した（Appendix参照）。[5]　また通常の介入型研究はプレとポストの2回の測定であるが、本論では中間測定も加えることで動機づけの変動をより詳細に捉えられるようにした。[6]

質問紙は3部構成で、第1部は3つのレベルの内発的動機づけに関する質問項目で7件法である。測定項目内容は、特性レベルの動機づけ（α = .79, .77, .84）、英語授業レベルの動機づけ（α = .93, .91, .87）、そして授業活動レベルの動機づけである。授業活動レベルの動機づけは、授業内容に合わせてリスニング活動への動機づけ（α = .75, .72, .84）とスピーキングへの動機づけ（α = .88, .88, .88）とした。第2部は3欲求に関する質問項目で、自律性の欲求（α = .78, .80, .87）、有能性の欲求（α = .91, .91, .83）、関係性の欲求（α = .86, .94, .91）の3つの欲求を各4項目から7件法による測定である。第3部は自由記述形式で、授業への学習者の取り組みの自己評価、および授業の感想を問う形式である。

分析方法

量的分析

研究目的（1）から（3）を達成するために量的データの分析を行った。分析にはSPSS17.0Jを用いて、記述統計量の算出、一要因分散分析、相関係数の算出を行った。ただし有意差検定のp値はサンプルサイズの影響を受けるため、本論では効果量も算出した。また効果量の提示は今後の研究でのメタ分析のための情報提供も可能にする点でも有益であると判断した。効果量の算出には水本・竹内（2008）の分析シートを用いた。

質的分析

研究目的 (4) は仮説検証的な性質ではなく、仮説や理論の探索が目的なので、データからボトムアップで研究領域に密着した仮説や理論を生成できるグラウンデッド・セオリー・アプローチ（*Grounded Theory Approach*、以降、GTAと略記）を用いた。GTAにはGlaser and Strauss (1967) の方法、Strauss and Corbin (1990) の方法、さらに木下（2003）によるM-GTA（*Modified Grounded Theory Approach*）など複数ある。本論では「構造構成主義的質的研究法（*Structure-Construction Qualitative Research Method*、以降、SCQRMと略記）を使ったM-GTA」を用いた。「SCQRMを使ったM-GTA」とは、木下（1999、2003）のM-GTAの研究手続きを西條（2005、2007、2008）のSCQRMによって理論的に裏付けた比較的新しい質的研究法である。SCQRM自体は個別具体

的な質的研究の手法ではなく、むしろ質的研究法を理論的に補強するメタ理論である。そのためSCQRMを使ったM-GTAの具体的な研究手続きは木下（1999、2003）のM-GTAを基本としている。SCQRMを使ったM-GTA独自の特徴として、(1) 現実的制約を加味しながらM-GTAの研究手続きを研究目的に応じて柔軟に修正ができる、[7] (2) 量的研究との組み合わせに相性が良い、(3) より妥当な結果の一般化可能性を担保できる、という点が挙げられる（西條、2007、2008）。

　特に第1の現実的制約は、本論のような実践研究的側面のある研究デザインを実施していく上で重要な要素である。通常、M-GTAは自由記述形式の質問紙ではなく、インタビュー調査によってデータを収集する。なぜならば、インタビュー調査は調査者協力者との対話形式であり、複数回の実施が可能なので、データ分析の際に十分なヴァリエーションが担保できるようなデータ数を確保できるからである（木下、2003）。しかし本論はクラスルームでの実践的研究という現実的制約がある。つまり質的データを収集した段階が実験の最終日、つまり授業の最終日に該当し、その後は調査協力者は長期休業期間に入ってしまう。長期休業後は、調査協力者は習熟度クラス編成によってクラス替えがなされるため、再度、調査目的で調査協力者を集めることは困難であり、また授業での印象などの記憶が薄れている可能性が高い。もちろん長期休業期間に調査協力者の同意の上でインタビュー調査を行う手順も考えられるが、現実問題として調査協力者への負担が大きく、実施可能性は低いと言わざるを得ない。一方、自由記述形式の質問紙調査は短時間で一斉にデータ収集できるため、授業活動のまとめと評価として授業時間に実施しやすく、調査協力者への負担も少ないと考えられる。また調査者が授業担当教員であるため、有能性の欲求や有能感に関するデータを収集するには、調査者と面談形式になるインタビュー調査よりも質問紙調査の方が回答しやすい可能性もある。

　以上の現実的制約から、本論では質問紙調査によるデータ収集を行い、そのデータ分析にSCQRMを使ったM-GTAを用いる。ただし本論の研究目的が、これまで焦点化されてこなかった動機づけが高まる際の有能性の欲求の働きを詳細に捉えることである点も考慮して、本調査で得られた結果は比較的データ数の限られ状況から生成された概念やモデルを仮説として位置づけることとした。

SCQRMを使ったM-GTAの研究手続き

　本論のSCQRMを使ったM-GTAは、西條（2007、2008）をベースに、本論の研究目的をより妥当に達成できるように本論独自の手順を加えた7つのステップとした（表2参照）。

　まず西條（2007、2008）によるSCQRMを使ったM-GTAは関心の探索的明確化を行う。つまり研究目的を明確化した上で、その研究目的に沿った形で分析の視点を設定する。次にデータ収集を行い（Step 1）、それをテクスト化する（Step 2）。ただし本論は自由記述データを用いるため、インタビュー調査のような文字起こし作業は必要なく、データをほぼそのままの形で分析に使用できる。Step 3では、テクストデータから分析ワークシートを作成し、概念の抽出を行い、概念名を与える。次に抽出した概念の中で類似したものを上位概念であるカテゴリーに統合し、カテゴリー名を与える（Step 5）。最後に概念やカテゴリーを関係付けた概念図を作成し、それを基にストーリーラインを作成することで解釈を行う（Step 7）。これらの手続はM-GTAと大枠では変わらな

いが、すべての研究プロセスにおいて研究目的と相関的に意思決定することが明確化され、各研究プロセスで使用される用語が異なっている。以上が西條の提示するSCQRMを使ったM-GTAの概略である。次に、本論で独自に加えた手順について概説する。

<p style="text-align:center">**表2. SCQRMを使ったM-GTAの研究手続き**</p>

Step	手順	具外的手続き
0	関心の探索的明確化	研究目的の明確化、分析の視点の設定
1	関心相関的データ構築	データ収集
2	関心相関的テクスト構築	テクスト化
3	関心相関的分析ワークシート作成	概念化、新たに加えたデータと比較による概念の洗練
4	第3者による確認	応用言語学と質的研究の経験がある英語教員1名による確認作業
5	関心相関的理論構築	概念をまとめカテゴリーを作成
6	第3者による確認	応用言語学と質的研究の経験がある英語教員1名による確認作業
7	関心相関的理論構築（結果図・ストーリーラインの作成）	結果図の作成、解釈

質的データの解釈の妥当性の確保

　SCQRMを使ったM-GTAに限らず、質的データの解釈では分析者の独善的な解釈に陥らない配慮が必要である。そこで本論では独善的な解釈を回避する2つの方策をとった。

　1つ目が分析者に第3者を加える方策である。質的データの分析には本論の著者に加えて、応用言語学の知識を持つ英語教師1名の合計2名の合議で行った。両者とも質的研究に関して専門的な知識と経験を有しており、事前トレーニングは十分であると考えられる。質的分析の際は、解釈の恣意性を排除するために第3者に分析と解釈を委ねる場合もあるが、教室で起こっている学びを把握しているのは学習者と授業者であり、学習者のテクストデータからメッセージを読み取るには授業者の参加が不可欠である。よって本論では授業者である著者が分析を行い、分析は2名が個別に行うのではなく、2名の合議とした。具体的には、SCQRMを使ったM-GTAの研究手続きに第3者による解釈の確認作業をStep 4及びStep 6として追加することで、独善的な解釈を回避する試みを行った。ただし分析と確認を同時並行で行うため、Step 3とStep 4、及びStep 5とStep 6はそれぞれ同時並行とした。

　2つ目の方策が、解釈に至るプロセスの開示によって独善性を可能な限り排除する試みである。つまり、カテゴリー編成のプロセスを開示し、解釈に至る思考プロセス

のオープン化を目指した。質的分析を用いた多くの研究では、この思考プロセスがブラックボックス化されており、他者と研究結果に対する納得を共有できずに、結果が独善的に見えてしまうことがあった。本論ではこのような問題を回避するために、概念化のプロセスについて具体的に記述した。

結果

量的データの分析結果

3欲求と動機づけの変動

　　まず動機づけを高める要因である3欲求の検討を行った。記述統計量、および1要因分散分析によって、教育的介入が3欲求に与えた影響を検討した（表3及び図2参照）。

　　まず自律性は介入によって大きく上昇した（M_{diff} = 1.40, F (2, 110) = 41.90, p = .00, partial η^2 = .43）。第1時点から第2時点にかけて大きく上昇した後（M_{diff} = 0.98）、第2時点から第3時点にかけても継続的に上昇した（M_{diff} = 0.42）。ボンフェローニの方法による多重比較では第1時点と第2時点（p = .00, r = .63）、第1時点と第3時点（p = .00, r = .82）が5%水準で有意だった。[8]

　　次に有能性の欲求も介入によって上昇が見られた（M_{diff} = 0.79, F (2, 110) = 12.97, p = .00, η_p^2 = .19）。第1時点から第2時点にかけては微増であるが（M_{diff} = 0.25）、第2時点から第3時点にかけて上昇した（M_{diff} = 0.54）。ボンフェローニの方法による多重比較では、第2時点と第3時点（p = .00, r = .41）、第1時点と第3時点（p = .00, r = .63）は5%水準で有意だった。

表3. 3欲求の平均値と標準偏差、平均値の変化値と効果量

3欲求	M (SD)	M_{diff} (r)
自律性		
第1時点	3.83 (1.02)	
第2時点	4.82 (0.93)	0.98 (.63)
第3時点	5.23 (1.10)	1.40 (.82)
有能性		
第1時点	4.82 (1.08)	
第2時点	5.07 (0.86)	0.25 (.18)
第3時点	5.61 (0.76)	0.79 (.63)
関係性		
第1時点	4.92 (1.10)	
第2時点	5.28 (1.01)	0.36 (.31)
第3時点	5.99 (0.88)	1.06 (.75)

Note. M_{diff}は第1時点を基準に計算。

関係性においても大きな上昇が見られた（M_{diff} = 1.06, F (2, 110) = 31.20, p = .00, partial η 2 = .36）。第1時点から第2時点にかけては微増であるが（M_{diff} = 0.36）、第2時点から第3時点にかけて上昇した（M_{diff} = 0.70）。ボンフェローニの方法による多重比較では、第2時点と第3時点（p = .00, r = .57）、第1時点と第3時点（p = .00, r =.75）は5%水準で有意だった。以上の結果から、介入によって3欲求のすべてが高まっていることが示された。特に本研究の目的である有能性の欲求も介入によって上昇していた。

図2. 3欲求の平均値の変化.
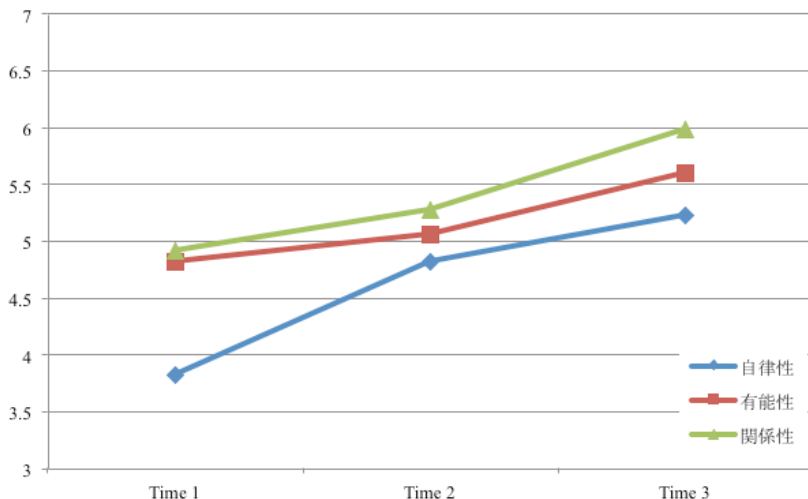
次に動機づけを高める方略の効果を検証するため、3つのレベルの内発的動機づけの検討を行った。3欲求と同様に、記述統計量、および1要因分散分析によって教育的介入が3欲求に与えた影響を検討した（表4及び図3参照）。

最も一般性の高いレベルの内発的動機づけである特性レベルの動機づけは、介入の前（第1時点）から後（第3時点）で、平均値の上昇が見られた（M_{diff} = 0.54, F (2, 110) = 7.43, p = .00, partial η^2 = .12）。第1時点から第2時点にかけて微増したのち（M_{diff} = 0.19）、第2時点から第3時点でもやや増加した（M_{diff} = 0.35）。ボンフェローニの方法による多重比較では、第1時点と第2時点（p = .00, r = .57）での上昇が5%水準で有意だった。

表4. 3つのレベルの内発的動機づけの平均値と標準偏差、および変化値

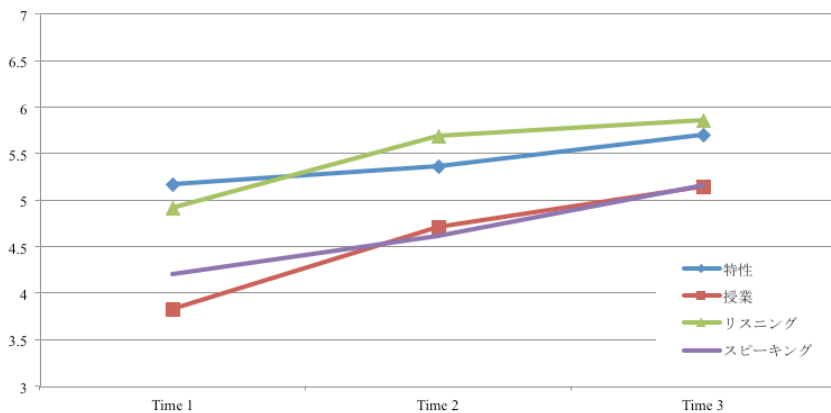| 内発的動機づけ	*M (SD)*	*M~diff~ (r)*
特性レベルの動機づけ		
第1時点	5.18 (0.97)	
第2時点	5.37 (0.81)	0.19 (.16)
第3時点	5.71 (0.86)	0.54 (.57)
英語授業レベルの動機づけ		
第1時点	3.83 (1.23)	
第2時点	4.71 (1.11)	0.89 (.51)
第3時点	5.15 (0.98)	1.32 (.77)
リスニング活動への動機づけ		
第1時点	4.92 (0.87)	
第2時点	5.69 (0.69)	0.77 (.59)
第3時点	5.86 (0.85)	0.94 (.74)
スピーキング活動への動機づけ		
第1時点	4.21 (1.27)	
第2時点	4.61 (1.06)	0.40 (.29)
第3時点	5.16 (1.12)	0.95 (.69)

Note. M_{diff}は第1時点を基準に計算。

図3. 3つのレベルの内発的動機づけの変化.

　　次に英語授業レベルの動機づけを検討する。英語授業レベルの動機づけは介入の前後で大きく上昇した（M_{diff} = 1.32, $F(2, 110)$ = 26.18, p = .00, η p2 = .32）。より詳細には、第1時点から第2時点にかけて著しく上昇し（M_{diff} = 0.89）、その後の第2時点から第3時点でも継続して上昇した（M_{diff} = 0.43）。ボンフェローニの方法による多重比較の結果から、第1時点と第3時点（p = .00, r = .77）、第1時点と第2時点（p = .00, r = .51）での英語授業レベルの動機づけの上昇が5%水準で有意だった。

　　最後に授業活動レベルの動機づけの変動を検討する。リスニング活動への動機づけは、介入の前（第1時点）から後（第3時点）で、平均値の大きな上昇が見られた（M_{diff} = 0.94, $F(2, 110)$ = 29.15, p = .00, η_p^2 = .35）。より詳細に検討すると、第1時点から第2時点にかけて大きく上昇した後（M_{diff} = 0.77）、第2時点から第3時点にかけては微増した（M_{diff} = 0.17）。ボンフェローニの方法による多重比較では、第1時点目から第2時点目（p = .00, r = .59）、第1時点目から第3時点目（p = .00, r = .74）でのリスニング活動への動機づけの上昇が5%水準で有意だった。

　　スピーキング活動への動機づけも介入の前後での上昇が見られた（M_{diff} = 0.95, $F(2, 110)$ = 17.95, p = .00, η_p^2 = .25）。より詳細には、第1時点から第2時点にかけて上昇した後（M_{diff} = 0.40）、第2時点から第3時点でも上昇が見られた（M_{diff} = 0.55）。ボンフェローニの方法による多重比較では、第1時点と第3時点（p = .00, r = .69）、第2時点と第3時点（p = .00, r = .41）での上昇が5%水準で有意だった。

3欲求と動機づけの関連性

　　次に動機づけの変動と3欲求の変動の関連性を検討する。3つのレベルの内発的動機づけの平均値の変化値と3欲求の平均値の変化値を用いて相関係数を算出することで、各要因の変化の関連性を検討した（表5参照）。

　　まず1時点目から3時点目の特性レベルの動機づけの変動と関連が強かったのは有能性の欲求であった（r = .45）。また1時点目から2時点目（r = .79）、2時点目から3時点目（r = .72）の特性レベルの動機づけも有能性の欲求と最も関連性を強かった。

　　同様の傾向は英語授業レベルの動機づけにおいて見られた。具体的には、1時点目から3時点目はr = .54、1時点目から2時点目はr = .72、2時点目から3時点目はr = .65だった。

　　一方、活動レベルの動機づけでは、1時点目から3時点目のリスニング活動への動機づけと関連が強かったのは自律性への欲求であった（r = .57）。1時点目から2時点目（r = .58）、2時点目から3時点目（r = .72）では、有能性の欲求が最も関連が強かった。スピーキング活動への動機づけでは、1時点目から3時点目で関連が強かったのは自律性への欲求であった（r = .45）。1時点目から2時点目（r = .54）、および2時点目から3時点目（r = .59）では関係性の欲求が最も関連が強かった。

表5. 3つのレベルの内発的動機づけと3欲求の相関係数

変数	1	2	3	4	5	6	7
1. 特性	-						
2. 授業	.18 (.73*/.54*)	-					
3. リスニング	.36* (.70*/.54*)	.32* (.58*/.51*)	-				
4. スピーキング	.35* (.35*/.38*)	.47* (.45*/.45*)	.52* (.37*/.37*)	-			
5. 自律性	.41* (.48*/.34*)	.49* (.60*/.60*)	.57* (.54*/.56*)	.45* (.32*/.30*)	-		
6. 有能性	.45* (.79*/.72*)	.54* (.72*/.65*)	.45* (.58*/.72*)	.42* (.43*/.48*)	.56* (.44*/.53*)	-	
7. 関係性	.14 (.31*/.40*)	.24 (.40*/.48*)	.23 (.34*/.44*)	.38* (.54*/.59*)	.21 (.27*/.47*)	.28* (.31*/.50*)	-

Note. 表記は1時点-3時点（1時点-2時点/ 2時点-3時点）である。
*は5%水準で有意。

質的データの解釈による分析過程と結果

データ分析に至るまでの過程

　次にSCQRMを使ったM-GTAを用いた質的研究の分析過程と結果について記述する。本論の質的研究は、4つ目の研究目的（動機づけが高まる時の、英語授業での有能性の欲求の働きを質的に捉える）に基いた分析者の視点で行われた（Step 1）。データ収集はコンピューターベースの質問紙調査であったため（Step 2）、自由記述データはテキスト化の必要はなく、そのまま分析対象のデータとして扱った（Step 3）。その結果、総文字数は7,503語であった。ただし、調査協力者から得られた自由記述データのすべてが有能性の欲求に関連するものではなかったため、有能性の欲求に関連する記述の選別を行った。

　自己決定理論における有能性の欲求の定義（Deci & Ryan, 2002）や、先行研究における自由記述データの分析結果（田中、2009b、2010）を基に「有能感」のカテゴリーには「分かるようになった」、「英語力に自信がもてるようになった」といった内容に該当する記述の分類を行った。[9]　具体例として「TOEICのリスニングの点が上がったのは、この授業のおかげだと思います。」や「教材がよかったので、しっかり聞こうとできたので、リスニングは向上したと思います。」という記述を「有能感」に分類した。その結果、112個の記述の中から25個の記述[9]を「有能感」の記述として選び出し、分析対象とした。総文字数は1,787文字で、1名の記述の平均は68.40文字だった。

データ分析の過程

　Step 3からは具体的なデータの分析作業に入った。各記述の内容と文字数を検討した結果、1名の記述が1つの意味内容にまとまっていたため、1名の記述を1つのデータとして分析ワークシートに追加していった。1つのデータを読み込み、その内容を適切に凝縮した表現を考案することで概念生成を行った。

　例えば、データ番号20（表6参照）であれば、「習った内容を会話の中で活かせたことがとても勉強になった。」という箇所は学期末の発表活動のことを指していると思われる。また「ほかの人が自分たちのコミュニケーションをどういう目で見ているのかを知れたことは今後の英語学習に活かせると思う。」という箇所も同様に、学期末の発表活動を指しており、特に発表終了後のピア・フィードバックに関する記述である。このような発表活動に対して、このデータは「やりがい」「今後の英語学習に活かせる」という活動の有効性について述べていることから、〈発表活動の成果〉という概念名を与えた。

　このように1つの概念化が終了すれば、データを新たに加えながら概念の生成や統廃合を繰り返し、概念の洗練化（オープン・コーディング）を進めた。分析の初期段階では木下（2007）に従い、文字数の多さと研究目的に関連する記述内容の豊かさという2つの観点を十分に満たしたデータから分析の対象とし、順次データを追加しながら、新たな概念の増加が見られなくなるまで分析対象のデータを加えた。その結果、すべてのデータが分析の対象となった。次のStep 4では第3者による確認が行われるが、本論では第3者と分析者（著者）が協議の上で分析を行ったため、Step 3とStep 4は同時並行であった。以上のプロセスを経て生成された概念とデータは表7のとおりである。[10]

表6. データ番号20の記述

データ番号	記述
20	習った内容を会話の中で活かせたことがとても勉強になった。ペアでコミュニケーションすることはとても楽しく、とてもやりがいがあった。ほかの人が自分たちのコミュニケーションをどういう目で見ているのかを知れたことは今後の英語学習に活かせると思う。

Note. 破線部は、概念の内容が強く表されている箇所を示す。

表7. Step 3とStep 4で生成された概念

概念名	データ番号とテクストデータ
a. 発表活動の成果	20)習った内容を会話の中で活かせたことがとても勉強になった。ペアでコミュニケーションすることはとても楽しく、とてもやりがいがあった。ほかの人が自分たちのコミュニケーションをどういう目で見ているのかを知れたことは今後の英語学習に活かせると思う。 24)最後のspeakingのテストは面白かったです。授業の表現を実際に使える機会があるのは良いことだと思いました。 25)コミュニケーションテストでは、今まで習った表現を使って文章を考えたため、身につきやすかったと思います。
b. 実際に使うことで身に付く実感	15)自分でも実際に授業で習った英語を使ってみることができてよかったと思います。ペアワークは楽しかったです。 23)リスニングでどういうときにその表現が使えるかよくわかったので、応用しやすいなと思いました。
c. 話すことでの記憶への定着	18)声に出したほうが記憶に残りやすいのでよかったと思う。 19)ただ書いて覚えるのではなく、隣の人と話すことで、より例文が記憶されたと思います。
d. 日常で役に立つ内容とう認識	8)コメディを取り扱っていたので興味を持ってみることができたし、日常生活に直結した内容だったので非常に参考になった。 13)教材に使われていたのがドラマだったので、実際の日常会話のリスニングができて、すごくためになったと思います。 14)スピーキングでは、日常でよく使うような会話表現がたくさん出てきたのでいい勉強になったと思います。 17)日常で使いそうなフレーズを多く学ぶことができてよかったです。フレーズをペアで覚えた後、リスニングですぐまたそのフレーズを確認できるというのも、ネイティブの発音で復習できて、よかったです。 21)教科書に載らない様なコミュニケーションの言い回しも多く、実際に日常生活で英語が役立つならこの様な会話だと思いました。その中でも、大切な表現を学べて良かったです。
e. 学習内容への満足感	7)今までの英語の学習の中で一番楽しく勉強ができました。もし、自分が社会に出て海外に行くことがあったら（たぶんないけど）、この授業を受け良かったな、思うことがあるかもしれません。特に、アメリカンジョークなんか、今までは学ぶことはなかったし、いい経験になりました。 22)今まで知らなかった会話表現がたくさん学べて良かったです。

f. 喜び	1)楽しかったです。今まで外国の映画やドラマは日本語に吹き替えてあるものを見たり、音声は英語でも字幕の日本語ばかりを見ていたりしたので、新しい体験で、最初はできるかどうか心配でしたがなんとかちょっとずつできるようになってきて嬉しいです。
	2)この授業で字幕なしでドラマなどを見るようになってから、日常生活の中でも、洋楽や外国の映画、ドラマを字幕なしで見てみようと心がけるようになりました。そうすると、少しずつですが、英語のリスニング能力が上がった気がして、授業のおかげだと思いました。
	5)海外ドラマや映画を見るのが好きなので、楽しく勉強できました。家で映画とか見てて単語や短い会話が聞き取れるようになったのがうれしかったです。
	9)興味がわく内容だったので、とても楽しんで、かつ集中してリスニングすることができました。リスニングのコツも教えていただき、ちょっと得意になったような気がします!こんなに楽しい授業は初めてでした。
g. 苦手意識の軽減	3)リスニングは苦手意識があったけど、この授業を通して少しだけできるようになったのではないかと思う。海外ドラマという取り組みやすいテーマはリスニングをする上でとてもよかった。
	4)リスニングはすごく苦手だったけど面白い教材のおかげで毎回楽しんでできました。たぶん前より少しは聞き取れるようになったと思います。
	11)教材がよかったので、しっかり聞こうとできたので、リスニングは向上したと思います。リスニングが苦手なので、とにかく少しでも多く聞き取ろうと頑張りました。
	16)英語のスピーキングはとても苦手な分野ではあるが、個人的にはがんばれたと思うし、授業開始前と比べると少しではあるができるようになっていると思う。
h. 繰り返し聞くことでの学習成果	6)最後に何回も聞く時間が与えられたことで個人のペースで学習できたのですごくよかったです。聞きとれないところを何回も何回も聞けたので納得するまでやれたので満足です。おかげでTOEICのリスニングの点数はあがりました。
	10)題材が面白く、楽しみながら見て聞くことが出来ました。わからないところは何回も聞きなおして自分なりの答えを出すことが出来るようになり、何回かやるうちに少しずつ出来ていきました。
i. テスト得点の上昇	12)TOEICのリスニングの点が上がったのは、この授業のおかげだと思います。

　次に生成された概念の中から共通する内容を探すことで、より簡潔なモデルになるように概念をカテゴリーに統合する作業を行った(Step　5)。この際に概念名だけでカテゴリー化するのではなく、データに立ち返りながら検討することで、独善的な解

釈にならないように配慮した。またStep 6の確認作業と同時並行に行うことで、カテゴリーの妥当性を高める試みも行った。その結果、8つの概念が3つのカテゴリーに圧縮された（表8参照）。最後にカテゴリー同士の関係を基に関係図を作成し、ストーリーラインを作ることで結果の解釈を行った（Step 7）。

表8. Step 3とStep 4で生成されたカテゴリー

カテゴリー	頻度 (%)	概念	データ 番号	頻度 (%)	
A.授業中に使うことでの習得	7 (28%)	a. 発表活動の成果	20) 24) 25)	3	(12%)
		b. 実際に使うことで身に付く実感	15) 23)	2	(8%)
		c. 話すことでの記憶への定着	18) 19)	2	(8%)
B. 価値の内在化	7 (28%)	d.日常で役立つ内容という認識	8) 13) 14) 17) 21)	5	(20%)
		e. 学習内容への満足感	7) 22)	2	(8%)
C. 英語力向上の自覚	11 (44%)	f. 喜び	1) 2) 5) 9)	4	(16%)
		g. 苦手意識の軽減	3) 4) 11) 16)	4	(16%)
		h. 繰り返し聞くことでの成果	6) 10)	2	(8%)
		i. テスト得点の上昇	12)	1	(4%)

有能性の欲求が満たされるメカニズム

　以上のSCQRMを使ったM-GTAから得られた関係図（図4参照）を基に、動機づけが高まる時の英語授業での有能性の欲求の働きについて検討を行う。なお、以降で表記する〈 〉は概念名、《 》はカテゴリーを表し、調査協力者のテクストデータは「 」に示す。

　まず本論の授業で用いた教材は、外国ドラマ・映画に出てくる日常会話表現を習得し、それを用いたコミュニケーション活動を行う内容であった。「教科書に載らない様なコミュニケーションの言い回しも多く、実際に日常生活で英語が役立つならこの様な会話だと思いました。その中でも、大切な表現を学べて良かったです」（データ番号21））という記述に見られるように、教材の中で扱われる会話表現を学習者は〈日常で役立つ内容という認識〉を持ったと考えられる。そしてその内容に、「今まで知らなかった会話表現がたくさん学べて良かったです」（データ番号22））という記述に見られるような、〈学習内容への満足感〉も獲得していた。以上のことから、外国ドラマ・映画を用いたコミュニケーション活動を通して、学習者は学習内容が自分にとって役立つ内容であると認識し、それに対して満足感を覚えていることが示された。自分に役立つという認識は、学習事項の価値を自分の中に取り込んだ状態であり、《価値の内在化》が起こっている状態と言えよう。
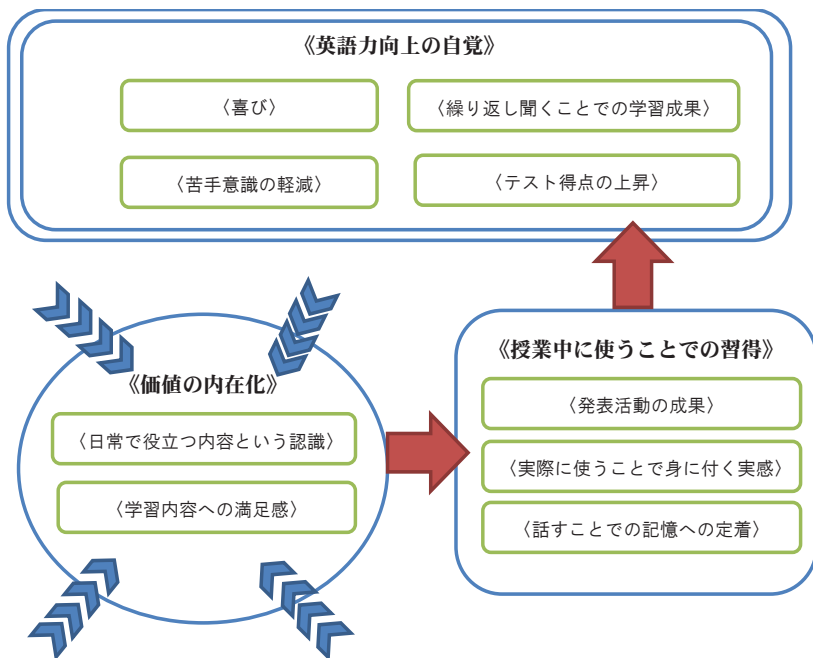
図4. 動機づけを高める授業での有能性の欲求が満たされるメカニズム.[11]

　学習事項の価値を内在化した学習者は、授業中のスピーキング活動やリスニング活動でこれらの学習事項を積極的に使おうとしたと考えられる。その結果、学習者は「ただ書いて覚えるのではなく、隣の人と話すことで、より例文が記憶されたと思います。」（データ番号19））という記述に見られるように、〈話すことでの記憶への定着〉や、「自分でも実際に授業で習った英語を使ってみることができてよかったと思います。ペアワークは楽しかったです。」（データ番号15））という〈実際に使うことで身に付く実感〉を得ることができたようだ。また学期末のペアによる発表活動では授業で習った事項を応用して、ペアでコミュニケーション活動の発表をする機会を学習者に与えた。「習った内容を会話の中で活かせたことがとても勉強になった。ペアでコミュニケーションすることはとても楽しく、とてもやりがいがあった」（データ番号20））という記述のように、発表活動においても成果が見られたようだ（〈発表活動の成果〉）。このように、学習者に授業中のコミュニケーション活動で学習内容を使う機会を与えることは、学習者に対して学習事項を活かす機会を与えることであり、それによって学習事項の習得を促すことが可能になったのだろう（《授業中に使うことでの習得》）。

　このような学習者の積極的な授業参加によって、学習者はやがて自分の《英語力向上の自覚》を覚えるようになったようだ。「海外ドラマや映画を見るのが好きなので、楽しく勉強できました。家で映画とか見てて単語や短い会話が聞き取れるようになったのがうれしかったです」（データ番号5））という記述のように、学習者は自分の

実力のつき具合に〈喜び〉を覚えたと考えられる。また「リスニングはすごく苦手だったけど面白い教材のおかげで毎回楽しんでできました。たぶん前より少しは聞き取れるようになったと思います」(データ番号4))という記述に見られるような、〈苦手意識の軽減〉もなされたようである。あるいは「最後に何回も聞く時間が与えられたことで個人のペースで学習できたのですごくよかったです。聞きとれないところを何回も何回も聞けたので納得するまでやれたので満足です。おかげでTOEICのリスニングの点数はあがりました」(データ番号6))に見られるように、授業中に繰り返しリスニングを復習する時間を多く取ったため、学習者は自分のペースで学習を進めることができ、〈繰り返し聞くことでの学習成果〉につながったと考えられる。

　以上を簡潔にまとめると、学習内容の価値の内在化によって、学習者は学習事項を積極的に授業で使おうと努力し、それが学習事項の習得へとつながり、その結果、学習者は英語力の向上を自覚する、というメカニズムが考えられる。このように獲得した有能感によって、学習者は有能性の欲求を満たしていると考えられる。

考察

動機づけを高める効果

　本論は改良版の外国ドラマ・映画を用いたコミュニケーション活動という動機づけを高める方略を用いて、特性レベルの動機づけを高める要因の検証、3欲求と3つのレベルの内発的動機づけの関係、そして外国ドラマ・映画を用いたコミュニケーション活動を用いた授業の中での有能性の欲求の働きについて検討した。

　この目的を達するためには、まず3つのレベルの内発的動機づけと3欲求すべてを高める必要があった。そこで改良版の外国ドラマ・映画を用いたコミュニケーション活動の3欲求と動機づけへの効果を検証した。その結果、本活動は英語授業レベルの動機づけや授業活動レベルの動機づけだけでなく、従来の外国ドラマ・映画を用いたコミュニケーション活動では高められなかった特性レベルの動機づけも高めることができた(M_{diff} = 0.54)。特性レベルの動機づけは1時点目から2時点目(M_{diff} = 0.19)、2時点目から3時点目(M_{diff} = 0.35)と継続的に上昇した。同様に英語授業レベルの動機づけも1時点目から2時点目(M_{diff} = 0.89)、2時点目から3時点目(M_{diff} = 0.43)と継続的に上昇し、変化値はM_{diff} = 1.32に達した。7件法の動機づけの質問紙において、M_{diff} = 1.32(r = .77)の変化は非常に大きく、改良版の方略の動機づけを高める効果が非常に大きかったと言えよう。授業活動レベルの動機づけにおいても、スピーキング活動への動機づけは1時点目から3時点目でM_{diff} = 0.94(r = .77)、リスニング活動への動機づけはM_{diff} = 0.95(r = .69)と、大きな上昇が見られた。このことから、改良版の外国ドラマ・映画を用いたコミュニケーション活動には動機づけを大きく高める効果があると言えよう。

3欲求を満たす効果

　3欲求においても、改良版の外国ドラマ・映画を用いたコミュニケーション活動は大きな効果が見られた。特に自律性への欲求に対しては1時点目から2時点目(M_{diff} = 0.98)、2時点目から3時点目(M_{diff} = 0.42)と継続的に効果を発揮し、1時点目から3時点目でM_{diff} = 1.40の上昇が見られた。7件法の質問紙によるデータであることか

ら、改良版の方略は極めて大きな効果があると言えよう。同様に関係性の欲求に対しても、1時点目から2時点目（M_{diff}=0.36）、2時点目から3時点目（M_{diff}=0.70）と継続的に効果を発揮し、1時点目から3時点目で関係性の欲求は大きく高まった（M_{diff}= 1.06）。このことから、改良版の方略は関係の欲求に対しても大きな効果があると言えよう。

　　また先行研究（田中、2009b）では効果が小さかった有能性の欲求（M_{diff} = 0.35）に対しては、1時点目から2時点目でM_{diff} = 0.25と微増だが、2時点目から3時点目ではM_{diff} = 0.54の上昇が見られ、その結果、1時点目から3時点目でM_{diff} = 0.79の上昇だった。1時点目から2時点目と比較して、2時点目から3時点目で有能性の欲求がより上昇したことから、2時点目の測定以降の発表活動が有能感に効果的だったと考えられる。質的データの分析結果では、調査協力者は学期末の会話テストで学習事項を実際にコミュニケーション活動に使うことで重要表現が身につき、会話に応用することにやりがいを感じていたことが示された。このことから学期末の発表活動が有能感を高める働きに貢献したと考えられる。以上のように、本論で用いた方略によって3つのレベルの内発的動機づけと3欲求のすべてを高めることができた。これらの内発的動機づけと3欲求が高まっていることは、目的（3）と（4）を達成する上での前提条件となっていた項目である。

動機づけの上昇と3欲求の関係

　　次に、3つのレベルの内発的動機づけの上昇に貢献した要因を検証した。先行研究（田中、2009b）では特性レベルの動機づけと有能感が十分に高まらなかったため、3欲求と3つのレベルの内発的動機づけを総合的に検討できなかったが、本論ではすべての動機づけと欲求が満たされていたため、このような検討が可能になった。

　　まず3時点での特性レベルの動機づけの上昇と最も相関関係が強かったのは、どの時点でも有能性の欲求だった（r = .45, .79, .72）。同様に英語授業レベルの動機づけの上昇に対しても、有能性の欲求の上昇が最も関連性が強かった（r = .54, .72, .65）。この結果は先行研究で行われたTanaka（2008）や廣森（2003）の構造方程式モデリングを用いた量的研究の結果とも一致することから、特性レベルの動機づけと英語授業レベルの動機づけを高めるには、学習者の有能感を効果的に高める必要があると言えよう。

　　一方、リスニング活動への動機づけでは、有能性の欲求だけでなく自律性の欲求も重要であった。第1時点から第2時点（r = .58）、第2時点から第3時点（r = .72）では有能性の欲求が最も関連性が強かったが、第1時点から第3時点では自律性の欲求が最も関連性が強かった（r = .54）。外国ドラマ・映画を用いたコミュニケーション活動でのリスニング活動は、学習者が一斉に取り組むのではなく、個人PCを用いて学習者自身の学習ペースにあわせて行われるため、自律性の欲求が重要な働きをすると考えられる。一方、リスニング活動中において学習者はPCと向き合っているため、どうしてもクラスメイトや教師との関係は希薄になるため、関係性の欲求の働きが小さくなったと考えられる。

　　それに対してスピーキング活動はペアで行われるため、関係性の欲求が重要な働きをしたと考えられる。第1時点から第2時点（r= .54）、第2時点から第3時点（r= .59）にかけては、スピーキング活動への動機づけの上昇と関係性の欲求が最も関連が強かった。一方、第1時点から第3時点までで2つの活動レベルの動機づけと関連が強

かったのは自律性の欲求であった（*r* = .57、*r* = .45）。田中（2009b）と異なるのは、介入の後半からペア活動の割合が多くなったためであったと考えられる。

　以上の点から、学習者の3つのレベルの内発的動機づけをすべて高めようとする場合は、3欲求すべてを高める必要があると言えよう。しかし個々のレベルだけを対象にする場合は、最も重視する要因が異なるだろう。つまり特性レベルと英語授業レベルの動機づけに対しては有能性の欲求、スピーキング活動への動機づけに対しては関係性の欲求と自律性の欲求、リスニング活動への動機づけに対しては有能性の欲求と自律性の欲求と言えよう。

有能性の欲求が満たされるメカニズム

　さらに本論では質的データの分析を通じて、動機づけを高める授業の中での有能性の欲求のメカニズムにも着目した。

　質的データの解釈による分析から、外国ドラマ・映画を用いたコミュニケーション活動を取り入れた授業での有能性の欲求が満たされるメカニズムは以下のように考えられる。学習内容の価値の内在化によって、学習者は学習事項を積極的に授業で使おうと努力し、それが学習事項の習得へとつながり、その結果、学習者は英語力の向上を自覚できるというメカニズムである。このように獲得した有能感によって、学習者は有能性の欲求を満たしていると考えられる。この質的研究から得られた有能性の欲求が満たされるメカニズムは、Elliot, McGregor, and Thrash（2002）によって裏付けが可能である。この研究によると、有能感には課題準拠（*task-referential*）、過去準拠（*past-referential*）、そして他者準拠（*other-referential*）の3つの側面があり、3つの有能感が高まることで有能性の欲求も高いレベルで満たされるとしている。

　まず外国ドラマ・映画を用いたコミュニケーション活動では、口語表現の習得が重要な課題である。《授業中に使うことでの習得》に見られるように、スピーキング活動において調査協力者は「ただ書いて覚えるのではなく、隣の人と話すことで、より例文が記憶された」（データ番号19））と感じたり、また学期末の発表活動で「今まで習った表現を使って文章を考えたため、身につきやすかった」（データ番号25））と感じたように、学習事項を会話の中に取り入れた活動で重要表現が定着していると感じ、その結果、調査協力者は課題準拠の有能感を獲得したと考えられる。

　また《英語力向上の自覚》のように、調査協力者は自分自身の英語力の伸びを実感し自信を持ち始めた。「今まで知らなかった会話表現がたくさん学べて良かったです」（データ番号22））や「少しずつですが、英語のリスニング能力が上がった気がして、授業のおかげだと思いました」（データ番号2））のように、会話表現が分かってきたという以前の学力からの伸びを実感していた。これは過去準拠の有能感と言えよう。

　以上の点から、本論での質的研究の結果は課題準拠と過去準拠の有能感に一致すると考えられる。

教育的示唆

　本論では「外国ドラマ・映画を用いたコミュニケーション活動」によって学習者の内発的動機づけを高めることに成功した。ただし、外国ドラマや映画を単に授業時間内に視聴することで動機づけが高まったわけではない。仮に外国ドラマや映画を単に

授業時間内に視聴しただけならば、学習者の中には講義を聞かなくて済むために授業が楽しいと感じる学習者もいるかもしれない。しかしその場合には、本論で扱ったリスニングやスピーキングといった授業活動レベルの動機づけや有能感など、英語学習に直結する動機づけや3欲求を高めることは不可能である。外国ドラマや映画は教材にすぎず、学習者の3つの欲求を満たす仕掛けを組み込んだ授業を設計することが重要である。つまり、3欲求を満たす仕掛けが組み込まれていれば、教材は外国ドラマや映画でなくとも構わない。本論では目標志向学習による自律性の欲求への働きかけ、肯定的フィードバックによる有能性の欲求への働きかけ、ペアでの協力的タスク活動による関係性の欲求への働きかけなどを行った。これらが3欲求を満たす仕組みを具体化したものであり、内発的動機づけを高める具体的要因である。他の研究例を挙げるならば、動機づけを高める授業の教材としてニュースを用いた岩中（2011）がある。この研究では、授業者が学習者との面談を行い、これまでの英語学習に関する情報収集から授業者と学習者の人間関係作りを行うという形式の関係性支援を行った。このようにSDTでは授業者や研究者の発想やアイデアを取り入れた3欲求への働きかけの具体化ができる。

　また本論では、英語授業レベルの動機づけを高めるには有能性の欲求が重要だと示された。コミュニケーション活動を主体とした授業で有能性の欲求を満たすには、学習者に学習事項をコミュニケーション活動の中で使わせることで、表現を定着させる必要がある。そのためには、モデルダイアローグに沿った単純な会話練習だけではなく、自分の表現したい内容に対して学習事項を応用する活動が必要である。また学習成果に対する肯定的フィードバックの付与も内発的動機づけには重要である。本論では、特にクラスメイトによる肯定的フィードバックが動機づけを高めるために効果的である可能性が示された。学習の成果を学習者自身が実感できるような働きかけを授業内に取り入れることで、有能感を高め、英語授業レベルの動機づけも高まると考えられる。

限界点と今後の課題

　最後に、本論の限界と今後の課題として、以下の4点を指摘しておく。

　第1に本論では外国ドラマ・映画を用いたコミュニケーション活動の限界点をいくつか克服できた。しかしこの活動によって、学習者の英語学習への取り組みが向上するのか、また英語習熟度は向上するのか、といった動機づけを超えた段階での検証はされておらず、今後の課題である。この方略が単に動機づけを高めるだけではなく、実際に学習者の授業時間外学習をどの程度増やし、また習熟度の向上にどの程度貢献するのかが明らかになれば、この方略の使用価値が高まると考えられる。

　第2に本論では特性レベルの動機づけだけではなく、英語授業レベルと授業活動レベルというよりミクロなレベルの動機づけを扱った。レベルをよりミクロにすればするほど、英語授業に特化した動機づけが扱える一方、コンテクストや状況に依存する割合が大きくなる。今後、ケースを重ねることでより一般性の高い結果が得られると考えられる。

　第3に、本論では動機づけを高める理論としてSDTを用いた。確かにSDTは動機づけを高めるために有益な理論であるが、唯一の理論ではない。今後は多様な理論を取り入れながら、包括的に内発的動機づけを高める方略の効果を検証する必要がある。

　第4に、本論の質的研究で得られた結果はまだ仮説の段階である。今後はこの結果を質問紙に反映するなどして、量的な検証を行う必要がある。特にElliot, McGregor, and Thrash (2002)らの3つの有能性の欲求の側面の内、質的データから他者準拠の有能感に関する記述は得られていない。今後は質問紙調査を行うなど、さらなる検証が必要だろう。

　以上のように本論ではいくつかの課題を残したものの、先行研究では扱えなかった問題をいくつか克服できた。[12]　今後の研究の積み重ねで、上述の問題も解決する必要があろう。

注釈

1. Dörnyei (1998)は授業の中で学習者の動機づけを高め、維持する方法やテクニックと定義している。以降、本論では単に方略と記述した場合も、動機づけを高める方略を意味する。

2. 本論で提示する「外国語学習における動機づけの階層モデル」は、田中（2009a）版のモデルの用語を一部整理したものである。なお、このモデルが扱う「特性レベルの動機づけ」、「英語授業レベルの動機づけ」、「学習活動レベルの動機づけ」はすべて内発的動機づけの下位概念である。本文中や先行研究において内発的動機づけを強調する場合は「特性レベルの内発的動機づけ」と記載しているが「特性レベルの動機づけ」と同義である。

3. 特性レベルの動機づけの初期値は7件法の質問紙で$M = 4.89$、英語授業レベルは$M = 3.68$である。

4. 田中（2013）ではスピーキング活動やリスニング活動といった、学習活動レベルでの有能性の欲求について検討したので、本論ではそれより1つ上のレベルである英語授業レベルで有能性の欲求を質的に把握する。

5. 第1の改良点は、「外国ドラマ・映画を用いたコミュニケーション活動」を用いた授業内容をより反映できるように項目を改めた点である。特に有能性の欲求は、田中（2009b）の質的研究の結果を反映させた項目の改訂を行なった。2点目は田中の質問紙は態度要因と意識要因の両方が項目に含まれている箇所があったが、本論では意識要因のみの項目に限定した。

6. 本論のデータは田中（2013）のデータに新規データを加え上で、新たな分析を行った。また第3部では、調査協力者に授業の感想と授業への取組みの2つの記述を求めたが、調査協力者はこの2つを混同して記述している場合が多かった。よって本論では2つを区別せずに分析した。

7. ここで言う研究手続きとは、研究の関心を明確化する段階から研究計画の立案、データ収集と分析、結果の解釈に至るまでの一連の研究全体の手続きを指す。

8. 水本・竹内（2008）によると、効果量のrは相関係数に基づく効果量で0～1の間の値を取る。$r = .10$は効果量が小、$r = .30$で効果量が中、$r = .50$で効果量が大とみなす。partial η^2には明確な基準は設定されていない。

9. 有能感に関する肯定的な内容だけでなく、否定的な内容も対象とした。

10. 本論では25個の記述内容すべてを論文内で公開することで、質的分析の結果の妥当性を再検証できるようにした。これは量的研究で言うところの、分析対象となった生データ（raw data）の公開である。記述内容は表7のテクストデータを参照。

11. 《価値の内在化》に付随する4つの矢印は価値が中に取り込まれているイメージを表すものである。

12. 本論は科学研究費助成（課題研究番号：25770217）の研究成果の一部である。

　田中博晃（Hiroaki Tanaka）は広島国際大学看護学部に所属し、専門は動機づけや研究手法（量的と質的）などである。

引用文献

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

Deci, E. L., & Ryan, R. M. (2002). *Handbook of self-determination research*. Rochester, NY: University of Rochester Press.

Dörnyei, Z. (1990). Conceptualizing motivation in foreign language learning. *Language Learning, 40*, 46-78. http://dx.doi.org/10.1111/j.1467-1770.1990.tb00954.x

Dörnyei, Z. (1998). Motivation in second and foreign language learning. *Language Teaching, 31*, 117-135. http://dx.doi.org/10.1017/S026144480001315X

Dörnyei, Z. (2001). *Motivational strategies in the language classroom*. Cambridge: Cambridge University Press.

Ehrman, M., & Oxford, R. (1989). Effects of sex differences, career choice, and psychological type on adult language learning strategies. *The Modern Language Journal, 72*, 1-13.

Elliot, A. J., McGregor, H. A., & Thrash, T. M. (2002). The need for competence. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 361-388). Rochester, NY: University of Rochester Press.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine Publishing Company.

廣森友人．（2003）．「学習者の動機づけは何によって高まるのか－自己決定理論による英語学習動機づけの検討」．*JALT Journal*, *25*, 173-186.

Hiromori, T. (2006). The effects of educational intervention on L2 learners' motivational development. *JACET Bulletin, 43*, 1-14.

廣森友人・田中博晃．（2006）．「英語学習における動機づけを高める授業実践：自己決定理論の視点から」．*Language Education and Technology, 43*, 111-126.

岩中貴裕．（2011）.「学習意欲の向上に貢献する教室活動－考慮すべき3つの心理欲求–」.『JACET中国・四国支部研究紀要』, *8*, 1-16.

加藤澄恵．（2012）.「学習活動が英語学習者の内発的動機に与える影響の検証」.『千葉大学言語教育センター言語文化論叢』, *6*, 9-22.

Kimura, Y., Nakata, Y., & Okumura, T. (2001). Language learning motivation of EFL learners in Japan—A cross-sectional analysis of various learning milieus. *JALT Journal, 23*, 47-68.

木下康仁．（1999）.『グラウンデッド・セオリー・アプローチ－質的実証研究の再生－』. 東京都：弘文堂.

木下康仁．（2003）.『グラウンデッド・セオリー・アプローチの実践－質的研究への誘い－』. 東京都：弘文堂.

木下康仁．（2007）.『ライブ講義M-GTA　実践的質的研究法－修正版グラウンデッド・セオリー・アプローチのすべて－』. 東京都：弘文堂.

Maekawa, Y., & Yashima, T. (2012). Examining the motivational effect of presentation-based instruction on Japanese engineering students—From the viewpoints of the ideal self and self-determination theory. *Language Education and Technology*, *49*, 65-92.

水本篤・竹内理.（2008）.「研究論文における効果量の報告のために－基礎的概念と注意点－」.『英語教育研究』、*31*, 57-66.

Nishida, R. (2012). A longitudinal study of motivation, interest, can-do and willingness to communicate in foreign language activities among Japanese fifth-grade students. *Language Education and Technology*, *49*, 23-46.

西條剛央．（2005）.『構造構成主義とは何か―次世代人間科学の原理』.　京都：北大路書房.

西條剛央．（2007）.『ライブ講義・質的研究とは何か―SCQRMベーシック編』.　東京：新曜社.

西條剛央．（2008）.『ライブ講義・質的研究とは何か―SCQRMアドバンス編』.　東京：新曜社.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Los Angeles, CA: Sage.

田中博晃.（2005）.「どのようにすれば学習者の動機づけは高められるのか?: 学習者の動機づけを促進する方略の効果検証」.*JLTA Journal, 7*, 163-176.

Tanaka, H. (2008). What facilitates students' active engagement in learning English in the classroom? Causal analysis between psychological needs, English classroom motivation and engagement. *JLTA Journal, 11*, 96-108.

田中博晃.(2009a).「動機づけが低い学習者の動機づけを高める実践研究」.『JACET 中国・四国支部研究紀要』, 6, 69-82.

田中博晃.(2009b).「3つのレベルの内発的動機づけを高める: 動機づけを高める方略の効果検証」. JALT Journal, 31, 227-250.

田中博晃.(2010).「英語の授業で内発的動機づけを高める研究」. JACET Journal, 50, 63-80.

田中博晃.（2013）.「動機づけを高める方略の修正：特性レベルの内発的動機づけを高める教育的介入」. JACET Journal, 56, 81-106

田中博晃・廣森友人.(2007).「英語学習者の内発的動機づけを高める教育実践的介入とその効果の検証」. JALT Journal, 29, 59-80.

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Briere, N. M., Senecal, C., & Vallieres, E. F. (1992). The academic motivation scale: A measure of intrinsic, extrinsic, amotivation in education. *Educational and Psychological Measurement, 52*, 1003-1017.

Vallerand, R. J., & Ratelle, C. F. (2002). Intrinsic and extrinsic motivation: A hierarchical model. In E. L. Deci and R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 37-63). Rochester, NY: University of Rochester Press.

Williams, M., & Burden, R. (1997). *Psychology for language teachers; A social constructivist approach*. Cambridge: Cambridge University Press.

Appendix
質問紙調査で用いた質問項目

【特性レベルの動機づけ】
1. 英語を勉強している時に、「あっそうか」や「なるほど」と思うような発見がある。
2. 英語を勉強することで、初めて気づくことがあると嬉しい。
3. 英語圏の人々や、彼らの生活様式について知るのは楽しい。
4. 英語ができるようになると、今までとは違う自分の新しい一面を見ることができる。
5. 英語を勉強し続けていると、今まで聞き取れなかった単語や言葉がわかるようになるのが嬉しい。

【3欲求】
1. 英語の授業では、教材・授業の進め方・学習内容に関して、私達にある程度の選択の自由が、与えられていると思う。
2. 英語の授業では、先生は私たちの授業に関する意見を尊重してくれていると思う。
3. 英語の授業では、授業の進め方の希望などを、先生に伝える機会が与えられていると思う。

4. 英語の授業では、プレッシャーを感じずに勉強をすることができると思う。
5. 英語の授業では、今まで出来なかったことが、少しずつでも出来るようになったと感じることがある。
6. 英語の授業では、回を重ねるごとに少しずつでも、分かるようになっていると感じることがある。
7. 英語の授業では、今まで知らなかったことが分かるようなったと感じることがある。
8. 英語の授業では、学習した多様な表現をどのように使うのかを分かったと感じることがある。
9. 英語の授業では、同じ教室の仲間と仲良くやっていると思う。
10. 英語の授業でのグループ活動・ペアワークでは、協力し合う雰囲気があると思う。
11. 英語の授業では、和気あいあいとした雰囲気があると思う。
12. 英語の授業では、同じ教室の仲間同士で学びあう雰囲気があると思う。

【授業活動レベルの動機づけ】
13. 英語の授業では、リスニングに集中しようとしている。
14. 英語の授業では、リスニングに熱心に取り組もうとしている。
15. 英語の授業では、英語をがんばって聞こうという気持ちがある。
16. 英語の授業では、楽しくリスニングをできると思う。
17. 英語の授業では、リスニング活動は好きな方だ。
18. 英語の授業では、スピーキングに集中しようとしている。
19. 英語の授業では、楽しくスピーキングをできると思う。
20. 英語の授業では、スピーキング活動は好きな方だ。
21. 英語の授業では、スピーキングに熱心に取り組もうとしている。

【英語授業レベルの動機づけ】
22. 英語の授業は、おもしろいと思う。
23. 英語の授業は、楽しくて時間が過ぎるのが早いと感じる。
24. 英語の授業を、楽しみにしている。
25. 英語の授業では、好奇心が刺激されると思う。

Reviews

***Replication Research in Applied Linguistics*. Graeme Porte (Ed.). Cambridge: Cambridge University Press, 2012. xiv + 286 pp.**

Reviewed by
Keith Adams
Tohoku Gakuin University

In the physical sciences, replication studies, which repeat a previous study exactly or with strictly controlled modifications, are highly valued because replication is seen as an essential process to evaluate the generalizability of the results of a study. However, replications are not held in such high regard in SLA research where, according to Porte, relatively little replication research has been conducted and published. This lack of enthusiasm for replications can be partly attributed to reservations by some researchers, such as Block (as cited by Charlene Polio on p. 50), about the suitability of replications in the domain of second language learning. However, in the Introduction, Porte maintains that a more influential factor inhibiting the acceptance of replications is the importance placed on *originality* in SLA research. In essence, replication research is philosophically perceived as being inferior to studies that claim to be original, innovative investigations (Valdman, 1993).

The first overall objective of *Replication Research in Applied Linguistics* is therefore to change negative attitudes so that replications are regarded as legitimate and vital contributions to SLA research. The second complementary objective is to present a practical framework for encouraging a wider understanding and dissemination of replication studies. To achieve these goals, specific reasons why journals, university promotion committees, and the general academic community tend to regard replications negatively are identified. In response, theoretical arguments and detailed practical guidelines are offered to encourage a reassessment of the value of replication research.

The book is divided into three parts containing a total of nine chapters written by a variety of authors. Each chapter begins with an introduction, which lays out the purpose of the chapter and the content areas to be discussed, and most conclude with a concise summary. I found both of these features to be very helpful in highlighting the focus of each chapter. In addition, references are provided at the end of each chapter, rather than combined into a potentially overwhelming collection at the end of the book.

Part I, "The case for replication studies," contains four chapters that offer a synthesis of perspectives to establish the theoretical basis for the central argument of the book. In the first chapter, "Why (or why not), when, and how to replicate research," Alison Mackey expands upon some of the main themes first introduced by Porte and then moves on to the critical issue of identifying features of previous studies which may or may not make them suitable candidates for replication.

In Chapter 2, Charlene Polio presents a detailed history and analysis of published studies in applied linguistics that claimed to be replications. Polio focuses on studies of written error correction and assesses the studies to determine whether they met the necessary requirements of a proper replication or fell short due to design or measurement flaws.

In the following chapter, Hossein Nassaji addresses misconceptions about using statistical significance tests in support of the generalizability of results. Nassaji calls for resampling of data through replications that use additional statistical measures, such as effect size, and even the somewhat controversial matter of internal replications by the original researcher.

In the final chapter of Part I, Luke Plonsky looks at meta-analysis, in which data from multiple replication studies are analyzed to determine generalizability. Plonsky includes an informative table of meta-analyses of several topics of interest to the practicing teacher, such as studies in corrective feedback and L2 strategy instruction, and a detailed instrument for evaluating the quality of L2 meta-analyses.

The two chapters in the second section of the book, "Replication studies in graduate programs," provide examples and models of how replication research can be integrated into instruction. In Chapter 5, Rebekha Abbuhl highlights practical skills such as critical reading instruction for evaluation of previous studies and a genre-based approach to writing to facilitate publication of a study. Chapter 6 focuses on a postgraduate research program at Swansea University. Tess Fitzpatrick first examines the use of replications to provide novice researchers with models to begin their own empirical studies quickly. She concludes by offering supervisors and students a general guide

for choosing, conducting, and publishing a replication study. An interesting feature of this chapter is the inclusion of extensive student comments about the value of replication studies in research training.

Part III, "Research studies in practice," contains three chapters that address replication studies in a holistic fashion. J. D. Brown begins his chapter, "Writing up a replication report," by looking at the structure and kinds of information that should be included in a replication study and how a replication may differ from the original study in research methods (e.g., quantitative or qualitative) or replication type (e.g., approximate or conceptual). Brown then uses this discussion as the framework for an in-depth analysis of a replication study of negotiation of meaning by Eckerth (2009). In a very entertaining twist in the format of the book, the reader is then able to read Eckerth's study in its entirety in the following chapter, which also includes comments on the replication by the author of the original study. This format added a very valuable, interactive aspect to the book, as Brown's guidance and framework gives the reader the tools for a "hands-on" (p. 195) experience of analyzing and evaluating a specific study.

The third chapter in this section by Susanne Rott looks at L2 vocabulary acquisition and provides an opportunity to apply the framework to a different replication study without the benefit of any specific prereading comments such as those made by Brown. I found this very enjoyable as I viewed it as a self-test of my analysis skills.

The book concludes with final comments by Porte that succinctly reiterate the main points made in the chapters in support of the philosophical justification and the means of effective implementation of replication research in applied linguistics. I would particularly recommend the book to inexperienced researchers in that it provides the framework to start a research project with a degree of confidence based on the chosen model. This is, of course, one of the central strengths of replications for research training as discussed in Part II. Additionally, as the author has achieved his aim of proposing a very comprehensive, convincing argument for the full acceptance of replication studies in applied linguistics, experienced researchers will find the book very instructive. Since replications are dependent on well-designed original studies which include sufficient information and data to enable replication, researchers might benefit by assessing both past and future studies with these factors in mind.

References

Eckerth, J. (2009). Negotiated interaction in the L2 classroom. *Language Teaching, 42*, 109-130. http://dx.doi.org/10.1017/S0261444808005442

Valdman, A. (1993). Replication study (editorial introduction). *Studies in Second Language Acquisition, 15*, 505*.*

Agendas for Language Learning Research. Lourdes Ortega, Alister Cumming, and Nick C. Ellis (Eds.). Malden, MA: John Wiley & Sons, 2013. vi + 215 pp.

Reviewed by
David Beglar
Temple University, Japan Campus

Language Learning has long been one of the premier journals in the field of SLA, and its importance shows no signs of diminishing given the quality of the research published in the journal and the researchers who make up its editorial board. The *Currents in Language Learning Series* is a biennial supplement to the journal, which was created to provide state-of-the-art overviews of important issues in the field and their applications to formal and informal first and second language acquisition. Articles in the supplement reflect a multidisciplinary perspective; this can easily be seen in the form of influences from diverse fields such as linguistics, psychology, sociology, and cognitive neuroscience. The first publication in the series, *Agendas for Language Learning Research* is an outstanding collection of articles written by editorial board members and editors of *Language Learning*.

The genesis of the articles is worthy of emulation, as the authors first met at an invitational conference where they made presentations. The presentations were followed by interactions among the authors that led to revisions of their early manuscripts. External reviewers and the journal editors then provided further feedback to the authors, and the resulting manuscripts make up the contents of this publication. One significant benefit of beginning with the conference is that a full set of podcasts of the talks the authors gave is available on the *Language Learning* website. I found listening to the podcasts illuminating, as they changed my perceptions of the information in the articles in subtle and significant ways.

The publication is made up of 10 articles. In the opening article, Ortega discusses progress in the field of SLA by delineating four major trends that have taken place over the past 15 years and then by arguing that the broadening of the field through what she calls *transdisciplinarity*, "the proclivity to pursue and generate knowledge that can be seen as relevant across many disciplinary boundaries" (p. 6), has made it increasingly informative to researchers in related fields. For instance, she notes that influences from SLA are found in leading journals in the fields of psycholinguistics and bilingualism. In the following article, Ellis, Matthew Brook O'Donnell, and Ute Römer describe ways in which usage-based, statistical learning, which involves cognitive abilities such as the acquisition of prototypes and a sensitivity to semantic and syntactic usage patterns, potentially contributes to and explains adult foreign language acquisition. Their view of language acquisition is strongly cognitive, as they propose that language learning is based on the same cognitive processes that apply to other forms of human learning and that language acquisition is the product of the complex, emergentist processes that occur in many natural phenomena. Robert DeKeyser focuses on one of the key issues in the field of language acquisition, the effect of age. DeKeyser explains why the issue is important, discusses conceptual and methodological problems that have plagued research in this area, and then provides concrete suggestions for moving forward with research on the effects of age. Kathleen Bardovi-Harlig discusses what she calls *acquisitional pragmatics*, or the development of pragmatic competence. She suggests five areas for future research, including the measurement of change in pragmatic competence over time and the interface of linguistic knowledge and pragmatic development. Scott Jarvis writes about measuring lexical diversity in learner speech and writing, an area that is but one aspect of the most vibrant research areas in the field: second language lexical acquisition. Jarvis argues that linguists should adopt the more theoretical approach used by ecologists and develop mathematical models that permit more accurate estimates of lexical diversity. Diane Larsen-Freeman takes up a topic that has been largely neglected in the field of SLA: transfer of learning. She provides an informative review of research conducted outside of the field of SLA on transfer of learning and then reframes the concept of transfer by partially relying on complex systems theory. Cumming concerns himself with the development of academic literacy among learners from diverse cultures by considering how language, literacy, and culture interact in second language classrooms, while in a related article on the acquisition of academic language proficiency, Mary Schleppegrell emphasizes the importance of meaningful interactions among learners, explicit attention to linguistic form, and the

teaching of metalanguage. Schleppegrell's approach is situated within what has been termed the social turn (Block, 2003) in SLA. Richard Young and Alice Astarita consider the role of the social environment as viewed through Practice Theory, a philosophical and methodological framework based on the ideas of postmodern thinkers such as Bourdieu and Foucault. Their goal is to move toward so-called "alternative approaches to SLA" by emphasizing the sociocultural context that learners bring to the task of language acquisition. In the final article in the volume, John Schumann discusses 11 strategies humans have developed to deal with the difficulties associated with adult second language acquisition. He then considers evolutionary reasons for why adults encounter these difficulties.

The strengths of this volume are many. Readers will encounter new ideas and get a glimpse of directions the field might move toward in the next few years. The breadth of the approaches to investigating SLA used by the authors is impressive, and to a degree, the articles in this volume represent many of the most important strands of thought moving in the field of SLA at present. However, like articles in *Language Learning*, understanding those in *Agendas for Language Learning Research* often requires knowledge of statistics, an appreciation for the scientific method, some knowledge of research outside of the field of SLA, and an understanding of a wide range of technical terminology. In addition, the papers are written by researchers for researchers; for this reason, practicing foreign language teachers will find little they can use in terms of pedagogical practice.

The authors of the papers are at the top of their profession and their writing is therefore dense, often abstract and theoretical, and for the most part, they have adopted a strongly scientific approach to studying and talking about SLA. These characteristics make the articles challenging to understand; however, for persons wishing to learn about cutting-edge research in the field, and for those conducting SLA research, this publication provides valuable information and shows multiple avenues for moving the field forward. In sum, the authors of the articles in this volume have done an admirable job of providing informative, state-of-the-art reviews of issues that will likely continue to be of considerable importance in the field of SLA in the coming years.

References

Block, D. (2003). *The social turn in second language acquisition*. Washington, DC: Georgetown University Press.

The Roles of Languages in CLIL. Ana Llinares, Tom Morton, and Rachel Whittaker. Cambridge: Cambridge University Press, 2012. vii + 344 pp.

Reviewed by
Howard Brown
University of Niigata Prefecture

Content and Language Integrated Learning (CLIL) is an educational approach growing in popularity in Japan and more and more CLIL programs are being started in all levels of education from primary to university. In fact, MEXT (2006) reports that programs that can be thought of as CLIL, or CLIL-like, are running at as many as a third of all universities in Japan. These programs are also being written about in the academic literature. In 2013, the *Asian EFL Journal* published a special issue on CLIL in Asia, featuring more than a dozen articles on CLIL practices in Japan, and a special edition of the *International CLIL Research Journal* focused entirely on Japan is in the works. Amid all this interest in CLIL amongst Japanese scholars and educators, *The Roles of Languages in CLIL* is a very timely book for readers of *JALT Journal*.

The first thing that strikes a reader about this book is the title, in particular the use of the plural *roles*. Llinares, Morton, and Whittaker explore language use through a 500,000-word corpus collected from English-medium CLIL classrooms in Spain, Austria, the Netherlands, and Finland. The writers delve into this corpus to look at multiple roles of language in both teachers' and students' usage. First, they explore registers of classroom language for regulation and instruction. They also look at language *of* learning (language needed to express content ideas), language *for* learning (language needed to participate in learning activities and classroom tasks), and language *through* learning (language which emerges as content learning stretches the students' ability to express ideas).

The book starts with an overview of how the writers define CLIL, which for them is primarily a social approach to learning, resting at the intersection of Halliday's systemic functional linguistics and Vygotsky's sociocultural theory, as well as upon social models of SLA from Lantolf and others. This view of CLIL, found throughout the book, informs the writers' analysis of classroom language.

The fundamental structure of this book is a three-part framework for the analysis of language use in CLIL based on classroom interaction, subject

literacies, and language development. The book itself parallels the structure of this framework. Following the introduction, the book is divided into three major parts.

Part 1 deals with English as a classroom language in CLIL contexts. The writers look at the interactions between teachers and students as well as among students. These interactions frame not only the social world of the classroom but also the development of language proficiency and the construction of content knowledge (i.e., the students' understanding of the discipline-specific knowledge being taught). In this part the writers also examine the teacher's use of language for instruction. They argue that teachers should shift away from authoritative, instructional language to more dialogic questioning. They look at how both teachers and other students use language to take on the role of "expert other" to support the construction of meaning and development of understanding. This scaffolding of meaning and understanding is a key element of the understanding of CLIL as social learning.

Part 2 shifts focus to deal with the idea of subject literacies. The writers look at genre and register in disciplines commonly taught in CLIL classrooms: science, geography, and history. They focus both on genres that students will encounter in teaching materials and genres the students themselves will be expected to produce. They also look at the grammar and lexis of English used in CLIL classes and show how the learning of lexis is tied to development of content understanding.

In Part 3 the writers examine the learner's language development. They argue that CLIL students tend to develop language proficiency starting with more spoken interaction (both input and output) and move into text-based work later, as their language proficiency develops. They then explore how successful CLIL teachers use feedback in oral exchanges to help students develop along this path through focusing on meaning, form, and register. They also look at how CLIL activities can be structured to encourage the learners' development of both academic literacies and interpersonal communication proficiency.

The book ends with recommendations for integrating language and content assessment in CLIL. Llinares, Morton, and Whittaker show that in the contexts they studied, assessment in CLIL classes is often based on the content knowledge students are expected to develop. Language assessment, when it is done at all, is often done separately, based on assessment criteria from language classes. However, the authors argue that this conventional language assessment may not be fair to CLIL students who have developed

their language proficiency differently than have students in language classes. Instead, they make a case for formative assessment of language in CLIL based on the students' content knowledge and their ability to work with and express that knowledge in the L2 over the duration of the course.

One key feature of this book is the addition of discussion tasks at the end of each chapter. These questions help the reader reflect on the contents and connect what they have read to their own teaching context. As such, these tasks seem designed to make the academic text of the book more accessible to practicing CLIL teachers.

A possible weakness of the book, at least for Japan-based readers, is the strong focus on CLIL as a European phenomenon. In the introduction, the writers make it clear that they see CLIL as European. Their corpus is based on samples from four European settings and when they define CLIL, they describe it as a teaching approach that grew out of a series of bottom-up innovations in language and content teaching (largely in primary and secondary schools) in Europe over the past 20 years. While it is part of the larger trend towards second language (often English) content instruction around the world, and the acronym CLIL is often used as an umbrella term for all kinds of second language content instruction, the European vision of CLIL has a distinct identity for these writers.

However, readers in Asia or other contexts will not find this focus on Europe overly limiting. The writers are careful in the introduction to clearly outline the sociolinguistic, sociopolitical, and educational situation that CLIL occupies and this clear description allows teachers in other contexts to compare their own situation with what is happening in Europe. This comparison gives the reader a lens through which to understand the rest of the book and apply its lessons. This makes *The Roles of Language in CLIL* an essential text for anyone involved in second-language-medium education.

References

MEXT (Japan Ministry of Education, Culture, Sports, Science and Technology). (2006). *Curriculum kaikaku no jisshijoukyou* [Current state of curriculum renewal]. Retrieved from http://www.mext.go.jp/b_menu/houdou/20/06/08061617/001.htm

***Materials and Methods in ELT: A Teacher's Guide (3rd edition).*
Jo McDonough, Christopher Shaw, and Hitomi Masuhara.
Chichester, UK: Wiley-Blackwell, 2013. xiv + 334 pp.**

Reviewed by
Tyler Burden
Tokyo University of Agriculture and Technology

Materials and Methods in ELT is an introductory book on the principles and practice of teaching English as a second or foreign language. It is intended to be of use to both practicing teachers who wish to keep up-to-date with developments in ELT and to those enrolled in language teaching courses. There are several aims, principally to provide readers with a bridge between theoretical considerations and the practical design of materials and methods. The authors also aim to show how trends in materials and methods in ELT have changed as time has passed and state in the preface that they hope to equip readers to be able to critically assess materials for their own teaching situations from a more informed perspective.

This is the third edition of a book that was first published in 1993. In its original form, it was cowritten by two authors, Jo McDonough and Christopher Shaw. For the third edition, Hitomi Masuhara joined the team to help update the material and Diane Slaouti contributed a chapter on technology in ELT. Since publication of the first edition, various changes have been made. The references throughout the book have been updated and more illustrative samples from current teaching materials have been included. Also, some material to illustrate a task-based learning approach has been incorporated. Although the back cover states that there are new sections on assessment and feedback, it appears that these have not been added.

The book is divided into three sections. The first is titled "Topics in the design of materials and methods." This includes a discussion of the various factors influencing teachers' selection of materials, a look at current approaches to materials and methods, and in particular, a discussion of the ubiquity of the communicative approach. It includes chapters on evaluating and adapting ELT materials along with the aforementioned chapter on technology that points the reader towards various online resources. It doesn't, however, provide guidance on actually designing materials in the sense of giving advice on how to create new materials. It is focused, instead, on analyzing the design of existing materials.

In the second section of the book, the authors look at teaching language skills. There are five chapters, one for each of the four skills of reading, writing, listening, and speaking and an additional chapter on integrated skills teaching. In these chapters, the authors look at the issues associated with, and the rationale behind, the teaching of these skills, before offering a few practical examples to illustrate these ideas. The examples are taken mainly from contemporary textbooks. The integrated skills chapter includes ideas on how combinations of the four skills can be woven together to form individual lessons or longer schemes of study.

The third section of the book deals with classroom methods. There are four chapters, one on group work and pair work, one on individualization and learner training, and another on observation in the classroom. The fourth chapter in the section is a discussion of the teacher's role and offers some ideas for professional development and small-scale teacher-led research.

The authors of *Materials and Methods in ELT* all appear to be lecturers on MA courses in TESOL or TEFL and the first thing that strikes the reader about this book is that it feels as if it has been designed to function as a textbook for such a course. For the general reader, this can be both a good and a bad thing. On the one hand, the referencing and further reading information is excellent and, in areas where the reader's interest is piqued, it is easy to find follow-up titles to explore these areas in more depth. On the other hand, almost every page features insets where the reader is invited to reflect (and often discuss with colleagues) about the various issues raised. This would no doubt work well in a classroom setting where a group of learners may generate a range of ideas, but for individual readers it may add little to the experience. Generally, when the reader is asked to reflect, little feedback is provided.

Large parts of the first and third sections of this book were rather heavy on verbiage and light on content. I found myself rereading several parts in order to try and work out what, if anything, was being said. The book was also not helped by the use of redundant diagrammatic representations. If we are told that reading is a two-way process between reader and text, do we really need to see a picture with a two-way arrow pointing to the words *reader* and *text* (as on p. 113)? Many of the diagrams in this book seemed similarly unnecessary.

The authors state that they aim to help their readers evaluate materials and they devote an entire chapter to this. However, I found the ideas within the chapter to be largely commonsensical. The authors unnecessarily intro-

duce jargon such as *micro-* and *macro-evaluation* to simply refer to a detailed look at the contents of a textbook as opposed to a quick look at the index. The reader is then informed that if we judge from the macro-evaluation that the book does not meet our needs then we can skip the micro-evaluation. There are plenty of examples of similarly unnecessary statements.

On the positive side, I found the new chapter on technology in ELT to be very interesting and often noted various websites that I had previously been unaware of. All teachers are restricted to some extent regarding the degree to which they can use technology in their classrooms. However, even if one just wishes to make more attractive or better informed worksheets, a number of websites are suggested for this. I particularly liked the introduction to the Wordle website (http://www.wordle.net/).

The second section of this book, dealing with teaching the skills, was also useful. The issues and theory are succinctly explained and interesting, practical examples of classroom activities are provided to illustrate the ideas. However, in order to more fully meet the stated aim of providing a bridge between theory and the practical design of materials, the chapters required far more examples than were provided in this section, I felt.

One aim that the authors have successfully met is in showing how trends in methods and materials in ELT have changed over the years. These changes are highlighted throughout the book. It is almost possible to trace the changes in methodology over the years simply by the titles of popular textbooks of the day. The message that comes across is that what is considered best practice in ELT has changed considerably in the past and, no doubt, will continue to change in the future.

Materials and Methods in ELT is now into its third edition, which clearly suggests it has a market. Given that most of this book is not really about materials at all, one wonders if *An Introduction to ELT* wouldn't have been a more appropriate title. I had hoped that this book would improve my own attempts at writing materials, but felt I gained little from reading it. If there is merit in this book, I think it is as a general introduction for novice teachers but there is little to recommend for experienced teachers.

Teaching and Learning Second Language Listening: Metacognition in Action. Larry Vandergrift and Christine C. M. Goh. New York: Routledge, 2012. xx + 315 pp.

Reviewed by
Robby Caughey
Kanto International Senior High School

Decades of research in SLA have demonstrated that comprehensible input is a driving factor in SLA. Unfortunately, in traditional listening pedagogy students are merely asked to answer comprehension questions based on listening texts but this fails to teach students *how* to listen. In *Teaching and Learning Second Language Listening: Metacognition in Action*, Vandergrift and Goh present a metacognitive approach to listening that has a firm theoretical foundation and seeks to teach students how to listen.

The book is divided into three parts. In Part I, the authors build up the theoretical foundation of their metacognitive approach. After they take readers through the history of listening pedagogy in Chapter 1, they provide a thorough discussion of top-down and bottom-up processing. Chapter 2 then shows how these relate to Anderson's (1995) three cognitive processes of listening: (a) perception (creating a phonological representation of the stream of speech in working memory), (b) parsing (segmenting the resulting phonological representation into meaningful units), and (c) utilization (linking the meaningful units with relevant knowledge from long-term memory). In Chapter 3, the authors map these three cognitive processes and metacognition onto Levelt's (1993) model of speech production and comprehension. Chapter 4 reviews studies on the cognitive and affective factors that influence foreign language listening comprehension.

Part II consists of six chapters that provide readers with a detailed look into a complete listening program in which learners are shown not only how to listen but also how to take control of their learning—an approach which incorporates cognitive and sociocultural views of language learning. After explaining the importance of metacognition and how it controls strategy use, in Chapter 5 the authors introduce four metacognitive processes: (a) planning for listening tasks, (b) monitoring comprehension, (c) solving comprehension problems, and (d) evaluating the approach and outcomes. Chapters 6 through 10 present an array of activities that teachers can introduce for learners to do both in and out of class to develop these metacognitive processes while improving their listening ability.

Chapter 6 presents an evidence-based metacognitive pedagogical sequence for listening instruction. This pedagogical sequence takes learners through five stages: (a) making predictions, (b) checking predictions and adding notes, (c) revising and adding more notes, (d) listening and reading the transcript, and (e) reflecting on their listening experience and setting goals for the next listening task. Research has shown that this pedagogical sequence builds metacognitive knowledge and improves listening comprehension. Because strategy instruction is embedded in the sequence, learners develop their ability to deploy appropriate strategies to comprehend texts. The authors provide a general worksheet (p. 113) that can be used with any listening text, and they also take readers through the steps of developing a customized worksheet for their own specific listening content. Chapter 7 takes a broader approach to teaching listening by introducing a number of activities that enable learners to focus on their self-concept, motivation, and anxiety as well as developing metacognition and listening abilities.

After introducing these activities, Chapters 8-10 contain descriptions of each activity in greater detail. In Chapter 8, the authors recommend providing learners with listening perception activities in the postlistening phase of the lesson to develop their bottom-up listening skills as well as language analysis activities to propel language development. In Chapter 9, the authors show readers how to prepare units of work that develop learners' one-way and interactive listening competence through task-based learning. The authors suggest principles for the selection of appropriate listening texts so that students are exposed to a broad range of real-life listening events and build up their knowledge of the discourse structure of these events. They also provide useful lists that assist teachers and course designers in selecting prelistening activities that prepare students for the listening task, one-way and interactive tasks that elicit desired listening skills and outcomes, and postlistening activities that enable students to elaborate on information they obtained from the listening task. Since learners know they will have to use the information they obtain from the listening task, they will be more engaged with the listening and use a broader repertoire of skills and strategies to process the input. The sample lesson plans at the end of the chapter show readers clearly how all of these elements can be brought together in the classroom.

As classroom instruction cannot possibly provide students with an adequate amount of input for significant development of listening abilities, in Chapter 10 the authors recommend activities students can do out of class. The authors do not support simply telling students to listen to English out-

side of class. They suggest providing students with worksheets that guide them in their independent listening studies. In self-directed listening, learners respond to prompts that focus on different aspects of metacognition and are guided through repeated listening of the same text. The authors also recommend listening diaries with prompts to direct learner reflections on different aspects of metacognitive knowledge. In this way, students understand what they should write about and develop insights into progress with their listening.

Part III is made up of two chapters in which the authors explore multimedia options and discuss assessment. Chapter 11 reviews research on the effectiveness of multimedia in improving listening ability. Although the research results are mixed, the authors shed some light on the role of technology in listening pedagogy and end the chapter with guidelines for using multimedia. In the final chapter of the book, they attempt to tackle the difficult issues involved in assessing listening. Although the authors could not do justice to the issues of validity, reliability, and washback in a single chapter, they do provide an overview of these concepts and the importance of formative as well as summative feedback. The discussion on how teachers can use the Common European Framework of Reference for languages to enable learners to plan, monitor, and evaluate their learning may be particularly appealing to those interested in fostering student autonomy.

Teaching and Learning Second Language Listening: Metacognition in Action should be read by teachers who want to teach their students how to listen rather than simply check their comprehension. Teachers will be delighted with the wealth of practical ideas, sample lesson plans, and accompanying worksheets. Students in teacher-training programs will also appreciate the classroom vignettes at the beginning of each chapter that bring the issues to life, and they will benefit from the questions and recommendations for further reading at the end of each chapter. This is an indispensable book for anyone interested in listening pedagogy or metacognition.

References

Anderson, J. R. (1995). *Cognitive psychology and its implications* (4th ed.). New York: Freeman.

Levelt, W. J. M. (1993). Language use in normal speakers and its disorders. In G. Blanken, J. Dittmann, H. Grimm, J. C. Marshall, & C.-W. Walesch (Eds.), *Linguistic disorders and pathologies* (pp. 1-15). Berlin, Germany: De Gruyter.

Computer-Assisted Language Learning: Diversity in Research and Practice. **Glenn Stockwell (Ed.). New York: Cambridge University Press, 2012. xii + 228 pp.**

Reviewed by
Gilbert Dizon Jr.
Himeji Dokkyo University

While technology has become ubiquitous in our everyday lives, the myriad of technologies available has made it difficult to effectively integrate computer-assisted language learning (CALL) into language teaching and research. Combine this technological diversity with the variability amongst learners and institutions and the task can be daunting. As a result, in *Computer-Assisted Language Learning: Diversity in Research and Practice*, the authors aim to introduce readers to these diversities so that they can make informed decisions that are appropriate for their own instructional and research contexts.

The book is a compilation of 10 chapters that cover the diversity in CALL. This includes the technologies and environments in which they are used, the pedagogies employed, and the users involved, as well as the research methods utilized. After the introductory chapter, Chapters 2-9 follow a consistent format, first introducing readers to the general issues related to the chapter theme, followed by at least one practical example from the authors' own experience or relevant literature. Potential teaching and research implications based on the chapter theme are then discussed along with examples so that readers can consider how the options presented may be of use in their particular settings.

Stockwell begins Chapter 1 by providing an overview of some of the key issues in CALL. As he notes, the affordances of technology, or how technology affects the learning process, are especially important to consider, as these affordances will impact whether technology supports or inhibits language learning.

The next three chapters are concerned with learners, who Stockwell describes as the starting point for CALL. The second chapter, by Robert Fischer, highlights how learners use CALL materials. Based on the research discussed in the chapter, Fischer concludes that learners often use technologies for purposes different than those intended by the developers and may even overuse a single component in a program. In Chapter 3, Philip Hubbard

and Kenneth Romeo provide an overview of the possibilities for learner training. Nevertheless, the authors suggest that due to limited research in this area, an optimal training process has yet to be developed. In Chapter 4, Hayo Reinders and Pornapit Darasawang discuss ways to increase learner autonomy though CALL materials. A recurring theme in these chapters is the significance of learner training. Even though today's students are sometimes referred to as "digital natives" (Prensky, 2001), these authors caution us that such learners may not know how to use technology for language learning purposes. Therefore, the conclusions arrived at are that it is essential to provide students with technical (how), strategic (what), and pedagogical (why) training before they begin using technology to support language learning.

Stockwell and Nobue Tanaka-Ellis state that although the technologies that are chosen influence one's teaching context, "the environment itself will determine the degree to which these technological affordances are applicable" (p. 86). In the fifth chapter, they discuss these affordances in relation to the four environments in which computers are used for language learning and teaching: face-to-face, blended, distance, and virtual. Face-to-face is clarified to mean when learners "interact directly with the computer individually, or work together in pairs or small groups at a single computer to orally discuss any information that they read from or input into the computer" (p. 69). Subsequently, an example of CALL being used in a face-to-face environment with an online element is given in which Australian high school students learning Japanese interacted with high school students in Japan through a bulletin board system (BBS). Before posting to the BBS, the Australian students interacted with each other and the teacher to read messages from the Japanese students and draft messages to them. Thus, the authors found that technology facilitated language use in two distinct ways: orally through peer and teacher discussion around the computer and in text-based communication via the BBS.

In Chapter 6, E. Marcia Johnson and John Brine examine how open educational resources (OERs) together with open source software (OSS) tools can influence educational settings by making relevant content more accessible. Despite their potential benefits, such tools and resources may not be effectively incorporated at the institutional level due to a lack of structured implementations, the absence of realistic goals, and insufficient e-learning training. Therefore, the authors stress that ongoing teacher support, effective pedagogical approaches, and clear goals based on local contexts and constraints are essential in order to take full advantage of OERs and OSS.

The seventh chapter is about the impact that modalities have on our understanding of CALL. Through an analysis of four metastudies, Marie-Noëlle Lamy illustrates that modality is a misunderstood concept that is often ignored in language research. Given this, she advocates for a cultural approach to the understanding of technology-mediated learning based on semiotics or how modes work together to convey meaning.

Chapter 8 focuses on the technologies used in language teaching. Gordon Bateson and Paul Daniels highlight current trends and sort these technologies into four categories: single-server, multi-server, single-computer, and mobile. Afterward, three examples are presented that outline how these technologies were used in a learning context. Given the rise of smartphones, the two mobile technology examples are the most intriguing. In the examples, not only did mobile technologies increase learner autonomy but they also allowed students to create authentic content outside of the classroom.

The penultimate chapter, by Stockwell, investigates the complex relationship between technology, research, and practice by examining the research approaches taken in a sample of articles from a 10-year span (2001-2010) in four prominent English language CALL journals. One approach that rarely appears in the literature but should be taken more seriously is the pedagogy-based approach (Colpaert, 2006). Compared with other research approaches that are focused on a predetermined technology, this approach identifies the specific needs of a language-learning environment before making technological choices.

In the final chapter, Stockwell examines diversity at the learner, institutional, and societal levels. Rather than operating independently from one another, these factors work together to influence the choices researchers and practitioners make about technology according to the contextual factors involved. Consequently, Stockwell emphasizes that understanding how each level of diversity affects a setting is critical for educators, as this will determine not only the learning objectives but also the technologies to be used.

The book concludes with the following question: Should the diversity in CALL be embraced or is it more of a hindrance? Ultimately, the answer to this question will vary depending on the specific needs of the learners involved as well as the institutional resources that are available. Nevertheless, this book will help readers tackle this question with a better understanding of the most important issues pertaining to language learning through CALL. Therefore, I recommend it as a guide to teachers who are interested in using technology to enhance language learning and to researchers who would like to contribute further to the understanding of CALL environments and ap-

plications. However, as the total message of these chapters points out, when deciding to implement technology in a given context, it is important to recognize that technology will never replace sound pedagogy. In other words, it is not *what* technologies are used that determines successful language learning but *how* they are used.

References

Colpaert, J. (2006). Pedagogy-driven design for online language teaching and learning. *CALICO Journal*, *23*, 477-497.

Prensky, M. (2001). Digital natives, digital immigrants. *On the Horizon*, *9*(5), 1-6.

Teacher Research in Language Teaching. Simon Borg. Cambridge: Cambridge University Press, 2013. xiii + 253 pp.

Reviewed by
Nick Doran
Rikkyo University

Simon Borg of the University of Leeds has published extensively in the field of teacher cognition—namely what language teachers know, believe, think, and do (Borg, 2003). His work has provided insights into the thought processes involved when teachers make teaching decisions as illustrated by a recent *JALT Journal* article in which he looked at the grammar teaching practices and teacher cognitions of three Japanese high school English teachers (Nishimuro & Borg, 2013). In this book, *Teacher Research in Language Teaching*, he delves much deeper into teacher cognition by examining the attitudes towards research and the research practices of over 1700 ELT professionals in the roles of teachers and managers worldwide.[1] In this book based on empirical data, Borg does a commendable job of collating vast amounts of quantitative data and presenting them to the reader in a clear and understandable way. There is also a wealth of qualitative data provided in the form of respondent comments, offering comprehensive insights into the minds of those involved in English language teaching that were collected across multiple studies in his programme of research.

Chapter 1 establishes various definitions of teacher research and provides a brief overview of its origins. Some of the recurrent themes that appear

throughout the book are introduced here including the benefits of teacher research and the various barriers that could hinder research. Finally, various critiques of teacher research are presented. These uncover the disparaging views of some academics that such research can be too localized and of poor quality.

Chapter 2 introduces the methodology used for eliciting opinions from language teachers and managers on teacher research. Borg carefully takes the reader through the various methodological instruments used to collect data (available in the book's appendices) and provides a clear explanation of the processes involved in analyzing the data. For each stage of data collection, the methodological challenges that were encountered are discussed together with how they were overcome. Overall, this chapter succinctly sets out how to conduct a quality teacher cognition research project and is invaluable for those who are interested in the nuts and bolts of doing this type of research.

Chapters 3-7 focus on differing themes of language teacher research engagement. In Chapter 3, Borg deals with what teachers and language program managers conceptualize as research by presenting a list of 10 scenarios and asking teachers and school managers to decide if they count as research or not. The list includes examples such as an informal action research project or collecting student feedback in order to make course improvements. Results show that teachers' conceptions of research generally take the more standard view associated with hypothesis testing and statistical analyses. Given the challenges involved in testing hypotheses and conducting a quantitative-based research project, Borg believes that this restricted conceptualization could hinder teachers' broader engagement in research.

Chapter 4 presents results on how often language teachers read research reports and the kinds of research they read about. Teachers appear to be most interested in anything that provides practical teaching ideas, with publications such as *The Modern English Teacher* a popular choice. More academic journals such as *ELT Journal* are often viewed as dry, meaning they are full of jargon or difficult statistics and are often perceived as irrelevant to what goes on in the classroom. Throughout this text, direct quotes from teachers and managers are used to exemplify the main points, and these are particularly interesting to read.

Chapter 5 deals with the frequency with which teachers engage in research and the reasons for not doing so. It also deals with the impact that research has had on their teaching practice. Borg does an excellent job of

carefully laying out the benefits to teachers engaging in research. However, it is evident that considerable work is needed to overcome the many barriers that face teachers, which are further explored in chapters 6 and 7.

In Chapter 6, Borg examines the relationships between teacher research engagement and teaching quality. A range of positions are discussed with the negative stances towards research most interesting, such as opinions that effective teaching could be achieved without any knowledge of research and that many teachers who are drawn to research lack basic teaching skills. Such opinions were unsurprising—reflecting the opinions of many colleagues who consider research to lack any personal value. Borg tries his utmost to argue against these negative positions by asserting that research "can contribute to a wider range of knowledge, skills, and sensitivities which are needed for effective practice" (p. 217). The difficulties in engaging teachers in research are further detailed in Chapter 7, which deals with how work environments can support (or not) teacher research. Borg makes practical suggestions for schools to facilitate teacher research such as giving teachers paid time off or funding to attend conferences. However, personal experience in Japan would suggest that many of these are unfeasible considering the economic realities faced by many institutions, particularly those outside of tertiary education. The harsh realities for teachers (e.g., part-time contracts, heavy teaching loads) that could hinder research engagement are also presented. Eventually, Borg asserts that "collective action is needed to elevate TEFL from . . . its current status as a domain of activity that lacks a strong professional ethos" (p. 219). Apart from encouraging teachers to engage more in research, it is difficult to envisage how this can be achieved.

In Chapter 8, Borg looks at a number of projects to engage language teachers in research. Some reasons why most were unsuccessful are explained by two examples of good practice that Borg was personally involved in. Following these case studies, Chapter 9 provides a list of 18 recommendations to enable teacher research with a greater chance of success. Surprisingly, in my own teaching context, which I previously felt was conducive to promoting teacher research, only three recommendations were satisfied.

Overall, this is a book with fascinating insights into the collective minds of EFL teachers and school managers worldwide. However, the question remains: Who will read this? Hopefully the answer is teachers, as it might spur them into engaging in research. However, this is not a how-to book and teachers interested in learning how to do research are advised to look at alternative texts such as Burns (2010) or Dörnyei (2007). Judging by the book's findings, that most teachers are put off from either reading about

or doing research, it seems that teachers will not naturally be drawn to this title. Those involved in teacher development such as school managers, principals, directors of studies, and faculty development committee members could read this book and take positive steps to instigate teacher research projects. However, if faced with insufficient research budgets, it could be difficult for educational institutions to implement many of Borg's suggestions. The real target audience for this book is people who are seriously committed to promoting teacher research or those who are interested in the burgeoning area of teacher cognition. For these readers, this book is recommended as essential reading.

Notes

1. At the IATEFL Conference 2013, Borg presented the main findings of this book, available at <http://iatefl.britishcouncil.org/2013/sessions/2013-04-09/cup-signature-event-research-engagement-and-teaching-quality>

References

Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching, 36*, 81-109. http://dx.doi.org/10.1017/S0261444803001903

Burns, A. (2010). *Doing action research in English language teaching: A guide for practitioners*. New York: Routledge.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.

Nishimuro, M., & Borg, S. (2013). Teacher cognition and grammar teaching in a Japanese high school, *JALT Journal, 35*, 29-50.

Information for Contributors

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

Editorial Policy

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

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

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

Guidelines

Style

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

Format

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

Submission Procedure

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

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

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

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

Restrictions

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

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

Please send submissions in these categories or general inquiries to:

jj-editor@jalt-publications.org

Melodie Cook, *JALT Journal* Editor

Japanese-Language Manuscripts

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

jj-editorj@jalt-publications.org

Ken Urano, *JALT Journal* Japanese-Language Editor

Reviews

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

jj-reviews @jalt-publications.org

Greg Rouault, *JALT Journal* Reviews Editor

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

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

日本語論文投稿要領

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

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

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

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

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

投稿原稿送り先またはお問い合わせ:

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

JALT2014

Conversations Across Borders
November 21–24, 2014 in Tsukuba, Ibaraki

JALT2014: Conversations Across Borders

40th Annual International Conference on Language Teaching and Learning & Educational Materials Exhibition

Friday, Nov 21 – Monday, Nov 24, 2014

Tsukuba International Congress Center (Epochal Tsukuba)

Tsukuba, Ibaraki, Japan

Language learning and language teaching carry us across all sorts of borders: national, cultural, disciplinary, psychological, and, of course, linguistic. JALT's 40th annual conference will celebrate this phenomenon with its theme, "Conversations Across Borders."

Plenary Speakers

- *Claire Kramsch - Professor, UC Berkeley*
- *Thomas Farrell, Professor, Brock University*
- *Momoko Nakamura - Professor, Kanto Gakuin University*
- *Bill Harley - Grammy-award winning musician, story-teller, and author*

jalt.org/conference